

# **ORAL LEUKOPLAKIA – AN INVESTIGATION OF ITS MICROBIOME AND OF THE ASSOCIATION OF CLINICAL AND MICROBIAL FEATURES WITH THE DEGREE OF DYSPLASIA**

A thesis submitted to the University of Dublin in fulfilment of the  
requirements for the degree of Doctor of Philosophy

Sheila Galvin  
February 2024

Department of Oral Medicine  
Dublin Dental University Hospital  
School of Dental Science  
Trinity College Dublin, the University of Dublin

# Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

A handwritten signature in black ink, appearing to read 'Sheila Galvin', is centered within a light gray rectangular box.

Sheila Galvin  
24 February 2024

## Summary

The primary aim of this study was to investigate if oral leukoplakia (OLK) was associated with changes in the microbiome and whether these changes differed with differing degrees of dysplasia. Secondary aims included examining which clinical and patient factors were associated with the degree of dysplasia of OLK, establishing how other factors such as smoking affected the microbiome and finally, if changes in the microbiome could be used to predict whether an OLK was low (LROLK) or high risk (HROLK).

Following ethical approval, swabs were taken from histologically proven OLKs and contralateral normal (CLN) sites in patients who met the inclusion criteria. Healthy controls (HC) were also sampled at various mucosal sites. After DNA extraction, the V1-V3 region of the 16S rRNA gene was sequenced. RStudio was used for statistical analysis.

A total of 186 patients with 225 OLKs were included in the analysis of clinical and patient features. The degree of dysplasia was found to be significantly associated with OLK site and clinical appearance, with location on the tongue or floor of mouth and non-homogenous clinical appearance both significantly associated with severe dysplasia ( $p=0.0005$ ). OLKs on the dorsum/lateral tongue had a 38.6 fold, and on the floor of mouth/ventral tongue a 70.4 fold, increased relative risk of being severely dysplastic compared with OLKs on the buccal mucosa, while non-homogeneity was associated with a 25 fold increased odds of having moderate or severe dysplasia.

Two hundred and two CLN and 120 HC swabs from 241 subjects were then examined to investigate the influence of smoking, alcohol consumption, oral hygiene practices and mucosal site on the microbiome of healthy mucosa. Analysis using the R package MaAsLin2 found that smoking and oral site had the most significant influence on the oral microbiome. Smoking was associated with a shift to a more periodontal pathogen-heavy microbiome dominated by *Prevotella*, *Actinomyces*, *Porphyromonas*, *Fusobacterium*, *Tannerella*, *Parvimonas*, *Filifactor*, *Clostridiales* and *Peptostreptococcus* species, and a reduction in healthy mucosa-associated species such as those of *Neisseria*, *Streptococcus* and *Rothia* genera. Site specificity was also identified with tongue sites dominated by *Rothia* and *Veillonella* spp. and palatal sites dominated by *Prevotella* spp., *Actinomyces* spp. and *Streptococcus* spp. Floor of mouth/ventral tongue and buccal mucosa/alveolus/gingiva sites displayed very similar microbial communities dominated by streptococcal species.

Next, 216 OLK sites and 202 CLN sites from 182 patients were compared to establish if there were microbiome changes specific to OLK. After adjusting for multiple variables using the R package MaAslin2, OLK was consistently associated with increased abundance of five taxa: *Bergeyella* sp. HMT322, *Streptococcus australis infantis* clade 431, *Gemella morbillorum*, *Leptotrichia* sp. HMT215 and *Leptotrichia* sp. HMT392. Adjusting for oral site and smoking identified several other OLK-associated species, including two *Fusobacterium* species, *Porphyromonas* sp. HMT284, *Lachnoanaerobaculum umeaense*, *Haemophilus parainfluenzae* and *Prevotella* sp. HMT396.

OLKs were then divided into LROLK (no or mild dysplasia [n=91]) and HROLK (moderate or severe dysplasia [n=125]) to investigate if differing degrees of dysplasia were associated with variations in microbial colonisation. The microbial communities of LROLK and HROLK were found to exhibit significant changes. LROLKs were consistently associated with significantly increased abundances of *Gemella morbillorum* and *Mogibacterium diversum*, while HROLKs were associated with significantly increased abundances of *Bergeyella* sp. HMT322, *Streptococcus australis infantis* clade 431 and *Leptotrichia* sp. HMT215. The sampled cohort also included 10 OSCCs which presented clinically as OLK. These OSCCs were associated with significantly increased abundances of *Parvimonas* sp. HMT393 and *Fusobacterium nucleatum* subsp. *polymorphum hwasookii*.

Finally, predictive modelling using the top 50 microbial features from a preliminary modelling analysis could differentiate HROLK from LROLK with a sensitivity of 85.5% and specificity of 76.5%, while a combined model using both clinical/patient and microbial features could predict LROLK or HROLK with a sensitivity of 87.4% and specificity of 76.5%. Remarkably these models were considerably more accurate than models using clinical/patient factors alone.

In conclusion, the microbiome of OLK was shown to differ significantly from that of clinically normal mucosa. OLKs with different degrees of dysplasia were also found to exhibit significant differences in microbial composition, and these differences could be used to predict whether an OLK was low or high-risk. Secondary findings included that, aside from OLK, smoking and oral site were the most significant influencers of the oral microbiome. Finally, OLK location on tongue/floor of mouth and non-homogeneity were associated with severe dysplasia.

# Acknowledgements

There are several people that I would like to thank for their help and support, without whom it would not have been possible to complete this PhD thesis. Firstly, I would like to express my sincere thanks to my supervisors Dr Gary Moran, Associate Professor in Microbiology, and Prof Claire Healy, Professor in Oral Medicine, for their unwavering support, invaluable advice and encouragement throughout the course of this PhD.

I would also like to thank Ms Sviatlana Anishchuk, research hygienist in the Dublin Dental University Hospital (DDUH), who contributed significantly to swab sampling and initial data collection, ensuring that the intended sample size was met.

A special thanks to Dr Bahman Honari, statistician in the Dublin Dental University Hospital, for his invaluable help with the statistical modelling detailed in Chapter 9.

I am also very grateful for the support provided by the Microbiology Research Unit, Division of Oral Biosciences, DDUH, and the funding from the Irish Health Research Board, grant number ILP-POR-2019-030.

A final thanks to my family who remind me daily to keep things in perspective.

## **Publications related to this thesis**

1. Galvin S, Moran GP, Healy CM. Influence of site and smoking on malignant transformation in the oral cavity: Is the microbiome the missing link? *Front Oral Health*. 2023 Mar 23;4:1166037. doi: 10.3389/froh.2023.1166037. PMID: 37035251; PMCID: PMC10076759 (**Appendix G Published Review**)
2. Galvin S, Anishchuk S, Healy CM, Moran GP. Smoking, tooth loss and oral hygiene practices have significant and site-specific impacts on the microbiome of oral mucosal surfaces: a cross-sectional study. *J Oral Microbiol*. 2023 Oct 2;15(1):2263971. doi: 10.1080/20002297.2023.2263971. PMID: 37795170; PMCID: PMC10547447 (**Appendix H Published paper based on the findings detailed in Chapter 6**)

# Abbreviations

|                    |                                                  |
|--------------------|--------------------------------------------------|
| <b>ACH</b>         | Acetaldehyde                                     |
| <b>ADH</b>         | Alcohol dehydrogenase                            |
| <b>AIC</b>         | Akaike information criterion                     |
| <b>Alv/Gin</b>     | Alveolus/Gingiva                                 |
| <b>ASV</b>         | Amplicon sequence variant                        |
| <b>AUC</b>         | Area under the curve                             |
| <b>BIC</b>         | Bayesian information criterion                   |
| <b>BM</b>          | Buccal mucosa                                    |
| <b>bp</b>          | base pairs                                       |
| <b>Buc/Alv/Gin</b> | Buccal mucosa/Alveolus/Gingiva                   |
| <b>CDC</b>         | Centers for Disease Control                      |
| <b>CHC</b>         | Chronic hyperplastic candidiasis                 |
| <b>CI</b>          | Confidence interval                              |
| <b>CLN</b>         | Contralateral normal                             |
| <b>CRC</b>         | Colorectal carcinoma                             |
| <b>DDUH</b>        | Dublin Dental University Hospital                |
| <b>DNA</b>         | Deoxyribonucleic acid                            |
| <b>DT</b>          | Dorsum tongue                                    |
| <b>eHOMD</b>       | expanded Human Oral Microbiome Database          |
| <b>EMT</b>         | Epithelial-mesenchymal transition                |
| <b>FEP</b>         | Fibroepithelial polyp                            |
| <b>FDR</b>         | False discovery rate                             |
| <b>FOM/VT</b>      | Floor of mouth/ventral tongue                    |
| <b>GM-CSF</b>      | Granulocyte-macrophage colony stimulating factor |
| <b>HC</b>          | Healthy control                                  |
| <b>HCC</b>         | Hepatocellular carcinoma                         |
| <b>HNSCC</b>       | Head and neck squamous cell carcinoma            |
| <b>HP</b>          | Hard palate                                      |
| <b>HROLK</b>       | High-risk OLK                                    |
| <b>IARC</b>        | International agency for cancer research         |
| <b>IBD</b>         | Inflammatory bowel disease                       |

|                  |                                                                                         |
|------------------|-----------------------------------------------------------------------------------------|
| <b>ITS2</b>      | Internal transcribed spacer 2 gene                                                      |
| <b>LBT</b>       | Lateral border tongue                                                                   |
| <b>LM</b>        | Linear model                                                                            |
| <b>LROLK</b>     | Low-risk OLK                                                                            |
| <b>ML</b>        | Machine learning                                                                        |
| <b>NGS</b>       | Next generation sequencing                                                              |
| <b>NHOLK</b>     | Non-homogenous OLK                                                                      |
| <b>NMDS</b>      | Non-metric multidimensional scaling                                                     |
| <b>NPV</b>       | Negative predictive value                                                               |
| <b>OE</b>        | Oral erythroplakia                                                                      |
| <b>OHI-S</b>     | Simplified oral hygiene index                                                           |
| <b>OLK</b>       | Oral leukoplakia                                                                        |
| <b>OPMD</b>      | Oral potentially malignant disorder                                                     |
| <b>OPSCC</b>     | Oropharyngeal squamous cell carcinoma                                                   |
| <b>OR</b>        | Odds ratio                                                                              |
| <b>OSF</b>       | Oral submucous fibrosis                                                                 |
| <b>OSCC</b>      | Oral squamous cell carcinoma                                                            |
| <b>OTU</b>       | Operational taxonomic unit                                                              |
| <b>OVH</b>       | Oral verrucous hyperplasia                                                              |
| <b>PAL</b>       | Palate                                                                                  |
| <b>PCA</b>       | Principal component analysis                                                            |
| <b>PCR</b>       | Polymerase chain reaction                                                               |
| <b>PERMANOVA</b> | Permutational Multivariate Analysis of Variance                                         |
| <b>PPV</b>       | Positive predictive value                                                               |
| <b>PR</b>        | Precision recall                                                                        |
| <b>PVL</b>       | Proliferative verrucous leukoplakia                                                     |
| <b>ROC</b>       | Receiver operator characteristic                                                        |
| <b>RR</b>        | Relative risk                                                                           |
| <b>rRNA</b>      | Ribosomal ribonucleic acid                                                              |
| <b>RRR</b>       | Relative risk ratio                                                                     |
| <b>SD</b>        | Standard deviation                                                                      |
| <b>SIAMCAT</b>   | Statistical Inference of Associations between Microbial Communities And host phenoTypes |



|                     |                             |
|---------------------|-----------------------------|
| <b>SP</b>           | Soft palate                 |
| <b>sp. and spp.</b> | Species singular and plural |
| <b>TSE</b>          | Tris/sodium/EDTA            |
| <b>VIF</b>          | Variance inflation factor   |
| <b>WHO</b>          | World Health Organisation   |

# Table of contents

|                                                                |            |
|----------------------------------------------------------------|------------|
| <b>Declaration.....</b>                                        | <b>ii</b>  |
| <b>Summary.....</b>                                            | <b>iii</b> |
| <b>Acknowledgements.....</b>                                   | <b>v</b>   |
| <b>Publications related to this thesis .....</b>               | <b>vi</b>  |
| <b>Abbreviations .....</b>                                     | <b>vii</b> |
| <b>Table of contents .....</b>                                 | <b>x</b>   |
| <b>Table of figures.....</b>                                   | <b>xv</b>  |
| <b>Table of tables.....</b>                                    | <b>xix</b> |
| <b>1 Introduction .....</b>                                    | <b>1</b>   |
| <b>2 Literature Review .....</b>                               | <b>3</b>   |
| 2.1 Role of microorganisms in cancer.....                      | 3          |
| 2.2 Risk factors for OSCC.....                                 | 5          |
| 2.3 Effect of smoking and alcohol on the oral microbiome ..... | 7          |
| 2.3.1 Smoking swab studies.....                                | 8          |
| 2.3.2 Smoking saliva studies.....                              | 9          |
| 2.3.3 Smoking plaque studies .....                             | 10         |
| 2.3.4 Alcohol studies.....                                     | 14         |
| 2.4 The microbiome in OSCC.....                                | 15         |
| 2.4.1 OSCC swab studies.....                                   | 15         |
| 2.4.2 OSCC tissue studies.....                                 | 17         |
| 2.4.3 OSCC saliva studies.....                                 | 20         |
| 2.5 Oral potentially malignant disorders .....                 | 32         |
| 2.6 The microbiome in OPMDs .....                              | 36         |
| 2.6.1 OPMD swab studies.....                                   | 39         |
| 2.6.2 OPMD tissue studies.....                                 | 41         |
| 2.6.3 OPMD saliva studies.....                                 | 43         |
| 2.6.4 Systematic review .....                                  | 45         |
| 2.7 Models of microbial carcinogenesis.....                    | 46         |
| 2.7.1 Epithelial change.....                                   | 46         |
| 2.7.2 Acetaldehyde.....                                        | 47         |
| 2.7.3 Inflammation.....                                        | 49         |

|          |                                                                                                                        |           |
|----------|------------------------------------------------------------------------------------------------------------------------|-----------|
| 2.8      | The role of <i>Candida</i> .....                                                                                       | 50        |
| 2.9      | Conclusion .....                                                                                                       | 54        |
| <b>3</b> | <b>Aims and Objectives .....</b>                                                                                       | <b>56</b> |
| <b>4</b> | <b>Materials and Methods.....</b>                                                                                      | <b>57</b> |
| 4.1      | Sample collection.....                                                                                                 | 57        |
| 4.2      | DNA extraction.....                                                                                                    | 61        |
| 4.3      | DNA sequencing.....                                                                                                    | 62        |
| 4.4      | Data analysis .....                                                                                                    | 63        |
| <b>5</b> | <b>Demographics of patient cohort and clinical characteristics and associations of OLK .....</b>                       | <b>67</b> |
| 5.1      | Introduction.....                                                                                                      | 67        |
| 5.2      | Methods .....                                                                                                          | 67        |
| 5.3      | Results.....                                                                                                           | 71        |
| 5.3.1    | Smoking and alcohol consumption.....                                                                                   | 71        |
| 5.3.2    | Disease sites.....                                                                                                     | 77        |
| 5.3.3    | Oral hygiene .....                                                                                                     | 78        |
| 5.3.4    | Degree of dysplasia .....                                                                                              | 78        |
| 5.3.5    | Clinical appearance.....                                                                                               | 87        |
| 5.4      | Follow up .....                                                                                                        | 88        |
| 5.5      | Discussion.....                                                                                                        | 89        |
| 5.6      | Conclusion .....                                                                                                       | 93        |
| <b>6</b> | <b>The microbiome of healthy mucosal surfaces and the influence of various factors on microbial colonisation .....</b> | <b>94</b> |
| 6.1      | Introduction.....                                                                                                      | 94        |
| 6.2      | Methods .....                                                                                                          | 94        |
| 6.3      | Results.....                                                                                                           | 95        |
| 6.3.1    | Smoking.....                                                                                                           | 97        |
| 6.3.2    | Alcohol consumption.....                                                                                               | 104       |
| 6.3.3    | Tooth brushing.....                                                                                                    | 106       |
| 6.3.4    | Oral hygiene .....                                                                                                     | 109       |
| 6.3.5    | Denture wear.....                                                                                                      | 111       |
| 6.3.6    | Mouthwash use .....                                                                                                    | 115       |
| 6.3.7    | Missing teeth .....                                                                                                    | 116       |
| 6.3.8    | Site.....                                                                                                              | 120       |
| 6.4      | Multivariate model analysis.....                                                                                       | 131       |

|          |                                                                                            |            |
|----------|--------------------------------------------------------------------------------------------|------------|
| 6.4.1    | Smoking, alcohol consumption and multiple measures of oral hygiene .....                   | 131        |
| 6.4.2    | Age and Gender .....                                                                       | 133        |
| 6.4.3    | Smoking and site.....                                                                      | 136        |
| 6.5      | Discussion: .....                                                                          | 142        |
| 6.5.1    | Influence of plaque, oral hygiene measures and tooth number on the mucosal microbiome..... | 142        |
| 6.5.2    | Influence of age and gender on the mucosal microbiome .....                                | 146        |
| 6.5.3    | Influence of mucosal site on the mucosal microbiome.....                                   | 147        |
| 6.5.4    | Influence of smoking on the mucosal microbiome.....                                        | 151        |
| 6.6      | Conclusion.....                                                                            | 154        |
| <b>7</b> | <b>The microbiome of OLK and the identification of bacterial taxa enriched on OLK.....</b> | <b>158</b> |
| 7.1      | Introduction .....                                                                         | 158        |
| 7.2      | Methods.....                                                                               | 158        |
| 7.3      | Results .....                                                                              | 160        |
| 7.3.1    | Comparison of the microbiome of OLK with CLN and HC sites .....                            | 161        |
| 7.3.2    | Microbiome associated with OLK adjusted for smoking .....                                  | 168        |
| 7.3.3    | Microbiome associated with OLK adjusted for alcohol .....                                  | 172        |
| 7.3.4    | Microbiome associated with OLK adjusted for brushing habit .....                           | 173        |
| 7.3.5    | Microbiome associated with OLK adjusted for oral hygiene .....                             | 175        |
| 7.3.6    | Microbiome associated with OLK adjusted for denture wear .....                             | 177        |
| 7.3.7    | Microbiome associated with OLK adjusted for mouthwash use .....                            | 179        |
| 7.3.8    | Microbiome associated with OLK and missing teeth .....                                     | 179        |
| 7.3.9    | Microbiome associated with OLK adjusted for mucosal site .....                             | 180        |
| 7.3.10   | Microbiome associated with OLK adjusted for age and gender.....                            | 183        |
| 7.3.11   | Microbiome associated with OLK adjusted for clinical appearance .....                      | 184        |
| 7.3.12   | Microbiome associated with OLK adjusted for multiple measures of oral hygiene              | 184        |
| 7.3.13   | Microbiome associated with OLK adjusted for site and smoking .....                         | 185        |
| 7.4      | Discussion .....                                                                           | 186        |
| 7.5      | Conclusion.....                                                                            | 192        |
| <b>8</b> | <b>The microbiome of low-risk and high-risk dysplastic OLKs and OSCC.....</b>              | <b>193</b> |
| 8.1      | Introduction .....                                                                         | 193        |
| 8.2      | Methods.....                                                                               | 193        |

|          |                                                                                                                                      |            |
|----------|--------------------------------------------------------------------------------------------------------------------------------------|------------|
| 8.3      | Results.....                                                                                                                         | 196        |
| 8.3.1    | Microbiome of low-risk and high-risk dysplasia and OSCC .....                                                                        | 196        |
| 8.3.2    | Microbiome associated with LROLK, HROLK, OSCC adjusted for smoking<br>207                                                            |            |
| 8.3.3    | Microbiome associated with LROLK, HROLK, OSCC adjusted for alcohol<br>211                                                            |            |
| 8.3.4    | Microbiome associated with LROLK, HROLK, OSCC adjusted for brushing<br>habit 212                                                     |            |
| 8.3.5    | Microbiome associated with LROLK, HROLK, OSCC adjusted for oral<br>hygiene 213                                                       |            |
| 8.3.6    | Microbiome associated with LROLK, HROLK, OSCC adjusted for denture<br>wear 214                                                       |            |
| 8.3.7    | Microbiome associated with LROLK, HROLK, OSCC adjusted for<br>mouthwash use ..... 215                                                |            |
| 8.3.8    | Microbiome associated with LROLK, HROLK, OSCC adjusted for missing<br>teeth 215                                                      |            |
| 8.3.9    | Microbiome associated with LROLK, HROLK, OSCC adjusted for mucosal<br>site 217                                                       |            |
| 8.3.10   | Microbiome associated with LROLK, HROLK, OSCC adjusted for clinical<br>appearance of OLK ..... 220                                   |            |
| 8.3.11   | Microbiome associated with LROLK, HROLK, OSCC adjusted for multiple<br>measures of oral hygiene..... 221                             |            |
| 8.3.12   | Microbiome associated with LROLK, HROLK, OSCC adjusted for site and<br>smoking model..... 222                                        |            |
| 8.4      | Microbiome of dysplastic OLK graded by the 3-tier WHO classification .....                                                           | 224        |
| 8.4.1    | Microbiome associated with dysplasia graded by the 3-tier WHO<br>classification adjusted for multiple measures of oral hygiene ..... | 224        |
| 8.4.2    | Microbiome associated with dysplasia graded by the WHO 3-tier<br>classification adjusted for site and smoking..... 227               |            |
| 8.5      | Discussion.....                                                                                                                      | 228        |
| 8.6      | Conclusion .....                                                                                                                     | 233        |
| <b>9</b> | <b>Predictive modelling .....</b>                                                                                                    | <b>234</b> |
| 9.1      | Introduction.....                                                                                                                    | 234        |
| 9.2      | Methods .....                                                                                                                        | 235        |
| 9.3      | Results.....                                                                                                                         | 237        |
| 9.3.1    | Modelling with clinical and patient factors .....                                                                                    | 237        |
| 9.3.2    | Initial modelling with microbiome data .....                                                                                         | 240        |
| 9.3.3    | Further modelling with microbiome data .....                                                                                         | 245        |

|           |                                                                                                             |            |
|-----------|-------------------------------------------------------------------------------------------------------------|------------|
| 9.3.4     | Modelling with combined data .....                                                                          | 248        |
| 9.4       | Discussion .....                                                                                            | 251        |
| 9.5       | Conclusion.....                                                                                             | 255        |
| <b>10</b> | <b>Summary and Conclusions .....</b>                                                                        | <b>256</b> |
| 10.1      | Influence of clinical and patient-related factors on the degree of dysplasia .....                          | 256        |
| 10.2      | Microbiome of clinically normal mucosa and the influence of patient factors on microbial colonisation ..... | 258        |
| 10.3      | The microbiome of OLK.....                                                                                  | 261        |
| 10.4      | The microbiome of OLKs of different degrees of dysplasia and OSCC .....                                     | 264        |
| 10.5      | Predictive modelling .....                                                                                  | 265        |
| 10.6      | The future .....                                                                                            | 266        |
| 10.7      | Limitations .....                                                                                           | 267        |
| 10.8      | Conclusion.....                                                                                             | 268        |
| <b>11</b> | <b>Bibliography.....</b>                                                                                    | <b>270</b> |
| <b>12</b> | <b>Appendix A Full results from Table 36.....</b>                                                           | <b>288</b> |
| <b>13</b> | <b>Appendix B Full results from Table 39 .....</b>                                                          | <b>291</b> |
| <b>14</b> | <b>Appendix C Patient information leaflet .....</b>                                                         | <b>296</b> |
| <b>15</b> | <b>Appendix D Patient consent form.....</b>                                                                 | <b>298</b> |
| <b>16</b> | <b>Appendix E Data recording form .....</b>                                                                 | <b>299</b> |
| <b>17</b> | <b>Appendix F OHI-S.....</b>                                                                                | <b>300</b> |
| <b>18</b> | <b>Appendix G Published Review .....</b>                                                                    | <b>301</b> |
| <b>19</b> | <b>Appendix H Published paper based on the findings detailed in Chapter 6.....</b>                          | <b>302</b> |

# Table of figures

|                                                                                                                                                                                                                                                                                                                                            |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Figure 1.</b> OSCC on the right lateral border of the tongue consisting of a 1cm ulcer with raised irregular margins. There is a non-homogenous leukoplakia distal to the ulcer. Photograph courtesy of Dr. S. Galvin.....                                                                                                              | 5   |
| <b>Figure 2.</b> Homogenous leukoplakia anterior hard palate extending to the incisive papilla. Photograph courtesy of Dr. S. Galvin. ....                                                                                                                                                                                                 | 33  |
| <b>Figure 3.</b> Proliferative verrucous leukoplakia right and left ventral tongue. Photograph courtesy of Dr. S. Galvin. ....                                                                                                                                                                                                             | 34  |
| <b>Figure 4.</b> Example of H&E stains of mild dysplasia at 10x and 100x magnification. Photos courtesy of Dr Róisín O'Connor, St. James's Hospital, Dublin.....                                                                                                                                                                           | 59  |
| <b>Figure 5.</b> Examples of H&E stains of moderate dysplasia at 20x magnification. Photos courtesy of Dr Esther O'Regan, St. James's Hospital, Dublin. ....                                                                                                                                                                               | 59  |
| <b>Figure 6.</b> Example of H&E stains of severe dysplasia at 10x and 100x magnification. Photos courtesy of Dr Róisín O'Connor, St. James's Hospital, Dublin.....                                                                                                                                                                         | 60  |
| <b>Figure 7.</b> Graphic of the 16S rRNA gene with variable regions highlighted (Image @ Pacific Biosciences) .....                                                                                                                                                                                                                        | 63  |
| <b>Figure 8.</b> Flowchart of patient cohort.....                                                                                                                                                                                                                                                                                          | 68  |
| <b>Figure 9:</b> Selection of included homogenous OLKs. Photographs courtesy of Dr. S Galvin .....                                                                                                                                                                                                                                         | 69  |
| <b>Figure 10.</b> Selection of included non-homogenous (including ulcerated) OLKs. Photos courtest of Dr. S Galvin. ....                                                                                                                                                                                                                   | 70  |
| <b>Figure 11.</b> Location of OLKs with severe dysplasia. LBT=lateral border tongue, PAL=palate, Alv/Gin=Alveolus/Gingiva, BM=Buccal mucosa, DT=Dorsum tongue, FOMVT=Floor of mouth/ventral tongue .....                                                                                                                                   | 82  |
| <b>Figure 12.</b> Non-homogenous leukoplakia right lateral border of the tongue. Photograph courtesy Dr. S. Galvin. ....                                                                                                                                                                                                                   | 87  |
| <b>Figure 13.</b> Flowchart of sample cohort of contralateral normal and healthy control sites .....                                                                                                                                                                                                                                       | 95  |
| <b>Figure 14.</b> Bray Curtis NMDS plot of smoking status on axes 1 and 2 (padj=0.001).....                                                                                                                                                                                                                                                | 99  |
| <b>Figure 15.</b> Bray Curtis NMDS plot of smoking status on axes 2 and 3 (padj=0.001).....                                                                                                                                                                                                                                                | 100 |
| <b>Figure 16.</b> Heatmap showing taxa with significant changes in abundance associated with smoking with effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ . Colour intensity relates to effect size where increased abundance is indicated in red and reduced abundance in blue (compared to never smokers). .... | 104 |
| <b>Figure 17.</b> Bray Curtis NMDS plot of categories of alcohol consumption (Alcohol bin=units/week) (p=0.38).....                                                                                                                                                                                                                        | 105 |
| <b>Figure 18.</b> Alpha diversity plots by daily tooth brushing, p values denoted by numbers above square brackets. (<1=Brushing less than once per day, 1=Brushing once per day, >1=Brushing more than once per day) .....                                                                                                                | 106 |
| <b>Figure 19.</b> Bray-Curtis NMDS plot of brushing habits: <1=less than once per day, 1=once per day, >1=more than once per day (p=0.03) .....                                                                                                                                                                                            | 107 |
| <b>Figure 20.</b> Bray-Curtis NMDS plot of oral hygiene category (Good=oral hygiene index score (OHI-S)<0.5, Fair=OHI-S 0.5-1, Poor=OHI-S>1) (p=0.3).....                                                                                                                                                                                  | 109 |
| <b>Figure 21.</b> Heatmap showing taxa with significant changes in abundance associated with poor (OHI_Cat Poor) and fair (OHI_Cat Fair) oral hygiene. (Effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ . Red=increased, blue=reduced relative to good oral hygiene (OHI-S<0.5)) .....                            | 111 |
| <b>Figure 22.</b> Bray-Curtis NMDS plot of denture wear (p=0.007) .....                                                                                                                                                                                                                                                                    | 112 |
| <b>Figure 23.</b> Heatmap showing taxa with significant changes in abundance associated with denture wear. (Effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ . Red=increased, blue=reduced relative to no denture wear).....                                                                                       | 114 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Figure 24.</b> Bray-Curtis NMDS plot of mouthwash use (p=0.03).....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 115 |
| <b>Figure 25.</b> Bray-Curtis NMDS plot of population structure by number of missing teeth (Missing bin=number of missing teeth) (p=0.02) .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 117 |
| <b>Figure 26.</b> Alpha diversity plots by site using ACE and Chao1 indices, p value denoted by number above square brackets. (Buc=Buccal mucosa/alveolus/gingiva, FOM/VT=Floor of mouth/ventral tongue, PAL=Palate, Tongue=DT/LBT) .....                                                                                                                                                                                                                                                                                                                                                                                 | 121 |
| <b>Figure 27.</b> Alpha diversity plots by site using Observe, Shannon and Simpson indices, p value denoted by number above square brackets. (Buc=Buccal mucosa/alveolus/gingiva, FOM/VT=Floor of mouth/ventral tongue, PAL=Palate, Tongue=DT/LBT).....                                                                                                                                                                                                                                                                                                                                                                   | 122 |
| <b>Figure 28.</b> Axis 1 and 2 PCA plot by site showing separation of the microbial communities of different mucosal sites. The labelled arrows indicate the influence of specific genera in separating samples along the respective axis. DT/LBT (dorsum tongue/lateral border tongue) sites are clearly dominated by <i>Rothia</i> spp., FOM/VT (floor of mouth/ventral tongue) and Buc/Alv/Gin (buccal mucosa/alveolus/gingiva) sites appear to have similar colonisation patterns dominated by streptococci, while the separation of PAL (palatal) sites is influenced by the presence of <i>Prevotella</i> spp. .... | 124 |
| <b>Figure 29.</b> Axis 2 and 3 PCA plot by site showing separation of the microbial communities of different mucosal sites. The labelled arrows indicate the influence of specific genera in separating samples along the respective axis. DT/LBT (dorsum tongue/lateral border tongue) sites are clearly dominated by <i>Rothia</i> and <i>Neisseria</i> spp., FOM/VT (floor of mouth/ventral tongue) and Buc/Alv/Gin (buccal mucosa/alveolus/gingiva) sites have similar colonisation patterns, while the separation of PAL (palatal) sites is due in part to the presence of <i>Prevotella</i> spp. ....               | 125 |
| <b>Figure 30.</b> Heatmap showing taxa with changes in abundance associated with mucosal site on MaAsLin2 analysis (Effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ ). Red=increased, blue=reduced relative to Buc/Alv/Gin).....                                                                                                                                                                                                                                                                                                                                                                 | 130 |
| <b>Figure 31.</b> Heatmap of MaAsLin2 analysis of a multivariate model of smoking, alcohol, oral hygiene, tooth number, mouthwash use, denture wear and brushing showing taxa with significant changes in abundance associated with these variables. (Effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ ). Red=increased, blue=decreased. Never smoker, no alcohol, OHI-S<0.5, missing < 5 teeth, no denture wear and brushing >1/day used as references) (Brushing_2=brushing once or less/day) .....                                                                                             | 133 |
| <b>Figure 32.</b> Heatmap of MaAsLin2 analysis of a multivariate model of age and gender adjusted for smoking, tooth number and brushing habit showing taxa with significant changes in abundance associated with these variables. (Effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ ). (Red=increased, blue=decreased. Never smoker, female gender, age<40, brushing >1/day and missing <5 teeth used as references).....                                                                                                                                                                        | 135 |
| <b>Figure 33.</b> MaAsLin2 generated heatmap showing taxa with changes in abundance associated with smoking and oral site. (Effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ ). Buc/Alv/Gin and never smokers used as references. Red-increased, blue=reduced).....                                                                                                                                                                                                                                                                                                                               | 142 |
| <b>Figure 34.</b> Venn diagram showing overlap of the genera associated with smoking, OSCC and OLK from previous studies.....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 157 |
| <b>Figure 35.</b> Flowchart of sample cohort .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 159 |
| <b>Figure 36.</b> Alpha diversity plots of disease, contralateral normal and healthy controls using ACE, Chao1, Observe, Shannon and Simpson indices. Y axis shows index values, p values shown above square brackets and also in Table 44 below.....                                                                                                                                                                                                                                                                                                                                                                     | 162 |
| <b>Figure 37.</b> Axis 1 and 2 PCA plot of disease (OLK) and contralateral normal (CLN) sites showing separation of the microbial communities of OLK and CLN sites and the genera responsible for these population shifts. Contralateral sites show increased abundances of by streptococci while disease sites are characterised by increased abundances of <i>Rothia</i> , <i>Prevotella</i> , <i>Fusobacterium</i> , <i>Porphyromonas</i> , <i>Leptotrichia</i> and <i>Gemella</i> . ....                                                                                                                              | 164 |



|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Figure 38.</b> Barplot showing the most abundant phyla associated with CLN (Contralateral) and OLK (Disease) .....                                                                                                                                                                                                                                                                                                                                                                                         | 165 |
| <b>Figure 39.</b> Bar plot showing the most abundant genera associated with CLN and OLK sites.....                                                                                                                                                                                                                                                                                                                                                                                                            | 166 |
| <b>Figure 40.</b> Heatmap showing taxa with significant changes in abundance associated with OLK and smoking with effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ . (D.Control_exSCC=OLK excluding OSCC samples [n=10]. Red=increased, blue=decreased) .....                                                                                                                                                                                                                         | 171 |
| <b>Figure 41.</b> Heatmap showing taxa with significant changes in abundance associated with OLK and alcohol consumption. Effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ . (D.Control_exSCC=OLK excluding OSCC samples [n=10]. Alcohol_bin 1 to 10 = 1-10 units/week, 11-20 = 11 to 20 units/week, more than 20 = >20 units/week. Red=increased, blue=decreased).....                                                                                                               | 173 |
| <b>Figure 42.</b> Heatmap showing taxa with significant changes in abundance associated with OLK and brushing habit. Effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ . (D.Control_exSCC=OLK excluding OSCC samples [n=10], Brushing_2 = Brushing once or less than once per day. Red=increased, blue=decreased).....                                                                                                                                                                 | 175 |
| <b>Figure 43.</b> Heatmap showing taxa with significant changes in abundance associated with OLK and oral hygiene category. (Effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ , Red=increased, blue=decreased. D.Control_exSCC=OLK excluding OSCC samples [n=10], OHI_Cat=oral hygiene category) .....                                                                                                                                                                                | 177 |
| <b>Figure 44.</b> Heatmap showing the top 50 taxa with significant changes in abundance associated with OLK and brushing habit, oral hygiene category, denture wear, mouthwash use and number of missing teeth. Effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ . (D.Control_exSCC=OLK excluding OSCC samples [n=10], M.W=mouthwash use, Missing_bin=number of missing teeth, Brushing_2=brushing once or less/day, OHI_Cat=oral hygiene category. Red=increased, blue=reduced)..... | 185 |
| <b>Figure 45.</b> pT1 OSCC left floor of mouth .....                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 194 |
| <b>Figure 46.</b> pT2 OSCC left lateral tongue.....                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 195 |
| <b>Figure 47.</b> Alpha diversity plots of contralateral normal, none/mild dysplasia (LROLK), moderate/severe dysplasia (mod/sev [HROLK]) and OSCC (SCC) using ACE, Chao1, Observe, Shannon and Simpson indices. p values shown above square brackets .....                                                                                                                                                                                                                                                   | 197 |
| <b>Figure 48.</b> Axis 1 and 2 NMDS plot of LROLK, HROLK, and OSCC sites showing separation of the communities of LROLK, HROLK and OSCC.....                                                                                                                                                                                                                                                                                                                                                                  | 198 |
| <b>Figure 49.</b> Axis 2 and 3 NMDS plot of LROLK, HROLK and OSCC sites showing separation of the communities of LROLK, HROLK and OSCC.....                                                                                                                                                                                                                                                                                                                                                                   | 199 |
| <b>Figure 50.</b> Barplot showing the most abundant phyla associated with CLN, LROLK (Mild/None), HROLK (Mod/Sev) and OSCC (SCC). .....                                                                                                                                                                                                                                                                                                                                                                       | 200 |
| <b>Figure 51.</b> Barplot showing the most abundant genera associated with CLN, LROLK (Mild/None), HROLK (Mod/Sev) and OSCC (SCC). .....                                                                                                                                                                                                                                                                                                                                                                      | 201 |
| <b>Figure 52.</b> Heatmap showing taxa with significant changes in abundance associated with LROLK (Dysplasia_2 Mild/None), HROLK (Dysplasia_2 Mod/Sev), OSCC (SCC) and smoking status. Effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ (red=increased, blue=reduced). .....                                                                                                                                                                                                         | 210 |
| <b>Figure 53.</b> Heatmap showing the top 50 taxa with significant changes in abundance associated with no, mild and severely dysplastic OLK and multiple measures of oral hygiene. (Effect size calculated as $-\log(q\text{-value}) * (\text{coefficient})$ . Red=increased, blue=reduced. Missing_bin=number of missing teeth, Brushing_2=brushing once or less/day, M.W=mouthwash use, OHI_Cat=oral hygiene category) .....                                                                               | 226 |
| <b>Figure 54.</b> Receiver operating characteristic curve for RF modelling using clinical and patient factors (AUC=78.5%) .....                                                                                                                                                                                                                                                                                                                                                                               | 238 |
| <b>Figure 55.</b> Graphic of a Classification tree from RF modelling of clinical and patient factors. (Clin_app = clinical appearance, 1 = homogenous, 2 = non-homogenous; Age_cat = age category,                                                                                                                                                                                                                                                                                                            |     |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                             |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 1 = <40 years, 2 = 40-60 years, 3 = >60 years, Brushing: 1 = More than once/day, 2 = once or less/day, Site 1 = Buc/Alv/Gin, 2 = FOM/VT, 3 = DT/LBT, 4 = PAL, MW = mouthwash use, 1 = Yes, 2 = No).....                                                                                                                                                                                                                                                     | 239 |
| <b>Figure 56.</b> Plot showing the top 50 features, i.e. including bacterial taxa and clinical/patient features, from the best random forest model (Model 1: microbiome, clinical appearance, age, site, gender and alcohol consumption). Mild/None = mild/no dysplasia (LROLK). Mod/Sev = moderate/severe dysplasia (HROLK). Brown = increased, Green = decreased where colour intensity relates to the z-score (standard deviations from the mean). ..... | 242 |
| <b>Figure 57.</b> Effect size and relative feature weight of each feature in <b>Figure 56</b> .....                                                                                                                                                                                                                                                                                                                                                         | 243 |
| <b>Figure 58.</b> Plot of the top 50 features using LASSO modelling of the same model (microbiome, clinical appearance, age, site, gender and alcohol consumption) as <b>Figure 56</b> . Mild/None = mild/no dysplasia (LROLK). Mod/Sev = moderate/severe dysplasia (HROLK). Brown=increased, green=decreased, colour intensity relates to z-score (standard deviations from the mean). .....                                                               | 244 |
| <b>Figure 59.</b> Effect size and relative feature weight of each feature in <b>Figure 58</b> .....                                                                                                                                                                                                                                                                                                                                                         | 245 |
| <b>Figure 60.</b> Receiver operating characteristic curve for RF modelling using the z-scores of the top 50 microbial features from initial SIAMCAT analysis (AUC=78.8%). .....                                                                                                                                                                                                                                                                             | 246 |
| <b>Figure 61.</b> Graphic of a Classification tree from RF modelling using the z-scores of the top 50 microbial features from initial SIAMCAT analysis.....                                                                                                                                                                                                                                                                                                 | 247 |
| <b>Figure 62.</b> Graph of the Gini index, a measure of variable impurity, for the microbial z-score RF analysis. The higher the score on the index the greater the discriminatory value of the variable. ....                                                                                                                                                                                                                                              | 248 |
| <b>Figure 63.</b> ROC curve for RF modelling using the clinical/patient and microbial (z-scores) features (AUC=88%) .....                                                                                                                                                                                                                                                                                                                                   | 249 |
| <b>Figure 64.</b> Graphic of a Classification tree from RF modelling of clinical/patient factors and microbial z-scores. (Clin_app = clinical appearance, 1 = homogenous, 2 = non-homogenous) ..                                                                                                                                                                                                                                                            | 250 |
| <b>Figure 65.</b> Graph of the Gini index of the RF analysis of clinical/patient and microbial features. The higher the score on the index the greater the discriminatory value of the variable. ....                                                                                                                                                                                                                                                       | 251 |

## Table of tables

|                                                                                                                                                                                                                                                                                                                                                                                                                                                 |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Table 1.</b> Significant differences between the microbiomes of smokers and non-smokers from previous studies.....                                                                                                                                                                                                                                                                                                                           | 11 |
| <b>Table 2.</b> Significant differences between the microbiomes of OSCC and normal mucosa from previous studies.....                                                                                                                                                                                                                                                                                                                            | 28 |
| <b>Table 3.</b> The 11 oral potentially malignant disorders (OPMDs) .....                                                                                                                                                                                                                                                                                                                                                                       | 32 |
| <b>Table 4.</b> Significant differences between the microbiomes of OLK/OPMD and normal mucosa from previous studies (FFPE=formalin-fixed paraffin embedded).....                                                                                                                                                                                                                                                                                | 37 |
| <b>Table 5.</b> Demographic data of patient cohort .....                                                                                                                                                                                                                                                                                                                                                                                        | 72 |
| <b>Table 6.</b> Association of age with smoking habit.....                                                                                                                                                                                                                                                                                                                                                                                      | 73 |
| <b>Table 7.</b> Association of smoking status and site of OLK (% in brackets)<br>(Alv/Gin=alveolus/gingiva, BM=buccal mucosa, DT=dorsum tongue, FOM/VT=floor of mouth/ventral tongue, HP=hard palate, LBT=lateral border tongue, SP=soft palate) .....                                                                                                                                                                                          | 74 |
| <b>Table 8.</b> Results of multinomial regression investigating the association of smoking with site of OLK. Reference is never smokers and BM (buccal mucosa). Significant results in bold. CIs in brackets. (*DT combined with LBT due to small number (n=8) of DT OLKs, HP combined with SP due to small number (n=9) of HP OLKs). The adjusted model was adjusted for age, gender, oral hygiene, alcohol consumption and missing teeth..... | 74 |
| <b>Table 9.</b> Association of alcohol consumption with gender.....                                                                                                                                                                                                                                                                                                                                                                             | 75 |
| <b>Table 10.</b> Logistic regression analysis showing the odds of higher alcohol consumption in males. The reference category was female gender and 0 units of alcohol per week. Significant results in bold. The adjusted model was adjusted for smoking habit, age, oral hygiene, missing teeth and degree of dysplasia of OLK. ....                                                                                                          | 75 |
| <b>Table 11.</b> Association of smoking and alcohol consumption .....                                                                                                                                                                                                                                                                                                                                                                           | 76 |
| <b>Table 12.</b> Results of multinomial regression investigating the association of smoking with alcohol consumption. The reference category is never smokers and no alcohol. Significant results in bold. CIs in brackets.....                                                                                                                                                                                                                 | 76 |
| <b>Table 13.</b> Disease sites, histological findings and clinical appearance of OLK .....                                                                                                                                                                                                                                                                                                                                                      | 77 |
| <b>Table 14.</b> Distribution of clinical parameters by three-tier grade of dysplasia *Fisher's exact test with simulated p value for 2000 replicates (n= in brackets, significant results in bold) .....                                                                                                                                                                                                                                       | 79 |
| <b>Table 15.</b> Distribution of clinical parameters by binary category of dysplasia *Fisher's exact test with simulated p value for 2000 replicates (n= in brackets, significant results in bold) .....                                                                                                                                                                                                                                        | 81 |
| <b>Table 16.</b> Regression analysis showing odds of high-risk dysplasia by site (no/mild dysplasia and buccal mucosa used as reference category). Significant results in bold. (BM=Buccal mucosa, Alv/Gin=Alveolus/Gingiva, DT/LBT=Dorsum/Lateral tongue, FOM/VT=Floor of mouth/Ventral tongue, HP=Hard palate, SP=soft palate).....                                                                                                           | 83 |
| <b>Table 17.</b> Multinomial regression investigating the effect of site on the degree of dysplasia. The reference category is no dysplasia and buccal mucosa. Significant results in bold. CIs in brackets. ....                                                                                                                                                                                                                               | 84 |
| <b>Table 18.</b> Multinomial regression investigating the effect of site on the degree of dysplasia with dorsum, lateral and ventral tongue and floor of mouth (FOM) sites combined as FOM/tongue. The reference category is no dysplasia and buccal mucosa. Significant results in bold. CIs in brackets. ....                                                                                                                                 | 84 |
| <b>Table 19.</b> Regression analysis showing odds of high-risk dysplasia by clinical appearance .....                                                                                                                                                                                                                                                                                                                                           | 85 |
| <b>Table 20.</b> Results of multinomial regression showing association of age with the degree of dysplasia. The reference category is no dysplasia and age <40. Significant results in bold. CIs in brackets. ....                                                                                                                                                                                                                              | 86 |

|                                                                                                                                                                                                                                                                                                                           |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Table 21.</b> Logistic regression analysis showing odds of NHOLK at various sites (significant results in bold).....                                                                                                                                                                                                   | 88  |
| <b>Table 22.</b> Demographics of control and contralateral normal (patient) cohorts. Significant differences in demographics highlighted in bold. ....                                                                                                                                                                    | 96  |
| <b>Table 23.</b> Sites sampled (percentages in brackets).....                                                                                                                                                                                                                                                             | 97  |
| <b>Table 24.</b> Pairwise PERMANOVA results for smoking status.....                                                                                                                                                                                                                                                       | 100 |
| <b>Table 25.</b> Taxa with significant changes in abundance associated with smoking on MaAsLin2 analysis (negative coefficient values denote reduced abundance). ....                                                                                                                                                     | 101 |
| <b>Table 27.</b> Pairwise PERMANOVA results for toothbrushing (* $p < 0.05$ ).....                                                                                                                                                                                                                                        | 107 |
| <b>Table 28.</b> Taxa significantly associated with brushing once or less per day on MaAsLin2 analysis (negative values denote reduced abundance).....                                                                                                                                                                    | 108 |
| <b>Table 29.</b> Results of pairwise PERMANOVA for oral hygiene categories .....                                                                                                                                                                                                                                          | 109 |
| <b>Table 30.</b> Taxa with significant ( $p < 0.05$ ) changes in abundance associated with oral hygiene on MaAsLin2 analysis (negative values denote reduced abundance).....                                                                                                                                              | 110 |
| <b>Table 31.</b> Pairwise PERMANOVA results for denture wear (* $p < 0.05$ , $p_{adj} < 0.1$ ) .....                                                                                                                                                                                                                      | 113 |
| <b>Table 32.</b> Results of MaAsLin2 analysis showing taxa with significant changes in abundance associated with denture wear (negative values denote reduced abundance) (*= $p_{adj} < 0.1$ ).....                                                                                                                       | 113 |
| <b>Table 33.</b> Table showing p values from comparisons of alpha diversities of communities missing 5, 5 to 15 and more than 15 teeth. Significant results in bold.....                                                                                                                                                  | 116 |
| <b>Table 34.</b> Pairwise PERMANOVA results for missing teeth.....                                                                                                                                                                                                                                                        | 117 |
| <b>Table 35.</b> Results of MaAsLin2 analysis showing taxa with significant changes in abundance associated with number of missing teeth (negative coefficient values denote reduced abundance). ....                                                                                                                     | 119 |
| <b>Table 36.</b> Pairwise PERMANOVA results for site comparisons (DT/LBT=Dorsum tongue/lateral border tongue, Buc/Alv/Gin=Buccal mucosa/Alveolus/Gingiva, PAL=Palate, FOM/VT=Floor of mouth/ventral tongue).....                                                                                                          | 126 |
| <b>Table 37.</b> Results of MaAsLin2 analysis showing taxa with significant changes in abundance associated with site. Only results with $p_{adj} < 0.05$ shown due to volume of results. (Buc/Alv/Gin used as reference) (negative values denote reduced abundance). Full results available in <b>Appendix A</b> ) ..... | 128 |
| <b>Table 38.</b> Significant results from multivariate MaAsLin2 analysis of smoking, alcohol (units/week), brushing (times per day), oral hygiene score, denture wear, mouthwash use and number of missing teeth (negative coefficient values denote reduced abundance) .....                                             | 132 |
| <b>Table 39.</b> Taxa with significant changes in abundance associated with age and gender after adjusting for smoking habit, number of missing teeth and brushing habit (negative coefficient value denotes reduced abundance) (*= $p_{adj} < 0.1$ ) .....                                                               | 135 |
| <b>Table 40.</b> Results of MaAsLin2 analysis of a site and smoking model showing taxa with significant changes in abundance (only $p_{adj} < 0.05$ values shown due to volume of results, full results available in <b>Appendix B</b> ). (Negative coefficient values denote reduced abundance) .....                    | 138 |
| <b>Table 41.</b> Significant differences between the microbiomes of smokers and non-smokers from previous studies.....                                                                                                                                                                                                    | 152 |
| <b>Table 42.</b> Demographic data of sample cohort and healthy controls .....                                                                                                                                                                                                                                             | 160 |
| <b>Table 43.</b> OLK, CLN and HC sites and clinical appearance of OLK of final sample cohort .....                                                                                                                                                                                                                        | 161 |
| <b>Table 44.</b> Table showing p values from comparisons of alpha diversities of HC (Healthy controls), CLN (contralateral normal) and OLK sites on different indices (i.e., p values from <b>Figure 30</b> ). Significant results in bold. ....                                                                          | 163 |
| <b>Table 45.</b> Genera showing changes in abundance associated with OLK on MaAslin2 analysis (negative coefficient value denotes reduced abundance) *= $p_{adj} < 0.1$ .....                                                                                                                                             | 166 |

|                                                                                                                                                                                                                                                                                                                                                                                                     |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Table 46.</b> Taxa with significant changes in abundance associated with OLK on MaAsLin2 analysis.....                                                                                                                                                                                                                                                                                           | 167 |
| <b>Table 47.</b> Taxa with significant changes in abundance associated with OLK on DESeq2 analysis (negative log2 values denote reduced abundance) .....                                                                                                                                                                                                                                            | 168 |
| <b>Table 48.</b> Taxa with significant changes in abundance associated with OLK adjusted for smoking on MaAsLin2 analysis (negative coefficient values denote reduced abundance) .....                                                                                                                                                                                                              | 169 |
| <b>Table 49.</b> Taxa significantly associated with OLK adjusted for alcohol consumption on MaAsLin2 analysis (negative values denote reduced abundance) .....                                                                                                                                                                                                                                      | 172 |
| <b>Table 50.</b> Taxa significantly associated with OLK adjusted for brushing habit (per day) on MaAsLin2 analysis (negative coefficient values denote reduced abundance) .....                                                                                                                                                                                                                     | 174 |
| <b>Table 51.</b> Taxa significantly associated with OLK adjusted for oral hygiene category on MaAsLin2 analysis (negative coefficient values denote reduced abundance, OH=oral hygiene) .                                                                                                                                                                                                           | 176 |
| <b>Table 52.</b> Taxa significantly associated with OLK adjusted for denture wear on MaAsLin2 analysis (negative coefficient values denote reduced abundance) .....                                                                                                                                                                                                                                 | 178 |
| <b>Table 53.</b> Taxa significantly associated with OLK adjusted for mouthwash use on MaAsLin2 analysis (negative coefficient values denote reduced abundance) .....                                                                                                                                                                                                                                | 179 |
| <b>Table 54.</b> Taxa significantly associated with OLK after adjusting for missing teeth and OLK-associated taxa influenced by number of missing teeth on MaAsLin2 analysis (Negative coefficient values denote reduced abundance). .....                                                                                                                                                          | 180 |
| <b>Table 55.</b> Taxa significantly associated with OLK after adjusting for mucosal site on MaAsLin2 analysis. Site associated (if any) also shown with p value and adjusted p value (negative coefficient values denote reduced abundance). (Buccal mucosa/alveolus/gingiva used as reference, DT/LBT=dorsum tongue/lateral border tongue, FOM/VT=floor of mouth/ventral tongue, PAL=palate) ..... | 181 |
| <b>Table 56.</b> Results of DESeq2 analysis of OLK versus contralateral normal sites by mucosal site (negative log2 fold change denotes reduced abundance) (Buc/Alv/Gin=buccal mucosa/alveolus/gingiva, DT/LBT=dorsum tongue/lateral border tongue, FOM/VT=floor of mouth/ventral tongue, PAL=palate).....                                                                                          | 183 |
| <b>Table 57.</b> Results of MaAsLin2 analysis showing taxa significantly associated with OLK after adjusting for mucosal site and smoking. (Only OLK results shown).....                                                                                                                                                                                                                            | 186 |
| <b>Table 57.</b> Summary table showing taxa significantly associated with OLK alone and associated with OLK after adjusting for various potentially influencing variables on MaAslin2 analysis (X denotes significant association where padj<0.1) (m/wash=mouthwash) .....                                                                                                                          | 190 |
| <b>Table 60.</b> Final sample cohort sample sites and histological findings.....                                                                                                                                                                                                                                                                                                                    | 195 |
| <b>Table 61.</b> Table of alpha diversity measures from <b>Figure 41</b> (significant results in bold) .....                                                                                                                                                                                                                                                                                        | 196 |
| <b>Table 62.</b> Genera with significant changes in abundance associated with LROLK, HROLK and OSCC compared with CLN sites on MaAsLin2 analysis (negative coefficient values denotes reduced abundance).....                                                                                                                                                                                       | 202 |
| <b>Table 63.</b> Taxa with significant changes in abundance associated with LROLK, HROLK and OSCC compared with CLN sites on MaAsLin2 analysis. ....                                                                                                                                                                                                                                                | 203 |
| <b>Table 64.</b> Results of DESeq2 analysis comparing LROLK and CLN sites showing taxa with significant changes in abundance in LROLK relative to CLN sites (negative log2 fold change values denote reduced abundance).....                                                                                                                                                                        | 204 |
| <b>Table 65.</b> Results of DESeq2 analysis comparing HROLK and CLN sites showing taxa with significant changes in abundance in HROLK relative to CLN sites. (Yellow=increased in HROLK, green=decreased in HROLK) .....                                                                                                                                                                            | 205 |
| <b>Table 66.</b> Results of both MaAsLin2 and DESeq2 analysis of HROLK versus LROLK showing taxa with significant changes in abundance in HROLK relative to LROLK. (Yellow=increased in HROLK, green=decreased in HROLK) NA=not detected on analysis .....                                                                                                                                          | 206 |

|                                                                                                                                                                                                                                                                                                                   |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Table 67.</b> Results of DESeq2 analysis comparing OSCC with CLN, LROLK and HROLK .....                                                                                                                                                                                                                        | 207 |
| <b>Table 68.</b> Taxa with significant changes in abundance associated with LROLK, HROLK and OSCC after adjusting for smoking habit on MaAsLin2 analysis (smoking results not shown, negative coefficient values denote reduced abundance). .....                                                                 | 209 |
| <b>Table 69.</b> Taxa with significant changes in abundance associated with LROLK, HROLK and alcohol consumption on MaAsLin2 analysis. ....                                                                                                                                                                       | 211 |
| <b>Table 70.</b> Taxa significantly associated with LROLK, HROLK and OSCC after adjusting for brushing habit (per day) on MaAsLin2 analysis (negative coefficient values denote reduced abundance). ....                                                                                                          | 212 |
| <b>Table 71.</b> Taxa with significant changes in abundance associated with LROLK, HROLK and oral hygiene category on MaAsLin2 analysis. ....                                                                                                                                                                     | 213 |
| <b>Table 72.</b> Taxa significantly associated with LROLK, HROLK and OSCC after adjusting for denture wear on MaAsLin2 analysis (negative coefficient values denote reduced abundance)....                                                                                                                        | 214 |
| <b>Table 73.</b> Taxa with significant changes in abundance associated with LROLK and HROLK after adjusting for mouthwash use on MaAsLin2 analysis .....                                                                                                                                                          | 215 |
| <b>Table 74.</b> Taxa associated with LROLK, HROLK and OSCC after adjusting for missing teeth on MaAsLin2 analysis .....                                                                                                                                                                                          | 216 |
| <b>Table 75.</b> Taxa significantly associated with LROLK, HROLK and OSCC after adjusting for mucosal site on MaAsLin2 analysis. Site associated (if any) also shown with p value and adjusted p value. (Buccal mucosa/alveolus/gingiva used as reference) .....                                                  | 217 |
| <b>Table 76.</b> Results of DESeq analysis of HROLK versus LROLK for DT/LBT sites showing taxa with significant changes in abundance in HROLK relative to LROLK at these sites. MaAsLin2 associations also shown (padj=adjusted p value) (negative log2 fold change denotes reduce abundance). *= padj <0.1 ..... | 218 |
| <b>Table 77.</b> Results of DESeq2 analysis of HROLK v LROLK for PAL sites showing taxa with significant changes in abundance in HROLK relative to LROLK at these sites. MaAsLin2 associations also shown (padj=adjusted p value).....                                                                            | 219 |
| <b>Table 78.</b> Results of DESeq2 analysis of HROLK v LROLK for FOM/VT sites showing taxa with significant changes in abundance in HROLK relative to LROLK at these sites. MaAsLin2 associations also shown (padj=adjusted p value).....                                                                         | 219 |
| <b>Table 79.</b> Results of DESeq2 analysis of HROLK v LROLK for Buc/Alv/Gin sites showing taxa with significant changes in abundance in HROLK relative to LROLK at these sites. MaAsLin2 associations also shown (padj=adjusted p value, negative log2 fold change denotes reduced abundance) (*=padj<0.1) ..... | 220 |
| <b>Table 80.</b> LROLK- and HROLK-associated taxa showing significant changes in abundance on a MaAslin2 model adjusted for multiple measures of oral hygiene .....                                                                                                                                               | 221 |
| <b>Table 79.</b> Results of MaAsLin2 analysis showing taxa significantly associated with LROLK, HROLK and OSCC after adjusting for mucosal site and smoking. (Only LROLK, HROLK and OSCC results shown, negative coefficient values denote reduced abundance). ....                                               | 222 |
| <b>Table 80.</b> Taxa showing significant changes in abundance associated with degree of dysplasia on a MaAslin2 model adjusted for multiple measures of oral hygiene.....                                                                                                                                        | 225 |
| <b>Table 84.</b> Results of MaAsLin2 analysis showing taxa significantly associated with no, mild, moderate and severe dysplasia and OSCC after adjusting for mucosal site and smoking. (Only dysplasia and OSCC results shown. Negative coefficient values denote reduced abundance) ....                        | 227 |
| <b>Table 85.</b> Summary table showing taxa significantly associated with LROLK alone and associated with LROLK after adjusting for various potential confounding variables (X denotes significant association where padj<0.1).....                                                                               | 230 |
| <b>Table 86.</b> Summary table showing taxa with significant changes in abundance associated with HROLK alone and associated with HROLK after adjusting for various potential confounding variables (X denotes significant association where padj<0.1, *denotes reduced abundance). .....                         | 232 |

|                                                                                                                                                                                                                                                                                                                                                    |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Table 87.</b> Summary table showing taxa significantly associated with OSCC alone and associated with OSCC after adjusting for various potentially influencing variables (X denotes significant association where $p_{adj} < 0.1$ ).....                                                                                                        | 232 |
| <b>Table 88.</b> Results of best models with LASSO (100 iterations) and random forest (10 iterations) (AUC = area under the curve).....                                                                                                                                                                                                            | 241 |
| <b>Table 89.</b> Results of MaAsLin2 analysis showing taxa with significant changes in abundance ( $p_{adj} < 0.1$ ) associated with site. (Buc/Alv/Gin [buccal mucosa/alveolus/gingiva] used as reference) (negative values denote reduced abundance). DT/LBT=dorsum/lateral border tongue, FOM/VT=floor of mouth/ventral tongue, PAL=palate..... | 288 |
| <b>Table 90.</b> Results of MaAsLin2 analysis of a site and smoking model showing taxa with significant changes in abundance ( $p_{adj} < 0.1$ ) (Negative coefficient values denote reduced abundance) .....                                                                                                                                      | 291 |

# 1 Introduction

Leukoplakia, defined as a ‘white plaque of questionable risk having excluded (other) known diseases or disorders that carry no increased risk for cancer’<sup>1</sup>, is the most common of the 11 recognised oral potentially malignant disorders (OPMDs) of the oral mucosa in the Western world, with an estimated worldwide prevalence of 4.1%<sup>2</sup> and an average malignant transformation rate to oral squamous cell carcinoma (OSCC) of 9.8% over 10 years<sup>3,4</sup>. OSCC accounts for over 90% of all oral cavity cancers,<sup>5</sup> lip and oral cavity cancer being the 18<sup>th</sup> most common cancer worldwide in 2020, with 377,713 new cases (2% of all cancers) and 177,757 deaths<sup>6</sup>.

Worldwide, there is regional variation in the incidence of oral cancer, with particularly high incidences reported in parts of Northern France, Eastern Europe, South America and South-East Asia<sup>5</sup>. In Sri Lanka, India, Pakistan and Bangladesh, oral cancer is the most common cancer in men and accounts for up to a quarter of all new cancers diagnosed annually<sup>7</sup>. Unfortunately, despite improvements in overall cancer survival rates, the 5-year all-stage survival rate of oral cancer remains unchanged at approximately 50%<sup>7</sup>.

As for some other cancers, the evidence for the role of bacteria and microbial dysbiosis in OSCC is growing and the advent of next-generation DNA sequencing (NGS) for bacterial identification has led to an explosion in microbiome data. Sequencing of short fragments of bacterial DNA from mixed populations is now possible, allowing rapid identification and quantification of whole bacterial communities in a single sequencing run. The bacterial 16S rRNA gene, pioneered by Carl Woese and George E. Fox in 1977<sup>8</sup>, is the gene of choice for NGS due to its nine variable regions (V1-V9) which have sufficient sequence diversity to



allow classification to species level. The choice of V region(s) depends on the likely colonisers of the environment under investigation, with regions V1-V4 typically chosen for the oral microbiome, although regions V3-V4 have been recently shown to have limited resolution for determining taxonomy at species level<sup>9</sup>.

The emerging evidence for microbial dysbiosis in OSCC may in time lead to diagnostic and chemotherapeutic advances which in turn could help to improve the morbidity and mortality of this devastating disease.

## 2 Literature Review

### 2.1 Role of microorganisms in cancer

Microorganisms are estimated to cause up to 20% of fatal human malignancies, with an estimated 2.2 million infection-attributed cancers diagnosed worldwide in 2018<sup>10</sup>, an increase of 200,000 cases in 10 years<sup>11</sup>. Microbes are thought to contribute to carcinogenesis by altering the balance between the immune system, host cell proliferation and death, and host and microbial metabolism<sup>12</sup>. To date, 11 microbes have been designated as human carcinogens by the International Agency for Research on Cancer (IARC): *Helicobacter pylori* (gastric cancer, non-Hodgkin lymphoma), Hepatitis B virus (liver cancer), Hepatitis C virus (liver cancer, non-Hodgkin lymphoma), *Opisthorchis viverrini* (liver cancer), *Clonorchis sinensis* (liver cancer), Human papillomavirus (cervical, penile, vulval, vaginal, anal and oropharyngeal cancer), Epstein-Barr virus (nasopharyngeal cancer, Hodgkin and non-Hodgkin lymphoma), Human herpes virus 8 (Kaposi sarcoma), Human T-cell Lymphotropic Virus type 1 (non-Hodgkin lymphoma), *Schistosoma haematobium* (bladder cancer) and Human Immunodeficiency Virus-1 (cervical, conjunctival, anal, vulval, vaginal, penile, liver and non-melanoma skin cancer and Hodgkin and non-Hodgkin lymphoma)<sup>13</sup>.

However, this list has not been updated in more than a decade, despite recent significant advances in microbiome research which suggest that many other microbial species may play a role in carcinogenesis. One of the most prominent of these is the oral bacterial pathogen *Fusobacterium nucleatum*, which is increasingly being linked with colorectal and oesophageal carcinoma<sup>14–20</sup>. Increased abundance of *Fusobacterium* species has been found in colorectal adenomas compared with normal colonic mucosa and increasing abundance of *Fusobacterium nucleatum* has been demonstrated along the normal-adenoma-

adenocarcinoma sequence<sup>14-16</sup>. *Fusobacterium nucleatum* has also been isolated directly from colorectal cancer tissue samples at up to 400 times greater levels than those of healthy tissue<sup>15,16</sup>. Furthermore, the amount of *Fusobacterium* DNA isolated from colorectal cancers has been shown to correlate with lymph node and liver metastases<sup>15,18</sup>, shorter survival<sup>17</sup> and recurrence<sup>18,20</sup>. *Fusobacterium nucleatum* has also been detected in oesophageal squamous cell carcinoma and significantly associated with hypo-differentiation, more lymph node metastases and reduced five-year and median survival<sup>19</sup>. Interestingly, infection was positively correlated with male gender, smoking and alcohol consumption. *Fusobacterium nucleatum* has also recently been associated with breast cancer with Parhi *et al.* detecting the pathogen in 30% of 50 breast cancer cases sampled<sup>21</sup>. They related this association to increased expression of Gal-GalNAC in breast tumours to which the *Fusobacterium nucleatum* antigen Fap2 can bind, as in colorectal carcinoma. Nejman *et al.*<sup>22</sup>, who demonstrated that different cancer types have distinct microbiomes, also identified a very high prevalence of *Fusobacterium nucleatum* (20-25%) in breast cancer. *Streptococcus infantis* was also found to be highly prevalent (30-50%) in breast cancer in this study.

After the colon, the oral cavity has the most complex microbiome in the human body, with over 1000 species of bacteria identified, along with fungi, viruses, protozoa and archaea<sup>23,24</sup>. Conditions here are ideal for microbial growth, with plentiful nutrients, oxygen, moisture, and varied habitats available. Attention is therefore now turning to the possible role of the oral microbiome in oral cancer.

## 2.2 Risk factors for OSCC

OSCC is the most common oral cavity cancer diagnosed worldwide (**Figure 1**). It is often diagnosed late, with more than half of cases having regional metastases at diagnosis and is therefore associated with significant morbidity and mortality<sup>25</sup>.



**Figure 1.** OSCC on the right lateral border of the tongue consisting of a 1cm ulcer with raised irregular margins. There is a non-homogenous leukoplakia distal to the ulcer. Photograph courtesy of Dr. S. Galvin.

Established risk factors for OSCC include tobacco and alcohol consumption, areca nut use, increased age and presence of an OPMD. Tobacco is the most significant of these risk factors with the population attributable risk of OSCC from tobacco alone being almost 25%, the risk of OSCC increasing with the number of cigarettes smoked and up to 80% of OSCC occurring in smokers<sup>5,26,27</sup>. Females appear to be more susceptible to the effects of smoking, their risk of OSCC being approximately twice that of males who smoke<sup>27</sup>. Tobacco and

alcohol are known to act synergistically, increasing the risk of oral cancer up to 15-fold, particularly in the floor of the mouth<sup>5,28-30</sup>. However, up to 15% of OSCC cases in older adults, and up to 26% of cases in younger adults have little to no exposure to these traditional risk factors<sup>31</sup>. A recent retrospective cohort study in the United States showed a progressive decrease in the prevalence of tobacco and alcohol use in oral cancer patients over a 25-year period<sup>32</sup>, while statistics from the Centers for Disease Control (CDC) in the US confirm that efforts in promoting tobacco cessation have led to a decrease in oral cancer rates associated with tobacco use<sup>33</sup>. However, over the same time period, Tota *et al.* demonstrated an increase in the proportion of tongue and oropharyngeal cancers in adults under 50 years of age<sup>34</sup> agreeing with older studies showing that OSCCs in non-smokers were more likely to be in females and located on the tongue or gingiva<sup>35-38</sup>. Deneuve *et al.*'s recent retrospective study<sup>39</sup> of 185 tongue cancers in those aged 18-40 years, also supports these results, with two thirds of their population being non-smokers (less than 100 cigarettes in their lifetime). Non-smokers also had a significantly lower prevalence of alcohol use but a significantly higher prevalence of OLK. Rock *et al.*'s longitudinal cohort study looking at progression of OPMDs in 455 patients also concurs, finding that OPMDs in never-smokers were four times more likely to progress than those in current smokers, with those on the floor of the mouth and tongue at highest risk<sup>40</sup>.

More recently, low consumption of fresh fruit and vegetables has been associated with an increased risk of OSCC<sup>31,41-44</sup>. Finally, multiple epidemiological studies and meta-analyses have associated poor oral hygiene, tooth loss and periodontal disease with OSCC<sup>29,44-53</sup>. However, whether this association is because chronic periodontal disease shares risk factors with OSCC, namely smoking and alcohol use, is still unclear. Several studies have found that the association between periodontal disease and OSCC was less after adjusting for

smoking and alcohol consumption<sup>52</sup>. A recent pooled analysis of 2449 head and neck squamous cell carcinomas<sup>54</sup> found that an increased number of remaining natural teeth and more frequent dental visits were associated with greater overall survival, even after adjusting for tumour size, smoking, age and socioeconomic status, suggesting that the relationship between oral health and head and neck cancer is valid. Although, whether the association is mechanistic or is due to poorer oral health being a surrogate marker for poorer overall health<sup>55</sup> is, as yet, unclear.

## 2.3 Effect of smoking and alcohol on the oral microbiome

As discussed, smoking and alcohol consumption are established risk factors for OSCC. However, these agents have also been shown to alter the oral microbiome, independently of the presence of an OPMD or OSCC.

*In vitro* studies have demonstrated that nicotine inhibits the growth of *Porphyromonas gingivalis*<sup>56</sup>. Early culture-based studies on the effect of smoking on the oral microbiome found that smoking inhibits the growth of some species of *Neisseria*, *Staphylococcus aureus* and certain streptococci such as *Streptococcus pneumoniae*<sup>57,58</sup>, its effect on *Neisseria* being particularly potent, resulting in a change to a Gram-negative heavy microbiome<sup>58</sup>. Since then, numerous studies, using various different sampling methods and substrates including mucosal swabs, saliva and plaque collection, have been carried out to determine the effect of smoking and alcohol consumption on the oral microbiome.

### 2.3.1 Smoking swab studies

Charlson *et al.*, chose the V1-V2 region of the 16S rRNA gene in one of the first studies to use NGS to investigate the effect of smoking on the microbiome. They compared the oropharyngeal communities, sampled by mucosal swabs, of 29 smokers and 33 non-smokers (less than 100 lifetime cigarettes), sampled by mucosal swabs, and found that smokers had significantly increased abundances of *Megasphaera* and *Veillonella* spp. and significantly reduced abundances of *Neisseria*, *Fusobacterium* and *Capnocytophaga*<sup>59</sup>. A slightly smaller Brazilian study of 22 patients, using oral mucosal swab samples and the V1 region of the 16S rRNA gene, found significant increases in Fusobacteria and Actinobacteria and decreases in Firmicutes, Cyanobacteria, TM7 and BD1-5 in smokers, but, interestingly, a decrease in *Fusobacteria* in those who both smoked and consumed alcohol<sup>60</sup>. Amer *et al.*<sup>61</sup> used the V1-V2 region of 16S rRNA in a swab study of 36 patients with buccal and tongue OLKs and found reduced levels of *Fusobacterium nucleatum*, *Neisseria* spp. and *Leptotrichia* in smokers. However, a more recent Indian study found increased abundance of *Fusobacterium* in tobacco users<sup>62</sup>. Swab samples of the buccal mucosa were used to compare the microbial profiles of smokers, smokeless tobacco users and healthy controls, with the authors finding that tobacco use was associated with significantly greater alpha diversity compared with healthy controls. The beta diversities of tobacco users also differed significantly from healthy controls, with taxa from the genus *Fusobacterium* in particular driving these differences. However, it should be noted that no oral examinations were carried out to rule out mucosal pathology which would have been particularly important in the smokeless tobacco cohort as these products are often placed against the buccal mucosa.

### 2.3.2 Smoking saliva studies

Morris *et al.* used 16s rRNA profiling to examine the bacterial composition of oral wash and bronchial lavage samples of 19 smokers and 45 non-smokers and found *Neisseria*, *Gemella* and *Porphyromonas* significantly decreased in smokers<sup>63</sup>. Belstrøm *et al.*<sup>64</sup>, in a salivary study of healthy Danish subjects, using DNA-DNA hybridisation and micro-array profiling, found *Streptococcus sobrinus* and *Eubacterium brachy* significantly increased in smokers, although only 7.2% of their study population smoked. No significant differences were noted between the saliva of former and never smokers. Wu *et al.*'s large salivary study of 1204 adults<sup>65</sup>, drawn from a cohort of 147,449 adults involved in two large cancer prevention studies<sup>66,67</sup> found significant enrichment of Actinobacteria and Firmicutes and depletion of Proteobacteria in current smokers compared with never/former smokers. Fusobacteria was not found to be enriched in smokers in this study, although only 9% of the study population were current smokers. The overall composition of the microbiome did not differ between former and never smokers, as previously demonstrated in Belstrøm *et al.*'s salivary study<sup>64</sup>, suggesting that smoking-related changes were reversible. However, the study had several notable limitations: no oral examinations were carried out prior to sampling to rule out mucosal disease, sampling was done by many of the patients at home, no details were provided on pre-sampling antibiotic or steroid use, patients with diabetes, inflammatory bowel disease or immune-related disease were not excluded, and sample collection was 11 to 20 years prior to DNA extraction.

Hsiao *et al.*'s case control study of male Han Chinese<sup>68</sup> investigated the influence of smoking, alcohol, areca quid, and oral hygiene on the salivary microbiome of OSCC and analysed the effect of smoking and alcohol on the microbiome in the control group. They found that *Fusobacterium nucleatum* alone, and *Fusobacterium nucleatum*, *Prevotella*



*tanneriae* and *Prevotella intermedia* combined, were significantly increased among smokers. They postulated that smoking could lead to microbial dysbiosis which could increase the risk of OSCC<sup>68</sup>. Finally, Wirth *et al.*<sup>69</sup> used shotgun metagenomic sequencing to investigate the effect of cigarette smoking on the salivary microbiome, including 11 smokers and 11 non-smokers in their study. Significant differences were found between the smoking and non-smoking cohorts, with *Prevotella* and *Megasphaera* showing significantly increased abundance in smokers while *Neisseria*, *Oribacterium*, *Capnocytophaga* and *Porphyromonas* were significantly enriched in non-smokers.

### 2.3.3 Smoking plaque studies

Moon *et al.*<sup>70</sup> were one of the first to use NGS sequencing to investigate the effect of smoking on the microbiome of subgingival plaque. They compared the subgingival plaque of smoking and non-smoking Korean patients with moderate periodontitis using the V1-V3 region of the 16S rRNA gene and observed differences in bacterial communities between the two groups at all phylogenetic levels. At phyla level, Fusobacteria were significantly more abundant in smokers while Bacteroidetes were more abundant in non-smokers. At the genus level *Fusobacterium*, *Freitbacterium*, *Streptococcus*, *Veillonella*, *Corynebacterium*, *TM7*, *Filifactor*, *Tannerella*, *Lautropia*, *Eubacterium*, *Gemella*, *Granulicatella*, *Parvimonas*, *Mogibacterium*, *Bacteroides*, *Leptotrichia* and *Rothia* were 1.5 times more abundant in smokers, while at the species level, smoking was associated with a significant, and more than two-fold, increase in abundance of seven bacterial species, including *Fusobacterium nucleatum*, and a significantly decreased abundance of 11 other species, demonstrating that smoking has a significant effect on the subgingival microbiome.

Mason *et al.*<sup>71</sup> also compared the subgingival microbiome in smokers and non-smokers, but in healthy individuals with no periodontal disease. They found that the microbiome in smokers had a lower abundance of commensals and was enriched with anaerobic and periodontal pathogens, including *Fusobacterium nucleatum*, surmising that it was more closely aligned with a periodontal disease-associated community than a healthy one.

The key findings of the studies discussed exploring the differences in the microbiome between smokers and non-smokers are summarised in **Table 1**.

**Table 1.** Significant differences between the microbiomes of smokers and non-smokers from previous studies

| Author                                    | Sampling method                                                                                                               | Periodontal health                                                                       | Increased in smokers v non-smokers                                                                                               | Decreased in smokers v non-smokers                                                                                                                               |
|-------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Charlson <i>et al.</i> 2010 <sup>59</sup> | Oropharyngeal swabs<br>29 smokers<br>33 non-smokers                                                                           | Not documented<br>No antibiotics for 3 months prior to sampling                          | <i>Megasphaera</i><br><i>Veillonella</i>                                                                                         | <i>Neisseria</i><br><i>Fusobacterium</i><br><i>Capnocytophaga</i>                                                                                                |
| Thomas <i>et al.</i> 2014 <sup>60</sup>   | Mucosal swabs<br>13 smokers<br>9 never smokers                                                                                | Not measured                                                                             | <i>Fusobacteria</i><br><i>Actinobacteria</i><br><i>Capnocytophaga</i><br><i>Prevotella</i>                                       | Firmicutes<br>Cyanobacteria<br>TM7<br>BD1-5<br><i>Neisseria</i><br><i>Peptostreptococcus</i><br><i>Gemella</i><br><i>Granulicatella</i><br><i>Staphylococcus</i> |
| Amer <i>et al.</i> 2017 <sup>61</sup>     | Mucosal swabs (OLK study, subset analysis)<br>19 current smokers<br>10 former smokers<br>7 never smokers                      | Not measured                                                                             |                                                                                                                                  | <i>Fusobacterium nucleatum</i><br><i>Neisseria</i><br><i>Leptotrichia</i>                                                                                        |
| Gopinath <i>et al.</i> 2021 <sup>62</sup> | Buccal mucosa swabs<br>17 male smokers<br>13 male non-smokers                                                                 | Not measured, no oral examinations                                                       | <i>Fusobacterium</i> spp. incl.<br><i>Fusobacterium nucleatum</i><br><i>Saccheribacterium</i> spp.<br><i>Shuttleworthia</i> spp. |                                                                                                                                                                  |
| Morris <i>et al.</i> 2013 <sup>63</sup>   | Saliva<br>19 smokers<br>45 non-smokers                                                                                        |                                                                                          |                                                                                                                                  | <i>Neisseria</i><br><i>Gemella</i><br><i>Porphyromonas</i>                                                                                                       |
| Belstrøm <i>et al.</i> 2014 <sup>64</sup> | Saliva<br>21 smokers<br>123 former smokers<br>148 never smokers<br>Note no differences found between former and never smokers | Low levels of caries and periodontal disease – significant disease an exclusion criteria | <i>Streptococcus sobrinus</i><br><i>Eubacterium brachy</i>                                                                       |                                                                                                                                                                  |

|                                     |                                                                  |                                       |                                                                                                                                               |                                                                                                                                                                                                                                                                                    |
|-------------------------------------|------------------------------------------------------------------|---------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Wu <i>et al.</i> 2016 <sup>65</sup> | Saliva<br>112 smokers<br>571 former smokers<br>521 never smokers | Not measured, no<br>oral examinations | Actinobacteria<br>Firmicutes<br><i>Atopobium</i> <i>Streptococcus</i><br><i>Bifidobacterium</i><br><i>Lactobacillus</i><br><i>Veillonella</i> | Proteobacteria<br><i>Neisseria</i><br><i>Haemophilus</i><br><i>Aggregatibacter</i><br><i>Capnocytophaga</i><br><i>Corynebacterium</i><br><i>Porphyromonas</i><br><i>Prevotella</i><br><i>Peptostreptococcus</i><br><i>Leptotrichia</i><br><i>Abiotrophia</i><br><i>Selenomonas</i> |
|-------------------------------------|------------------------------------------------------------------|---------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                        |                                                                              |                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                               |
|----------------------------------------|------------------------------------------------------------------------------|-----------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Hsiao <i>et al.</i> 2018 <sup>68</sup> | Saliva<br>104 male smokers<br>47 male never smokers                          | Not measured                                                    | <i>Fusobacterium nucleatum</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                               |
| Wirth <i>et al.</i> 2020 <sup>69</sup> | Saliva<br>11 smokers<br>11 non-smokers (8 never, 3 former - > 1yr cessation) | Dental examinations, moderate and severe periodontitis excluded | <i>Prevotella</i><br><i>Megasphaera</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | <i>Neisseria</i><br><i>Capnocytophaga</i><br><i>Porphyromonas</i><br><i>Oribacterium</i>                                                                                                                                                                                                                                      |
| Moon <i>et al.</i> 2015 <sup>70</sup>  | Subgingival plaque<br>134 smokers<br>134 non-smokers                         | Moderate periodontitis                                          | <i>Fusobacteria</i><br><i>Fusobacterium</i><br><i>Freitbacterium</i><br><i>Streptococcus</i><br><i>Veillonella</i><br><i>Corynebacterium</i><br><i>TM7</i><br><i>Filifactor</i><br><i>Parvimonas</i><br><i>Tannerella</i><br><i>Rothia</i><br><i>Lautropia</i><br><i>Eubacterium</i><br><i>Gemella</i><br><i>Granulicatella</i><br><i>Mogibacterium</i><br><i>Bacteroides</i><br><i>Leptotrichia</i>                                                                                                                                             | <i>Bacteroidetes</i><br><i>Neisseria</i><br><i>Campylobacter</i><br><i>Haemophilus</i><br><i>Prevotella</i><br><i>Aggregatibacter</i><br><i>Selenomonas</i><br><i>Veillonella</i><br><i>Porphyromonas</i><br><i>Cardiobacterium</i>                                                                                           |
| Mason <i>et al.</i> 2015 <sup>71</sup> | Subgingival plaque<br>200 subjects                                           | Periodontally healthy                                           | <i>Fusobacterium nucleatum</i><br><i>Fusobacterium naviforme</i><br><i>Filifactor alocis</i><br><i>Dialister microaerophilus</i><br><i>Megasphaera sueciensis</i><br><i>Megasphaera geminatus</i><br><i>Megasphaera elsdenii</i><br><i>Megasphaera</i><br><i>micronuciformis</i><br><i>Acinetobacter johnsonii</i><br><i>Acinetobacter guillouiae</i><br><i>Acinetobacter schindleri</i><br><i>Acinetobacter baumannii</i><br><i>Acinetobacter haemolyticus</i><br><i>Pseudomonas pseudoalcaligenes</i><br><i>Pseudoramibacter alactolyticus</i> | <i>Streptococcus sanguinis</i><br><i>Streptococcus parasanguinis</i><br><i>Streptococcus oralis</i><br><i>Gemella elegans</i><br><i>Gemella adiacens</i><br><i>Acinetobacter viscosus</i><br><i>Acinetobacter israelii</i><br><i>Acinetobacter dentalis</i><br><i>Neisseria subflava</i><br><i>Haemophilus parainfluenzae</i> |

### 2.3.4 Alcohol studies

Hsiao's previously mentioned study<sup>68</sup> also found *Fusobacterium nucleatum*, *Prevotella tannerae* and *Prevotella intermedia* significantly increased in saliva samples among consumers of alcohol, independent of smoking, oral hygiene and OSCC. A larger saliva study of 1044 US adults found significant enrichment of *Fusobacteria*, *Bacteroidetes*, *Proteobacteria* and *Actinobacteria* and reduced abundance of *Lactobacillales* in moderate and heavy alcohol consumers compared with non-consumers<sup>72</sup>. However, this study's participants were drawn from the same cohort, and investigated using the same protocol, as that of Wu *et al.*<sup>65</sup> and so the same limitations highlighted above apply.

Chronic users of alcohol are known to have a higher prevalence of periodontal disease<sup>49,73,74</sup>, with smoking a significant co-factor<sup>74</sup>. Lages *et al.*<sup>75</sup> demonstrated that alcohol dependence was significantly associated with periodontitis and that patients with periodontitis and alcohol dependence had significantly higher levels of the periodontal pathogens *Prevotella intermedia*, *Fusobacterium nucleatum* and *Eikenella corrodens* in subgingival samples. Further, the abundances of these bacteria decreased sequentially from the alcohol-dependence/periodontitis group to the no alcohol/periodontitis group to the alcohol-dependence/no periodontitis group and finally to the no alcohol/no periodontitis group.

The oral microbiome is clearly affected by smoking and alcohol use, both of which are also risk factors for oral cancer and so it may be that the dysbiosis caused by these risk factors, along with that caused by periodontal disease and poor oral hygiene, has a role to play in oral carcinogenesis. However, it also appears that different sampling methods can impact on study findings.

## 2.4 The microbiome in OSCC

Early studies using culture-based methods showed that OSCC surface biofilms contained significantly increased numbers of aerobes and anaerobes compared to contralateral normal sites, with *Fusobacterium nucleatum* and *Porphyromonas gingivalis* in particular being enriched<sup>76,77</sup>. The advent of next generation sequencing (NGS)<sup>78</sup> using the 16S rRNA gene has led to a multitude of studies investigating the bacterial colonisation of OSCC using a variety of sampling methods and substrates, including mucosal swabs, tissue biopsies and saliva collection, and there is general agreement that the oral microbiome in OSCC differs to that in health<sup>9,67,79–104</sup>. One small systematic review and meta-analysis reported increased abundances of *Fusobacteria* and reduced abundances of *Streptococcus* associated with OSCC, although some of the studies included in the review included throat and tonsil cancers while others did not compare with contralateral normal sites<sup>105</sup>. As will be discussed, there is significant variation in the literature.

### 2.4.1 OSCC swab studies

Surface swabs enable direct and targeted sampling of oral mucosal surfaces. Being non-invasive, they pose no risk and cause no discomfort to patients. Their principal advantage however is that they allow for separate sampling of lesional and normal sites within the same individual.

Schmidt *et al.*<sup>79</sup> were one of the first groups to use NGS to investigate the microbiome of OSCC. They swabbed OSCCs and contralateral normal sites in 18 patients and, using the V4 region of 16S rRNA, found that abundances of both Firmicutes (particularly *Streptococcus* spp.) and Actinobacteria (particularly *Rothia* spp.) were significantly

decreased at OSCC sites relative to contralateral normal sites whereas there was a significant increase in abundance of *Fusobacterium* spp. at cancer sites. Zhao *et al.*<sup>80</sup> used the V4-V5 region of the 16S rRNA gene to investigate the bacterial colonisation of OSCC in 40 Chinese patients. Again, swabs were taken from the OSCC and a contralateral normal site in each patient. They found significant enrichment of *Fusobacterium*, *Dialister*, *Peptostreptococcus*, *Filifactor*, *Peptococcus*, *Catonella* and *Parvimonas* species and reduced abundance of Firmicutes and Actinobacteria at OSCC sites compared with contralateral normal sites. Mok *et al.*<sup>81</sup> also used mucosal swab sampling in their microbiome study. Their nine OSCC samples came from a biobank, however it was not specified how they were sampled. Healthy controls were swabbed on the tongue, floor of mouth, lower gingiva, buccal mucosa, hard and soft palate and upper gingiva, and all sites were combined for analysis. Patient with less than 20 teeth were excluded, but no other parameters of dental health were measured. The OSCC group were older than the controls but were matched for smoking and alcohol consumption. *Actinobacteria*, *Proteobacteria*, *Neisseria*, *Gemella* and *Granulicatella* were enriched in OSCC while *Streptococcus* and *Veillonella* showed increased abundance in normal controls.

More recently Zhang *et al.*<sup>82</sup> compared the microbiome, using the V3-V4 region, of 50 OSCCs with the contralateral normal buccal mucosa within the same patients. They found *Fusobacterium*, *Alloprevotella* and *Porphyromonas* species significantly increased ( $p < 0.05$ ) and *Streptococcus*, *Veillonella* and *Rothia* species significantly decreased ( $p < 0.05$ ) in OSCC. Differences were found in other species and genera, but p values were not given. It is worth noting that only patients who had taken antibiotics in the previous week were excluded, in contrast to most microbiome studies which exclude patients who have taken antibiotics in the previous two months. Limited data was also provided on how smokers,

alcohol drinkers and areca quid users were defined. Finally, Su *et al.*<sup>83</sup> examined the microbiome of 116 buccal mucosa OSCCs in males in Taiwan using the V4 region, comparing it to contralateral normal sites. 93% of those sampled were users of areca quid. They found significant increases in *Fusobacterium*, *Peptostreptococcus*, *Campylobacter*, *Prevotella*, and *Capnocytophaga* and a significant decrease in *Streptococcus* spp. at tumour sites (although p values were not given). Further analysis found that *Streptococcus* and *Fusobacterium* could differentiate tumour sites from normal sites with confidence intervals (CI) of 0.67-0.86 and 0.61-0.78 respectively. However, the generalisability of this study's findings is limited somewhat by single site sampling, male only gender, and high areca quid use.

#### 2.4.2 OSCC tissue studies

Direct tissue sampling enables investigation of invasive species within oral lesions. This method of sampling is more invasive, requiring mucosal biopsy, and so carries some risks for patients including post-operative pain, bleeding, and infection. Therefore, in recent studies, ethical considerations have prevented the biopsy sampling of normal tissue as a control. Nonetheless, tissue studies have shown some interesting results, in particular in relation to *Fusobacterium nucleatum* as shown by Zhang *et al.*<sup>106</sup> who used a *Fusobacterium nucleatum* specific probe and fluorescence *in situ* hybridisation to visualise *Fusobacterium nucleatum* within OSCC (n=10) tissues and found the species to be highly abundant within OSCC and present at much higher levels than in normal healthy tissue (n=10).

Hooper *et al.* were the first to investigate the presence of bacteria within OSCC tissue using biopsy<sup>107,108</sup>. They sampled both superficial and deep OSCC tissues in 20 patients and found



viable bacteria in both areas, with the same isolates being found in both deep and superficial specimens. Control samples were taken five centimetres from the OSCC site. Although no statistical differences between OSCC and normal tissue sites were found, several species, *Exiguobacterium oxidotolerans*, *Prevotella melaninogenica*, *Staphylococcus aureus*, *Veillonella parvula*, *Micrococcus luteus*, and *Fusobacterium naviforme*, were isolated only from OSCC while several others, *Moraxella osloensis*, *Prevotella veroralis*, and *Actinomyces* spp., were only found at normal sites. Most of the species isolated were saccharolytic and acid tolerant, in keeping with the hypoxic and acidic tumour environment. Pushalkar *et al.*<sup>85</sup> used electrophoresis followed by NGS to compare the bacteria within 10 OSCC with normal tissue samples taken 5 cm from the OSCC site. In contrast to other studies, they found Firmicutes highly prevalent at tumour sites, although not significantly so. Al-hebshi *et al.*<sup>88</sup> sequenced the V1-V3 region of the 16S rRNA gene in their biopsy study of the OSCC microbiome. They identified *Fusobacterium nucleatum* subsp. *polymorphum*, *Pseudomonas aeruginosa* and *Campylobacter* sp. oral taxon 44 as being significantly associated with OSCC, while *Streptococcus mitis*, *Rothia mucilaginosa* and *Haemophilus parainfluenzae* were associated with healthy controls, sampled by mucosal swabbing of anatomically matched sites. When they analysed their findings by site, *Fusobacterium nucleatum* subsp. *polymorphum* was significantly associated with tongue cancers, while *Pseudomonas aeruginosa* was associated with tongue and gingival cancers.

Perera *et al.*<sup>86,87</sup> documented the mycobiome and microbiome of OSCC in two separate reports, sampling 25 OSCC and 27 fibroepithelial polyps (FEP) as controls, using NGS and the ITS2 gene and the V1-V3 region of the 16S rRNA gene respectively. The mycobiome study identified the genera *Candida*, *Hannaella* and *Gibberella* as significantly more abundant in OSCC. *Candida albicans* was identified in all samples but was significantly

more abundant in OSCC than in FEP. The microbiome study of the same cohort found significantly increased abundances of *Campylobacter concisus*, *Prevotella salivae*, *Prevotella loeschii*, *Fusobacterium* oral taxon 204, *Fusobacterium nucleatum* subsp. *polymorphum*, *Streptococcus dysgalactiae*, *Citrobacter koseri*, and *Pseudomonas aeruginosa* in OSCC while seven *Streptococcus* species including *Streptococcus mitis*, two *Rothia* species, *Lautropia mirabilis*, *Leptotrichia* oral taxon 255, and *Staphylococcus epidermidis* were more abundant in FEP. The authors noted however that the OSCC cohort were older, reported more tobacco and alcohol use and had more severe periodontal disease than the control cohort which could have contributed to the findings. Zhou *et al.*<sup>89</sup> sampled 24 OSCC, comparing them with 24 tissue samples taken 2 cm from the tumour using the V3-V4 region of the 16S rRNA gene, and found increased abundances of Bacteroidetes, Firmicutes, Fusobacteria and Spirochaetes and reduced abundances of Actinobacteria and Cyanobacteria associated with OSCC. At the genera level *Carnobacterium*, *Filifactor*, *Fusobacterium*, *Parvimonas*, *Tannerella*, *Peptostreptococcus*, *Streptococcus* and *Treponema* showed increased abundance associated with OSCC.

Ye *et al.*<sup>92</sup> used the same sampling methodology as Zhou *et al.*<sup>89</sup> in their microbiome study of 23 tongue OSCC and found increased abundance of Bacteroidetes and reduced abundance of Actinobacteria, *Actinomyces*, *Streptococcus* and *Neisseria* associated with OSCC. They noted that patients had systemic periodontal treatment prior to sampling, although what this entailed was not explained. Sarkar *et al.*<sup>91</sup> also used the V3-V4 region of the 16S rRNA gene when investigating the microbiome of 50 OSCC and 50 adjacent histologically normal sites. Agreeing with previous studies, they found increased abundances of Bacteroidetes and Fusobacteria associated with OSCC. Abundances of *Prevotella*, *Pseudomonas*, *Corynebacterium*, *Deinococcus* and *Noviherbaspirillum* were also increased in OSCC while

abundances of *Actinomyces*, *Anoxybacillus*, *Serratia*, *Strenotrophomonas* and *Sutterella* were decreased. Finally, Herreros-Pomares *et al.*<sup>93</sup> compared the microbiomes of 10 OSCC, eight other OSCCs which had been preceded by proliferative verrucous leukoplakia (PVL-OSCC), an oral potentially malignant disorder with a high rate of malignant transformation (see **section 2.5**), and 11 healthy tissue samples. Using the V3-V4 region of the 16S rRNA gene, they found that Fusobacteria comprised 34% of the OSCC and 30% of the PVL-OSCC sequences, compared with 24% of those in the healthy cohort. Interestingly, the relative abundances of Proteobacteria and Campylobacteria differed significantly between the PVL-OSCC and OSCC cohorts, Proteobacteria showing significantly reduced (11% versus 19%) and Campylobacteria showing significantly increased (14% versus 4%) abundance in the PVL-OSCC cohort compared with the OSCC cohort. At the species level, OSCC was found to have increased abundance of several species of *Capnocytophaga*, *Metamycoplasma salivarium* and *Prevotella nanceiensis* while PVL-OSCC showed increased abundance of *Lachospiraceae* bacterium, *Selenomonas sputigena* and *Prevotella shahii*. The authors demonstrated for the first time that PVL-OSCC displays a different microbial community structure to OSCC that is not preceded by PVL, with both also differing from healthy controls.

### 2.4.3 OSCC saliva studies

The ease of collection of saliva has facilitated numerous studies on the microbiome in OSCC. However, saliva collection does not allow for sampling of specific sites/lesions and has also been shown to be strongly biased towards tongue and palate communities<sup>109</sup>. The saliva sampled can also be contaminated with food and other debris, and although there are

various protocols described for pre-sampling rinsing to minimise contamination, this contributes further to the issue of differing sampling methodology in the literature.

Before the era of NGS, Mager *et al.*<sup>110</sup> analysed the salivary microbiome of 45 OSCC patients and 229 controls using checkerboard DNA-DNA hybridisation and found significantly elevated levels of *Prevotella melaninogenica*, *Capnocytophaga gingivalis*, *Capnocytophaga ochracea* and *Streptococcus mitis* and significantly decreased levels of *Leptotrichia buccalis* and *Eubacterium saburreum* in OSCC patients compared with controls ( $p < 0.001$ ). In the first study to use NGS, Pushalkar *et al.*<sup>84</sup> sampled saliva from three OSCC and two controls and found that the most prevalent genera in the OSCC cohort were *Streptococcus*, *Gemella*, *Rothia*, *Peptostreptococcus*, *Porphyromonas* and *Lactobacillus*. In the control group, *Prevotella*, *Neisseria*, *Leptotrichia*, *Capnocytophaga*, *Actinobacillus*, *Oribacterium*, SR1 and TM7 were predominantly observed. Statistical significance was not reached however, likely due in part to the small sample size. Guerrero-Preston *et al.*<sup>102</sup> also used saliva to investigate the microbiome associated with head and neck squamous cell cancers (HNSCC). Their study included 6 OSCC, 11 oropharyngeal squamous cell carcinomas (OPSCC) and 25 controls. They used the V3-V5 region of the 16S rRNA gene and found that microbial communities from control samples had significantly higher richness ( $p < 0.001$ ) and diversity ( $p < 0.001$ ) than HNSCC samples and, in contrast with other studies, HNSCC samples had increased abundances of Firmicutes compared with controls. They also found that *Streptococcus* spp. and *Lactobacillus* spp. were significantly enriched in HNSCC samples versus controls ( $p < 0.05$ ), although it must be noted that the sample size was small and combined OSCC and OPSCC. Hu *et al.*<sup>94</sup> looked at the salivary microbiome of 16 patients with OSCC and 10 with oral leukoplakia (OLK), comparing them with 19 healthy controls. In contrast to most other studies, they found the

highest diversity in the OSCC group. Comparing phyla, Bacteroidetes and TM7 were the most, and Firmicutes the least, abundant in saliva of OSCC patients. At the genus level, *Bacillus* was the most abundant genus in the OSCC group, while *Streptococcus* and *Abiotrophia* were significantly more abundant in the control group. Details on smoking, alcohol or areca nut use were not provided.

Takahashi *et al.*<sup>111</sup> sampled saliva from 60 OSCC patients and 80 controls and, using the V3-V4 region of 16S rRNA, found significant differences between the cancer and control groups. *Peptostreptococcus*, *Fusobacterium*, *Alloprevotella*, and *Capnocytophaga* were more abundant, and *Rothia* and *Haemophilus* less abundant, in the cancer group compared to the control group ( $p < 0.01$ ). Lee *et al.*'s saliva study<sup>99</sup> was larger, involving 376 participants, comprised of 127 normal controls, 125 OSCC and 124 patients with either dysplasia, hyperplasia or hyperkeratosis. They labelled the latter patient group as a 'precursor' group, although some forms of hyperkeratosis such as frictional keratosis carry no increased risk for oral cancer. Using the V4 region of 16S rRNA they found Bacteroidetes and Escherichia were present in more than 60% of samples in the precursor and cancer groups compared with 37% and 14.4% of samples respectively in the control group. It should be noted that there was a very high prevalence of areca quid use in this study, with 90% of OSCC patients, compared with only 27% of controls, being current or former users.

In a large study in Taiwan, Yang *et al.*<sup>112</sup> collected saliva from 197 OSCC patients and 51 controls. Using the V3-V4 region of the 16S rRNA gene, they found that the abundance of *Fusobacterium* was significantly increased ( $p < 0.0001$ ) and the abundances of *Streptococcus*, *Haemophilus*, *Porphyromonas*, and *Actinomyces* significantly decreased ( $p$

< 0.0001) in OSCC compared with healthy controls. Interestingly they found a progressive significant increase in *Fusobacterium*, and a decrease in Bacteroidetes and Actinobacteria from normal oral mucosa to stage 1 OSCC to stage 4 OSCC. Hayes *et al.*<sup>113</sup> also analysed saliva in a nested case control study of 129 HNSCC patients, 41 of whom had OSCC. They found that a greater abundance of the commensal bacteria *Corynebacterium* and *Kingella* was associated with a reduced risk of HNSCC. This association was strongest in smokers. *Parvimonas micra* and *Neisseria sicca* were associated with a reduced risk of OSCC while an unnamed *Actinomyces* species was associated with increased risk of OSCC. They concluded that commensal bacteria and a healthy oral microbiome were associated with reduced risk for HNSCC. However, the study had several limitations including no baseline oral examinations to exclude occult malignancy or OPMD, saliva samples taken at home by the patients and sent to the investigators via post, no details on oral hygiene practices and no information on pre-sample antibiotic or steroid use. Also, in contrast to most other microbiome studies, diabetes, immune-system disease, and inflammatory bowel disease were not considered exclusion factors. Hashimoto *et al.*<sup>95</sup> sequenced the V4 region of the 16S rRNA from salivary samples of six OSCC patients, six patients with OLK and four controls, although they did not use the control samples for comparison due age differences from the OSCC and OLK cohorts. They found increased abundance of Bacteroidetes and reduced abundance of *Streptococcus* in OSCC relative to OLK. However, it should be noted that one of the OSCCs was in fact a carcinoma-*in-situ*, which is not invasive cancer but severe dysplasia<sup>114</sup>, and so should have been classified as OLK. Also, the histology was only provided for one OLK which showed mild dysplasia and so it is not certain that these were in fact true OLKs.

Ganly *et al.*<sup>96</sup> used the V3-V4 region of the 16S rRNA gene to investigate the salivary microbiome of 18 OSCC in non-smokers (never and ex-smokers of more than 10 years), and found the periodontal pathogens *Fusobacterium*, *Prevotella* and *Alloprevotella* significantly increased, and *Streptococcus* significantly decreased, in OSCC compared with controls. However, oral hygiene and periodontal disease status were not recorded. Gopinath *et al.*<sup>98</sup> compared the salivary microbiome of 30 OSCCs with 23 healthy controls, using the V3-V4 region of the 16S rRNA gene, finding increased abundance of *Porphyromonas* associated with OSCC, while *Streptococcus*, *Rothia* and *Fusobacteria* were more associated with controls. However, there were significant differences between the OSCC and control patients with regards to smoking and alcohol consumption, while the OSCC cohort was also significantly older than the control cohort. Oral hygiene and periodontal disease were not recorded. Chen *et al.*<sup>100</sup> carried out a salivary microbiome study including only males with what they termed ‘oral verrucous hyperplasia-associated oral cancer’, although it is not clear what this entity is. They found that the families *Flavobacteriaceae*, *Peptostreptococcaceae*, *Mycoplasmataceae*, *Carnobacteriaceae*, *Lachnospiraceae*, and *Peptoniphilaceae* and the species *Capnocytophaga sputigena*, *Catonella morbi*, and *Dilaster invisus* were specific to the oral cancer cohort. Sawant *et al.*<sup>101</sup> unusually chose to sequence the V6-V8 region of the 16S rRNA gene, with the result that species level taxonomy was not achieved, in their study comparing 10 OSCC with 10 healthy non-smoking controls, finding that *Prevotella* dominated the OSCC cohort while streptococci dominated the control group, which also showed increased abundance of *Rothia* relative to the OSCC cohort. The genera *Pseudomonas*, *Peptococcus*, *Morganella*, *Alloscardovia*, *Klebsiella*, *Bacteroides*, and *Propionibacterium* were found to only be present in the OSCC cohort. Ganly *et al.*<sup>97</sup> published a second salivary study on the microbiome of OSCC in non-smokers, this time using metagenomic shotgun sequencing, and found increased abundances of Bacteroidetes,

Synergistetes, Spirochaetes and *Fusobacterium* and reduced abundances of Firmicutes, Actinobacteria, *Corynebacterium*, *Streptococcus*, *Actinomyces*, *Cryptobacterium*, *Prevotella* and *Selenomonas* associated with OSCC. As in their previous study however, oral hygiene and periodontal status were not recorded. Finally, Zhu *et al.*<sup>9</sup> used the V3-V4 region of the 16S rRNA gene in their study of the salivary microbiome of OSCC, and found increased abundance of Fusobacteria, including *Fusobacterium nucleatum*, and Bacteroidetes, and reduced abundances of Actinobacteria, Elusimicrobia, *Streptococcus oralis* and *Rothia mucilanginosa* in OSCC compared with controls. As in previous studies, the choice of the V3-V4 region resulted in limited resolution of taxonomy to species level. Oral hygiene and periodontal disease were also not measured.

Two microbiome studies to date have used multiple methods of sampling. Zhou *et al.*<sup>90</sup> sampled OSCCs via biopsy, swab and saliva, healthy controls via saliva and also swabbed contralateral normal sites in the OSCC patients. The inclusion criteria included no alcohol or tobacco for the previous six months, and exclusion criteria concomitant fungal infection, mucosal disease, and presence of significant dental disease or calculus. Simpson and Shannon indices unsurprisingly differed between all five different sampling techniques/sites. *Veillonella* and *Fusobacterium* were more abundant and *Streptococcus*, *Neisseria*, *Prevotella* and *Haemophilus* less abundant in the OSCC cohort. It is unclear in the paper however whether the three OSCC sampling methods (biopsy, swab and saliva) were combined for analysis or analysed separately. Heng *et al.*<sup>103</sup> also used multiple methods of sampling when investigating the bacteriome and mycobiome of OSCC. They sampled the normal buccal mucosa (but not the OSCCs) via swab and also sampled plaque and saliva in 29 patients with OSCC and 30 controls. They found *Bacteroides*, *Alloprevotella* and *Prevotella* increased in OSCC patients in both buccal swab and plaque samples but not in saliva samples. Buccal samples in OSCC patients also showed increased



abundance of Fusobacteria, *Fusobacterium*, *Aggregatibacter*, *Capnocytophaga* and the fungus *Acremonium*, while plaque samples showed increased *Capnocytophaga* and the fungi *Aspergillus* and *Acremonium*. Saliva samples from OSCC patients showed increased abundance of *Neisseria*, *Aggregatibacter*, *Capnocytophaga* and *Aspergillus*. The only genera that showed increased abundance associated with OSCC in all three sampling methodologies was *Capnocytophaga*, highlighting the influence of the choice of sampling methodology on the results. There was more consistency of results between sampling methods when decreased abundance was considered with Firmicutes, *Streptococcus* and *Veillonella* showing reduced abundance associated with OSCC on buccal swabs, plaque and saliva. In addition, Ascomycota was decreased on buccal swabs, *Morchella* on plaque samples and Actinobacteria, Ascomycota, and *Morchella* in saliva. The key findings of these studies exploring the differences in the microbiome between OSCC and normal oral mucosa are summarised in **Table 2**. It is clear that sampling methods, as in the microbiome smoking and alcohol studies, can have a significant bearing on results (**Table 2**). Firstly, many more studies using swab and biopsy sampling of OSCC<sup>77,79,80,82,83,87–89,91,103</sup> than using saliva sampling<sup>9,90,96,97,104,111</sup> identify increased abundances of Fusobacteria and *Fusobacterium* associated with OSCC. Secondly, several studies using saliva sampling identified Firmicutes, streptococci and *Rothia* as more abundant in OSCC patients<sup>84,102,110</sup>, a finding not replicated in any of the swab sampling studies, where they are typically associated with controls<sup>79–83</sup>. Finally, it is also worth noting that the physical characteristics of oral cancers, often large and deep ulcers on the tongue or floor of mouth, provide environments amenable to anaerobic colonisation making it difficult to determine the significance of the OSCC microbiome in relation to cause or effect.

In conclusion, although the findings of studies differ, all agree that a shift in the microbiome is associated with OSCC. When this shift happens however is yet to be established.

**Table 2.** Significant differences between the microbiomes of OSCC and normal mucosa from previous studies

| Author                                     | Sampling method | Bacterial characterisation                                                      | OSCC/controls                                        | Increased in OSCC v control                                                                                                                                                                                                                                                                    | Decreased in OSCC v control or increased in control V OSCC                                                                                               |
|--------------------------------------------|-----------------|---------------------------------------------------------------------------------|------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|
| Nagy <i>et al.</i> 1998 <sup>77</sup>      | Swab (culture)  | Culture                                                                         | 21 OSCC<br>21 contiguous mucosa controls             | <i>Fusobacterium</i> spp.                                                                                                                                                                                                                                                                      |                                                                                                                                                          |
| Schmidt <i>et al.</i> 2014 <sup>79</sup>   | Swab            | 16S rRNA V4 sequencing                                                          | 18 OSCC<br>18 contralateral controls                 | <i>Fusobacterium</i> spp.                                                                                                                                                                                                                                                                      | <i>Streptococcus</i> spp.<br><i>Rothia</i> spp.                                                                                                          |
| Zhao <i>et al.</i> 2017 <sup>80</sup>      | Swab            | 16S rRNA V4-V5 sequencing                                                       | 40 OSCC<br>40 contralateral controls                 | <i>Fusobacterium</i><br><i>Dialister</i><br><i>Peptostreptococcus</i><br><i>Filifactor</i><br><i>Peptococcus</i><br><i>Catonella</i><br><i>Parvimonas</i>                                                                                                                                      | Firmicutes<br>Actinobacteria                                                                                                                             |
| Mok <i>et al.</i> 2017 <sup>81</sup>       | Swab            | 16S rRNA restriction fragment length polymorphism and cloned library sequencing | 9 OSCC<br>9 dysplastic OPMD<br>9 controls            | Actinobacteria<br>Proteobacteria<br><i>Neisseria</i><br><i>Gemella</i><br><i>Granulicatella</i>                                                                                                                                                                                                | <i>Streptococcus</i><br><i>Veillonella</i>                                                                                                               |
| Zhang <i>et al.</i> 2019 <sup>82</sup>     | Swab            | 16S rRNA V3-V4 sequencing                                                       | 50 OSCC<br>50 buccal mucosa controls                 | <i>Fusobacterium</i><br><i>Alloprevotella</i><br><i>Porphyromonas</i>                                                                                                                                                                                                                          | <i>Streptococcus</i><br><i>Veillonella</i><br><i>Rothia</i>                                                                                              |
| Su <i>et al.</i> 2020 <sup>83</sup>        | Swab            | 16S rRNA V4 sequencing                                                          | 116 buccal OSCC<br>116 contralateral controls        | <i>Fusobacterium</i><br><i>Peptostreptococcus</i><br><i>Campylobacter</i><br><i>Prevotella</i><br><i>Capnocytophaga</i>                                                                                                                                                                        | <i>Streptococcus</i>                                                                                                                                     |
| Pushalkar <i>et al.</i> 2012 <sup>85</sup> | Biopsy          | 16S rRNA V4-V5 sequencing                                                       | 10 OSCC<br>10 controls - biopsies 5cm from OSCC site | Firmicutes (not significant):<br><i>Streptococcus</i> spp.<br><i>Gemella</i><br><i>Johnsonella ignava</i><br><i>Peptostreptococcus</i> spp.                                                                                                                                                    | <i>Granulicatella adiacens</i>                                                                                                                           |
| Al-hebshi <i>et al.</i> 2017 <sup>88</sup> | Biopsy          | 16S rRNA V1-V3 sequencing                                                       | 20 OSCC<br>20 matched site control swabs             | <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum</i><br><i>Pseudomonas aeruginosa</i><br><i>Campylobacter</i> sp. Oral taxon 44                                                                                                                                                            | <i>Rothia mucilaginosa</i><br><i>Haemophilus parainfluenzae</i><br><i>Streptococcus mitis</i>                                                            |
| Perera <i>et al.</i> 2017 <sup>86</sup>    | Biopsy          | ITS2 sequencing                                                                 | 22 OSCC<br>25 Fibroepithelial polyps as controls     | <i>Candida</i><br><i>Hannaella</i><br><i>Gibberella</i><br>Species level:<br><i>Candida albicans</i> ,<br><i>Candida etchellii</i>                                                                                                                                                             | <i>Trametes</i><br><i>Alternaria</i>                                                                                                                     |
| Perera <i>et al.</i> 2018 <sup>87</sup>    | Biopsy          | 16S rRNA V1-V3 sequencing                                                       | 25 OSCC<br>27 Fibroepithelial polyps as controls     | <i>Campylobacter concisus</i><br><i>Prevotella salivae</i><br><i>Prevotella loeschii</i><br><i>Fusobacterium</i> oral taxon 204<br><i>Fusobacterium nucleatum</i> subsp. <i>polymorphum</i><br><i>Streptococcus dysgalactiae</i><br><i>Citrobacter koseri</i><br><i>Pseudomonas aeruginosa</i> | <i>Streptococcus</i> spp.<br><i>Rothia</i> spp.<br><i>Lautropia mirabilis</i><br><i>Leptotrichia</i> oral taxon 225<br><i>Staphylococcus epidermidis</i> |

|                                                    |                |                                    |                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                        |
|----------------------------------------------------|----------------|------------------------------------|--------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Zhou <i>et al.</i> 2020 <sup>89</sup>              | Biopsy         | 16S rRNA V3-V4 sequencing          | 24 OSCC<br>24 para-OSCC, histologically normal tissue                          | Bacteroidetes<br>Firmicutes<br>Fusobacteria<br>Spirochaetes<br><i>Carnobacterium</i><br><i>Filifactor</i><br><i>Fusobacterium</i><br><i>Parvimonas</i><br><i>Tannerella</i><br><i>Peptostreptococcus</i><br><i>Streptococcus</i><br><i>Treponema</i>                                                                                                                                                       | Actinobacteria<br>Cyanobacteria<br><i>Streptophyta</i><br><i>Brevundimonas</i><br><i>Paenibacillus</i><br><i>Microbacterium</i><br><i>Desulfovibrio</i><br><i>Mucispirillum</i><br><i>Arthrobacter</i> |
| Sarkar <i>et al.</i> 2021 <sup>91</sup>            | Biopsy         | 16S rRNA V3-V4 sequencing          | 50 OSCC<br>50 para-OSCC histologically normal tissue                           | Bacteroidetes<br>Fusobacteria<br><i>Corynebacterium</i><br><i>Deinococcus</i><br><i>Noviherbaspirillum</i><br><i>Prevotella</i><br><i>Pseudomonas</i>                                                                                                                                                                                                                                                      | <i>Actinomyces</i><br><i>Anoxybacillus</i><br><i>Serratia</i><br><i>Stenotrophomonas</i><br><i>Sutterella</i>                                                                                          |
| Ye <i>et al.</i> 2021 <sup>92</sup>                | Biopsy         | 16S rRNA V3-V4 sequencing          | 23 OSCC<br>23 para-OSCC histologically normal tissue                           | Bacteroidetes                                                                                                                                                                                                                                                                                                                                                                                              | Actinobacteria<br><i>Actinomyces</i><br><i>Streptococcus</i><br><i>Neisseria</i>                                                                                                                       |
| Herreros-Pomares <i>et al.</i> 2023 <sup>93</sup>  | Biopsy         | 16R rRNA V3-V4 sequencing          | 10 OSCC<br>8 PVL-OSCC<br>11 histologically normal tissue from healthy controls | <b>OSCC</b><br>Fusobacteria<br><i>Capnocytophaga leadbetteri</i><br><i>Capnocytophaga sputigena</i><br><i>Capnocytophaga gingivalis</i><br><i>Capnocytophaga showae</i><br><i>Metamycoplasma salivarium</i><br><i>Prevotella nanceiensis</i><br><br><b>PVL-OSCC</b><br><i>Campylobacter</i><br>Fusobacteria<br><i>Lachospiraceae bacterium</i><br><i>Selenomonas sputigena</i><br><i>Prevotella shahii</i> | <i>Streptococcus</i><br><i>Campylobacter gracilis</i><br><i>Filifactor alocis</i><br><i>Phocaicola absecessus</i><br><i>Treponema maltophilum</i>                                                      |
| Mager <i>et al.</i> 2005 <sup>110</sup>            | Saliva         | Checkerboard DNA-DNA hybridisation | 45 OSCC<br>229 controls                                                        | <i>Prevotella melaninogenica</i><br><i>Capnocytophaga gingivalis</i><br><i>Streptococcus mitis</i><br><i>Capnocytophaga ochracea</i>                                                                                                                                                                                                                                                                       | <i>Leptotrichia buccalis</i><br><i>Eubacterium saburreum</i>                                                                                                                                           |
| Pushalkar <i>et al.</i> 2011 <sup>84</sup>         | Saliva         | 16S rRNA V4-V5 sequencing          | 3 OSCC<br>2 controls                                                           | <i>Streptococcus</i><br><i>Gemella</i><br><i>Rothia</i><br><i>Peptostreptococcus</i><br><i>Lactobacillus</i><br><i>Porphyromonas</i>                                                                                                                                                                                                                                                                       | <i>Prevotella</i><br><i>Neisseria</i><br><i>Leptotrichia</i><br><i>Capnocytophaga</i><br><i>Actinobacillus</i><br><i>Oribacterium</i><br>SR1 and TM7                                                   |
| Guerrero-Preston <i>et al.</i> 2016 <sup>102</sup> | Saliva (HNSCC) | 16S rRNA V3-V5 sequencing          | 6 OSCC<br>11 OPSCC<br>25 controls                                              | Firmicutes<br><i>Streptococcus</i> spp.<br><i>Lactobacillus</i> spp.                                                                                                                                                                                                                                                                                                                                       | Bacteroidetes<br>Proteobacteria                                                                                                                                                                        |

|                                             |        |                                           |                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|---------------------------------------------|--------|-------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Hu <i>et al.</i> 2016 <sup>94</sup>         | Saliva | 16S rRNA sequencing, region not specified | 16 OSCC<br>10 OLK<br>19 controls                                                                                                              | <i>Bacillus</i>                                                                                                                                                                                                                                                                                                                                                  | <i>Streptococcus</i><br><i>Abiotrophia</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Takahashi <i>et al.</i> 2017 <sup>111</sup> | Saliva | 16S rRNA V3-V4 sequencing                 | 60 OSCC<br>80 controls                                                                                                                        | <i>Peptostreptococcus</i><br><i>Fusobacterium</i><br><i>Alloprevotella</i><br><i>Capnocytophaga</i>                                                                                                                                                                                                                                                              | <i>Rothia</i><br><i>Haemophilus</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Lee <i>et al.</i> 2017 <sup>115</sup>       | Saliva | 16S rRNA V4 sequencing                    | 125 OSCC<br>124 dysplasia/<br>hyperplasia/<br>hyperkeratosis<br>127 controls                                                                  | Bacteroidetes<br>Escherichia                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Yang <i>et al.</i> 2018 <sup>112</sup>      | Saliva | 16S rRNA V3-V4 sequencing                 | 197 OSCC<br>51 controls                                                                                                                       | Fusobacteria                                                                                                                                                                                                                                                                                                                                                     | <i>Streptococcus</i><br><i>Haemophilus</i><br><i>Porphyromonas</i><br><i>Actinomyces</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Hayes <i>et al.</i> 2018 <sup>113</sup>     | Saliva | 16S rRNA V3-V4 sequencing                 | 129 HNSCC<br>41 OSCC<br>254 controls                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                  | <i>Parvimonas micra</i><br><i>Neisseria sicca</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Hashimoto <i>et al.</i> 2019 <sup>95</sup>  | Saliva | 16S rRNA V4 sequencing                    | 6 OSCC*<br>6 OLK**                                                                                                                            | Bacteroidetes                                                                                                                                                                                                                                                                                                                                                    | <i>Streptococcus</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Ganly <i>et al.</i> 2019 <sup>96</sup>      | Saliva | 16S rRNA V3-V4 sequencing                 | 18 OSCC<br>8 OLK with<br>dysplasia<br>12 controls                                                                                             | <i>Fusobacterium</i><br><i>Alloprevotella</i><br><i>Prevotella</i> spp.                                                                                                                                                                                                                                                                                          | <i>Streptococcus</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Gopinath <i>et al.</i> 2021 <sup>98</sup>   | Saliva | 16S rRNA V3-V4 sequencing                 | 30 OSCC<br>20 OLK with<br>moderate/severe<br>dysplasia<br>23 controls                                                                         | <i>Porphyromonas</i>                                                                                                                                                                                                                                                                                                                                             | <i>Streptococcus</i><br><i>Rothia</i><br><i>Fusobacterium</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Chen <i>et al.</i> 2021 <sup>100</sup>      | Saliva | 16S rRNA V3-V4 sequencing                 | 27 OSCC (oral<br>verrucous<br>hyperplasia<br>associated oral<br>cancer)<br>21 'non-recurrent<br>oral verrucous<br>hyperplasia'<br>27 controls | Flavobacteriaceae<br>Peptostreptococcaceae<br>Mycoplasmataceae<br>Carnobacteriaceae<br>Lachnospiraceae<br>Peptoniphilaceae<br><i>Prevotella oris</i><br><i>Peptostreptococcus stomatis</i><br><i>Parvimonas micra</i><br><br><b>Present in OSCC but not in controls:</b><br><i>Capnocytophaga sputigena</i><br><i>Catonella morbi</i><br><i>Dilaster invisus</i> | Actinomycetaceae<br>Veillonellaceae<br>Prevotellaceae<br><br>Unclassified <i>Prevotella</i> sp.<br>Unclassified <i>Selenomonas</i> sp.<br><i>Saccharibacteria</i> TM7 G1<br><i>bacterium</i> HMT352<br><i>Leptotrichia</i> sp. HMT417<br><i>Actinomyces</i> sp. HMT180<br><i>Prevotella melaninogenica</i><br><i>Prevotella salivae</i><br><i>Prevotella pallens</i><br><i>Veillonella atypica</i><br><i>Veillonella dispar</i><br><i>Streptococcus salivarius</i><br><i>Haemophilus parainfluenzae</i><br><i>Megasphaera micronuciformis</i><br><i>Campylobacter concisus</i><br><i>Rothia mucilaginosa</i> |

|                                          |                                                                                        |                                    |                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|------------------------------------------|----------------------------------------------------------------------------------------|------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sawant <i>et al.</i> 2021 <sup>101</sup> | Saliva                                                                                 | 16S rRNA V6-V8 sequencing          | 10 OSCC<br>10 controls                                                                                                                                             | <i>Prevotella</i><br><br><b>Only present in OSCC:</b><br><i>Pseudomonas</i><br><i>Peptococcus</i><br><i>Morganella</i><br><i>Alloscardovia</i><br><i>Klebsiella</i><br><i>Bacteroides</i><br><i>Propionibacterium</i>                                                                                                                                                                                                                                                                                                      | <i>Streptococcus</i><br><i>Rothia</i>                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Ganly <i>et al.</i> 2022 <sup>97</sup>   | Saliva                                                                                 | Metagenomic shotgun sequencing     | 42 OSCC<br>45 controls<br>Non-smokers only                                                                                                                         | Bacteroidetes<br>Synergistetes<br>Spirochaetes<br><i>Fusobacterium</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Firmicutes<br>Actinobacteria<br><i>Corynebacterium</i><br>matruchoat<br><i>Actinomyces</i><br><i>Cryptobacterium</i><br><i>Streptococcus salivarius</i><br><i>Streptococcus cristatus</i><br><i>Streptococcus gordonii</i><br><i>Prevotella denticola</i><br><i>Prevotella oulorum</i><br><i>Prevotella histicola</i><br><i>Leptotrichia wadei</i><br><i>Veillonella parvula</i><br><i>Campylobacter gracilis</i><br>6 <i>Actinomyces</i> spp.<br><i>Selenomonas noxia</i> |
| Zhu <i>et al.</i> 2023 <sup>9</sup>      | Saliva                                                                                 | 16S rRNA V3-V4 sequencing          | 91 OSCC<br>91 controls                                                                                                                                             | <i>Fusobacteria</i><br>Bacteroidetes<br><i>Fusobacterium nucleatum</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Actinobacteria<br>Elusimicrobia<br><i>Streptococcus oralis</i><br><i>Rothia mucilanginosa</i>                                                                                                                                                                                                                                                                                                                                                                              |
| Zhou <i>et al.</i> 2021 <sup>90</sup>    | Swab, biopsy and saliva of OSCC, swab of contralateral normal site, saliva of controls | 16S rRNA V4-V5 sequencing          | <b>47 OSCC cohort:</b><br>47 saliva samples<br>18 subgingival plaque samples<br>21 OSCC swabs<br>16 contralateral normal<br>16 OSCC biopsy<br>46 controls (saliva) | <i>Veillonella</i><br><i>Fusobacterium</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <i>Streptococcus</i><br><i>Neisseria</i><br><i>Prevotella</i><br><i>Haemophilus</i>                                                                                                                                                                                                                                                                                                                                                                                        |
| Heng <i>et al.</i> 2022 <sup>103</sup>   | Buccal mucosa swabs, plaque and saliva                                                 | 16S rRNA V3-V4 and ITS1 sequencing | 29 OSCC<br>32 'pre-malignant lesions'<br>30 controls                                                                                                               | <b>Buccal swabs:</b><br>Bacteroidetes<br><i>Fusobacteria</i><br><i>Alloprevotella</i><br><i>Fusobacterium</i><br><i>Prevotella</i><br><i>Aggregatibacter</i><br><i>Capnocytophaga</i><br><i>Acremonium</i> (fungus)<br><br><b>Plaque:</b><br>Bacteroidetes<br><i>Alloprevotella</i><br><i>Prevotella</i><br><i>Capnocytophaga</i><br><i>Aspergillus</i> (fungus)<br><i>Acremonium</i> (fungus)<br><br><b>Saliva:</b><br><i>Neisseria</i><br><i>Aggregatibacter</i><br><i>Capnocytophaga</i><br><i>Aspergillus</i> (fungus) | <b>Buccal swabs:</b><br>Firmicutes<br>Ascomycota (fungus)<br><i>Streptococcus</i><br><i>Veillonella</i><br><br><b>Plaque:</b><br>Firmicutes<br><i>Streptococcus</i><br><i>Veillonella</i><br><i>Morchella</i> (fungus)<br><br><b>Saliva:</b><br>Firmicutes<br>Actinobacteria<br>Ascomycota (fungus)<br><i>Streptococcus</i><br><i>Veillonella</i><br><i>Morchella</i> (fungus)                                                                                             |

## 2.5 Oral potentially malignant disorders

As in other human cancers such as cervical and bowel cancer, oral cancer can be preceded by a precursor lesion or condition. In the mouth these are termed oral potentially malignant disorders (OPMD), a group of mucosal conditions associated with a statistically increased risk of transformation to cancer. There are currently 11 disorders of the oral mucosa that are considered OPMDs<sup>116</sup> (**Table 3**).

**Table 3.** The 11 oral potentially malignant disorders (OPMDs)

| <b>Oral potentially malignant disorders (OPMDs)</b> |
|-----------------------------------------------------|
| Leukoplakia (OLK)                                   |
| Proliferative verrucous leukoplakia (PVL)           |
| Erythroplakia (OE)                                  |
| Oral submucous fibrosis                             |
| Oral lichen planus                                  |
| Actinic keratosis (actinic cheilitis)               |
| Palatal lesions in reverse smokers                  |
| Oral lupus erythematosus                            |
| Oral lichenoid lesions                              |
| Oral graft versus host disease                      |
| Dyskeratosis congenita                              |



**Figure 2.** Homogenous leukoplakia anterior hard palate extending to the incisive papilla. Photograph courtesy of Dr. S. Galvin.

The global prevalence of OPMD is estimated at 4.5%, the majority being oral submucous fibrosis (OSF) and oral leukoplakia (OLK) (**Figure 2**) which has an estimated prevalence of 4.1%<sup>2</sup>. A significantly higher prevalence of OPMD (10.54%) is noted in Asians, likely due to regional use of areca nut<sup>2</sup>. In the Western world, the most significant OPMDs, in terms of prevalence and malignant transformation risk, are OLK, PVL and oral erythroplakia (OE)<sup>116</sup>, risk factors for which include tobacco smoking and consumption of alcohol<sup>2,117-121</sup>. The diagnosis of an OPMD is made by clinical and histopathological examination, with the histology of OLK, PVL and OE typically showing some degree of dysplasia (cytological and/or architectural atypia), which is graded as mild, moderate or severe<sup>114</sup>, although some can show no dysplasia. Binary grading of dysplasia, as low-grade and high-grade, shown by some to have higher inter-observer reliability than the current three-tier system and



advocated for research studies<sup>122–125</sup>, is also included in the most recent WHO grading system<sup>126</sup> although with the caveat that further validation is needed.



**Figure 3.** Proliferative verrucous leukoplakia right and left ventral tongue. Photograph courtesy of Dr. S. Galvin.

Iocca *et al.*<sup>127</sup>, estimated that the overall malignant transformation rate of OPMDs was 7.9%, although the rate varies substantially between conditions, with rates of 1.4% quoted for oral lichen planus up to rates of 49.5% for proliferative verrucous leukoplakia (PVL) (**Figure 3**), over periods of one to 20 years. OLK carries an overall risk of malignant transformation of 9.8%<sup>3</sup> (annual rate 1.56%), although the risk again varies, with systematic reviews showing that non-homogenous forms have a higher risk than homogenous forms<sup>3,128–130</sup>.

However, malignant transformation is difficult to predict. It is generally accepted that females<sup>3,128,129,131</sup>, non-smokers<sup>40,130–133</sup> and older persons are at increased risk<sup>3,134–136</sup>. OLKs of greater than 200 mm<sup>2</sup><sup>3,130,133,137,138</sup>, on the tongue and floor of the mouth<sup>3,40,129,130,133–136,139,140</sup> and with areas of redness<sup>3,128–132,136,138,141</sup> are also at increased risk of malignant transformation. The most accurate predictor of malignant transformation however is the degree of dysplasia on biopsy<sup>3,127,128,131,134,136</sup>, with lesions showing severe dysplasia having a 24-26% transformation rate over 10 years<sup>134,142</sup>. More recently, Kierce *et al.*<sup>143</sup> reported a malignant transformation rate of 16.7% in a retrospective chart review of 204 patients with oral epithelial dysplasia. Heavy smoking, excessive alcohol intake, size greater than 200mm<sup>2</sup>, moderate or severe dysplasia, progression of dysplasia grade, non-homogeneity and immunosuppression (including use of topical steroids) were all found to be risk factors for malignant transformation. Another recent retrospective review<sup>144</sup> of 120 patients with unifocal high-grade oral epithelial dysplasia (100 severe dysplasia and 20 moderate dysplasia, all re-classified as high-grade on retrospective slide review), who underwent incisional biopsy followed by excision (n=106) found a malignant transformation rate of 17%, with malignant transformation associated with increased age, non-homogenous clinical appearance, non-excision and positive excision margins. Surprisingly alveolar ridge and palatal locations were also associated with malignant transformation. Recurrence post-excision occurred in 34.7%, up to 12 years following excision, and was related to positive excision margins and clinical appearance of erythroplakia.

The largest systematic review on the malignant transformation of OLK<sup>3</sup> included 16,604 patients and found that non-homogeneity, presence and grade of dysplasia, female gender, age over 50 years and location on the tongue were significant risk factors for malignant

transformation. Smoking and increased size were not found to be associated with malignant transformation in this study.

Unfortunately, despite the recognition of these risk factors for malignant transformation of OPMDs, the prediction of malignant transformation is still an inexact science with a 4% transformation rate for OLKs with no or mild dysplasia reported<sup>134,145</sup>. The management of OLK, PVL and OE is also challenging, with no effective topical agents available and limited to no evidence that surgical resection can prevent malignant transformation<sup>137,142,146,147</sup>. In those that smoke, smoking cessation can lead to resolution<sup>148,149</sup> however it has been shown that up to 30% of lung and head and neck cancer patients are unable to maintain smoking cessation<sup>150</sup> and it is likely that this is also true of OLK patients. Finally, recurrence rates of 7-38% regardless of treatment modality are also reported, with recurrences and malignant transformation more likely in those who continue to smoke<sup>127,151</sup>

## 2.6 The microbiome in OPMDs

The microbiome shift that has been identified in OSCC has focussed attention on the possible role of the microbiome in the development of OSCC. This has resulted in several studies and one systematic review exploring the microbiome in OPMD to determine if there are significant differences in colonisation patterns between OPMD, OSCC and normal oral mucosa (**Table 4**).

**Table 4.** Significant differences between the microbiomes of OLK/OPMD and normal mucosa from previous studies (FFPE=formalin-fixed paraffin embedded)

| Author                                             | Sampling method                | Bacterial characterisation                                                      | Study cohort                                                                       | Increased in OLK/OPMD v normal control                                                                                                                                                                                                                                   | Decreased in OLK/OPMD v control or increased in control v OLK/OPMD                                                                                                                                                                                                                                                                                        |
|----------------------------------------------------|--------------------------------|---------------------------------------------------------------------------------|------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Schmidt <i>et al.</i> 2014 <sup>79</sup>           | Swab                           | 16S rRNA V4 sequencing                                                          | 10 'pre-cancer'<br>8 contralateral controls                                        | <i>Fusobacterium</i>                                                                                                                                                                                                                                                     | Firmicutes<br>Actinobacteria<br><i>Streptococcus</i> spp.                                                                                                                                                                                                                                                                                                 |
| Amer <i>et al.</i> 2017 <sup>61</sup>              | Swab                           | 16S rRNA V1-V2 sequencing                                                       | 36 OLK and contralateral normal<br>32 healthy controls                             | <i>Fusobacterium</i><br><i>Leptotrichia</i><br><i>Rothia mucilaginosa</i><br><i>Campylobacter</i>                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                           |
| Mok <i>et al.</i> 2017 <sup>81</sup>               | Swab                           | 16S rRNA restriction fragment length polymorphism and cloned library sequencing | 9 dysplastic OPMD<br>9 controls                                                    | Actinobacteria<br>Firmicutes<br>Bacteroidetes<br>Fusobacteria<br><i>Neisseria</i><br><i>Gemella</i><br><i>Granulicatella</i>                                                                                                                                             | <i>Streptococcus</i><br><i>Veillonella</i>                                                                                                                                                                                                                                                                                                                |
| Radaic <i>et al.</i> 2023 <sup>152</sup>           | Swab                           | 16S rRNA V4 sequencing                                                          | 4 dysplasia<br>4 histologically normal adjacent to dysplasia<br>4 healthy controls | Proteobacteria                                                                                                                                                                                                                                                           | Firmicutes<br><i>Streptococcus</i>                                                                                                                                                                                                                                                                                                                        |
|                                                    | Historical FFPE tissue samples |                                                                                 | 13 dysplasia<br>8 OSCC<br>8 histologically normal adjacent to OSCC                 | Proteobacteria                                                                                                                                                                                                                                                           | Firmicutes                                                                                                                                                                                                                                                                                                                                                |
| Decsi <i>et al.</i> 2018 <sup>153</sup>            | Biopsy                         | Metagenome sequencing                                                           | 7 hyperplasia/hyperkeratosis/dysplasia/ lichenoid<br>7 ipsilateral controls        | <i>Fusobacterium nucleatum</i>                                                                                                                                                                                                                                           | <i>Streptococcus mitis</i>                                                                                                                                                                                                                                                                                                                                |
| Herreros-Pomares <i>et al.</i> 2021 <sup>154</sup> | Biopsy                         | 16S rRNA V3-V4 sequencing                                                       | 10 PVL<br>5 controls                                                               | <i>Oribacterium</i> sp. oral taxon 108<br><i>Campylobacter jejuni</i><br>Uncultured<br><i>Eubacterium</i> spp.<br><i>Campylobacter</i><br><i>Tannerella</i><br><i>Porphyromonas</i><br><i>Peptoniphilus</i><br>Undefined<br><i>Mobiluncus</i><br><i>Anaerosalibacter</i> | <i>Streptococcus agalactiae</i><br><i>Mycoplasma adleri</i><br>Uncultured <i>Fusobacteria</i><br><i>Bacteroidales</i><br><i>Synergistetes</i><br><i>Mycoplasma</i><br><i>Bacteroidetes</i><br>Undefined<br><i>Acetobacteroidetes</i><br><i>Olsenella</i><br><i>Thermoflavifilum</i><br><i>Cellulosibacter</i><br><i>Pacearchaeota Incertae Sedis</i> AR13 |

|                                                   |        |                                           |                                                              |                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|---------------------------------------------------|--------|-------------------------------------------|--------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Herreros-Pomares <i>et al.</i> 2023 <sup>93</sup> | Biopsy | 16S rRNA V3-V4 sequencing                 | 9 OLK<br>12 PVL<br>11 healthy controls                       | <b>OLK</b><br><i>Streptococcus parasanguinis</i><br><i>Streptococcus salivarius</i><br><i>Fusobacterium periodonticum</i><br><i>Prevotella histicola</i><br><i>Porphyromonas pasteri</i><br><i>Megasphaera micronuciformis</i><br><br><b>PVL</b><br><i>Campylobacteria</i><br><i>Prevotella salivae</i><br><i>Campylobacter concisus</i><br><i>Dialister pneumosintes</i><br><i>Schaalia odontolytica</i> | <i>Campylobacter gracilis</i><br><i>Filifactor alocis</i><br><i>Phocaeicola abscessus</i><br><i>Treponema maltophilum</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| Hu <i>et al.</i> 2016 <sup>94</sup>               | Saliva | 16S rRNA sequencing, region not specified | 10 OLK (histology not given)<br>19 controls                  | Bacteroidetes<br>TM7<br><i>Haemophilus</i><br><i>Bacillus</i>                                                                                                                                                                                                                                                                                                                                             | Firmicutes<br><i>Streptococcus</i><br><i>Abiotrophia</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Lee <i>et al.</i> 2017 <sup>99</sup>              | Saliva | 16S rRNA V4 sequencing                    | 124 dysplasia/hyperplasia/hyperkeratosis<br>127 controls     | Bacteroidetes<br><i>Escherichia</i>                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Ganly <i>et al.</i> 2019 <sup>96</sup>            | Saliva | 16S rRNA V3-V4 sequencing                 | 8 OLK with dysplasia<br>12 controls<br>Non-smokers only      | <i>Fusobacterium</i><br><i>Veillonella</i>                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Gopinath <i>et al.</i> 2021 <sup>98</sup>         | Saliva | 16S rRNA V3-V4 sequencing                 | 20 OLK with moderate/severe dysplasia<br>23 controls         | <i>Megasphaera</i><br>Unclassified<br><i>Enterobacteria</i><br><i>Salmonella</i><br><i>Prevotella</i>                                                                                                                                                                                                                                                                                                     | <i>Streptococcus</i><br><i>Rothia</i><br><i>Fusobacterium</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Chen <i>et al.</i> 2021 <sup>100</sup>            | Saliva | 16S rRNA V3-V4 sequencing                 | 21 'non-recurrent oral verrucous hyperplasia'<br>27 controls | Leptotrichiaceae<br><i>Veillonella parvula</i><br><i>Rothia dentocariosa</i><br><br><b>Present in OLK but not in controls:</b><br><i>Anaeroglobus geminatus</i><br><i>Porphyromonas gingivalis</i><br><i>Prevotella oulorum</i><br><i>Saccharibacteria</i> TM7<br>G5 <i>bacterium</i> HMT356<br><i>Tannerella forsythia</i><br><i>Dilaster invisus</i>                                                    | Actinomycetaceae<br>Veillonellaceae<br>Prevotellaceae<br><br>Unclassified <i>Prevotella</i> sp.<br>Unclassified<br><i>Selenomonas</i> sp.<br><i>Saccharibacteria</i> TM7<br>G1 <i>bacterium</i> HMT352<br><i>Leptotrichia</i> sp.<br>HMT417<br><i>Actinomyces</i> sp.<br>HMT180<br><i>Prevotella melaninogenica</i><br><i>Prevotella salivae</i><br><i>Prevotella pallens</i><br><i>Veillonella atypica</i><br><i>Veillonella dispar</i><br><i>Streptococcus salivarius</i><br><i>Haemophilus parainfluenzae</i><br><i>Megasphaera micronuciformis</i><br><i>Campylobacter concisus</i><br><i>Rothia mucilaginosa</i> |

|                                        |                                                            |                                    |                                           |                                                                                                                                              |                                                                                                                                                          |
|----------------------------------------|------------------------------------------------------------|------------------------------------|-------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|
| Heng <i>et al.</i> 2022 <sup>103</sup> | Buccal mucosa swab (i.e. not lesional)<br>Plaque<br>Saliva | 16S rRNA V3-V4 and ITS1 sequencing | 32 'pre-malignant lesions'<br>30 controls | <b>Buccal swabs:</b><br>Proteobacteria<br><i>Neisseria</i><br><br><b>Plaque:</b><br>Proteobacteria<br><br><b>Saliva:</b><br><i>Neisseria</i> | <b>Buccal swabs:</b><br>Firmicutes<br><i>Streptococcus</i><br><br><b>Plaque:</b><br>Firmicutes<br><i>Veillonella</i><br><br><b>Saliva:</b><br>Firmicutes |
|----------------------------------------|------------------------------------------------------------|------------------------------------|-------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|

The lack of effective treatments for OPMD adds impetus to these studies as there would be chemotherapeutic possibilities if the microbiome was found to be implicated in the pathogenesis of OPMD and OSCC.

### 2.6.1 OPMD swab studies

Schmidt *et al.*<sup>79</sup> were one of the first groups to look at the microbiome of OPMDs, including OPMDs in their study on OSCC. They identified OPMDs as showing dysplasia on biopsy, although clinical diagnoses, e.g., OLK, were not provided. They found that, as for OSCC, abundances of *Fusobacterium* spp. were significantly increased and abundances of Firmicutes and Actinobacteria were significantly reduced at dysplastic sites relative to contralateral normal sites. At the genus level, *Streptococcus* spp. were significantly reduced on OPMDs. Amer *et al.*<sup>61</sup> investigated the microbiome of the most common OPMD, OLK, by comparing swabs of dysplastic OLKs with contralateral normal sites in the same individuals. They found significant enrichment of OLKs with *Fusobacterium*, *Leptotrichia*, *Rothia mucilaginosa* and *Campylobacter* species, relative to contralateral normal sites. They also found that *Candida* colonisation of oral leukoplakia was associated with increased levels of *Fusobacterium nucleatum*, suggesting an interplay between *Fusobacterium*

*nucleatum* and *Candida* which influences the bacterial microbiome of OLKs. Mok *et al.*<sup>81</sup> also included OPMDs in their OSCC microbiome study mentioned above but, like Schmidt *et al.*,<sup>79</sup> they identified OPMDs only as having dysplasia on biopsy without providing a clinical diagnosis. Actinobacteria, Firmicutes, Bacteroidetes, Fusobacteria, *Neisseria*, *Gemella*, and *Granulicatella* showed increased abundance associated with OPMDs, while *Streptococcus* and *Veillonella* were reduced in abundance.

Finally, Radaic *et al.*<sup>152</sup> recently carried out a small pilot study comparing the microbiomes of mild and moderately dysplastic tissue (which the authors termed low-grade dysplasia) with histologically normal adjacent tissue in four patients, and with histologically normal mandibular gingiva in four healthy controls. Clinical descriptions of the dysplastic areas were not given, oral sites were not matched, and smoking habit was not recorded for two of the patient cohort. Significant differences were found between both the alpha and beta diversities of patients and healthy controls and of histologically normal tissue adjacent to dysplasia samples and low-grade dysplasia. Phyla level analysis showed a significant increase in Proteobacteria and decrease in Firmicutes in both low-grade dysplasia and histologically normal tissue adjacent to dysplasia samples compared with the healthy controls while genus level analysis found reduced abundance of streptococci in the histologically normal tissue adjacent to dysplasia samples compared with healthy controls. Finally, at the species level, *Neisseria baciliformis* was significantly increased in abundance in histologically normal tissue adjacent to dysplasia samples compared with both healthy controls and low-grade dysplasia. The combining of mild and moderate dysplasia into a low-grade dysplasia group in this study is unusual and goes against recommendations that moderate dysplasia should be considered high grade or high-risk<sup>122,125</sup>.

### 2.6.2 OPMD tissue studies

There are only four OPMD microbiome studies to date that have used tissue sampling (**Table 4**), the first being by Decsi *et al.*<sup>153</sup> who included seven patients who had lesional biopsies for both histopathology and microbiome analysis and a third biopsy of ipsilateral normal tissue as control taken. OPMDs were not defined clearly and encompassed white or lichenoid patches showing various combinations of hyperplasia, hyperkeratosis, dysplasia, or lichenoid infiltrate on biopsy. The OPMDs were found to have increased abundance of *Fusobacterium nucleatum* and decreased abundance of *Streptococcus mitis*. Herreros-Pomares *et al.*<sup>154</sup> looked specifically at the microbiome of PVL, the OPMD with the highest rate of malignant transformation<sup>127</sup>, in their study, comparing 10 tissue samples from PVL with five tissue samples from healthy controls. They found significant enrichment of *Oribacterium* sp. oral taxon 108, *Campylobacter jejuni*, uncultured *Eubacterium* spp., *Campylobacter*, *Tannerella*, *Porphyromonas*, *Peptoniphilus*, and undefined *Mobiluncus* and *Anaerosalibacter* bacteria, while healthy samples were significantly enriched in *Streptococcus agalactiae*, *Mycoplasma adleri*, uncultured *Fusobacteria*, *Bacteroidales*, *Synergistetes*, *Mycoplasma*, *Bacteroidetes*, and undefined *Acetobacteroidetes*, *Olsenella*, *Thermoflavifilum*, *Cellulosibacter*, and *Pacearchaeota Incertae Sedis AR13*. PVL was confirmed histologically, and the cohort included none, mild, moderate, and severe degrees of dysplasia. Notably, seven patients developed at least one OSCC, with four developing more than one OSCC, within three years of sample collection. A second study by Herreros-Pomares *et al.*<sup>93</sup> compared the microbiomes of healthy tissue (n=11), homogenous OLK (n=9), PVL (n=12), OSCC (n=10) and OSCC preceded by PVL (n=8) (OSCC and PVL-OSCC findings are discussed in **section 2.4.2**). Interestingly, the microbiome of PVL differed significantly from both that of OLK and healthy controls, with PVL shown to have significantly increased abundance of Campylobacteria and significantly reduced abundance



of Proteobacteria compared to OLK and healthy controls. At the species level, PVL showed enrichment in *Prevotella salivae*, *Campylobacter concisus*, *Dialister pneumosintes* and *Schaalia odontolytica*, while OLK showed enrichment in *Streptococcus parasanguinis*, *Streptococcus salivarius*, *Fusobacterium periodonticum*, *Prevotella histicola*, *Porphyromonas pasteri* and *Megasphaera micronuciformis*. Controls were enriched in *Campylobacter gracilis*, *Filifactor alocis*, *Phocaeicola abscessus* and *Treponema maltophilum*. The authors demonstrated for the first time that the microbiomes of PVL and OLK, OPMDs with very distinct clinical behaviours, are different, which raises the question of the role of the microbiome in the natural history of these conditions. Finally, Radiac *et al.*<sup>152</sup> also analysed historical formalin-fixed paraffin-embedded (FFPE) tissue in their above-mentioned study. Archival FFPE tissue from 13 dysplastic lesions (without clinical description), eight OSCC and eight histologically normal mucosa adjacent to OSCC were compared. Six patients had matched samples available for analysis - OSCC and/or dysplasia and/or histologically normal tissue adjacent to OSCC. Significant differences in alpha and beta diversities were noted in the unmatched cohort between low-grade (mild-moderate) dysplasia and high-grade (severe) dysplasia and histologically normal tissue adjacent to OSCC. No significant differences in alpha or beta diversity were noted between the groups in the matched cohort suggesting that the differences found in the unmatched cohort may be due to confounding variables such as oral hygiene that had not been accounted for. Unmatched high-grade dysplasia samples were found to have significantly increased abundance of Proteobacteria and reduced abundance of Firmicutes compared to histologically normal adjacent samples, but no differences in abundances were found between the groups of matched samples, again suggesting the possible influence of confounding factors.

### 2.6.3 OPMD saliva studies

Hu *et al.*<sup>94</sup> also included OPMDs in the form of OLK in their salivary OSCC microbiome study, although histological data of the OLKs was not provided. They found that the relative abundances of Bacteroidetes and TM7 were greater, and the relative abundance of Firmicutes was less, in the OLK group than in the healthy control group. At the genera level, *Haemophilus* and *Bacillus* were more abundant, and *Streptococcus* and *Abiotrophia* less abundant, in OLK compared with controls. Lee *et al.*<sup>99</sup> investigated the salivary microbiome in patients with what they termed ‘precursor lesions’, defined as patches showing histological evidence of dysplasia, hyperplasia or hyperkeratosis, and found *Bacteroidetes* and *Escherichia* significantly increased compared with controls.

The previously mentioned salivary study of Ganly *et al.*<sup>96</sup> also included OLK and found *Fusobacterium* and *Veillonella* significantly increased in leukoplakias with dysplasia compared with controls, the increase showing a trend from normal to OLK to OSCC. Twenty OLKs with moderate or severe dysplasia were also included in the salivary study of Gopinath *et al.*<sup>98</sup> Bacteroidetes was the most abundant phyla in the OLK cohort and OLKs were also associated with increased abundance of *Megasphaera*, *Enterobacteria*, *Salmonella* and *Prevotella*, while healthy controls showed increased abundances of *Fusobacterium*, *Streptococcus* and *Rothia*. However, as discussed in relation to the OSCC cohort, the OLK cohort differed significantly from the healthy control group with regards to smoking and alcohol consumption.

Chen *et al.*<sup>100</sup> sampled the salivary microbiome of 21 cases of what they described as ‘non-recurrent oral verrucous hyperplasia’ (OVH) along with the OSCCs detailed earlier. It is not clear if these were cases of PVL as the histological diagnosis was not given, nor was it

specified what was meant by ‘non-recurrent’. As mentioned previously, all subjects were male and significantly more of the patient group than the control group used areca nut. Leptotrichiaceae, *Veillonella parvula* and *Rothia dentocariosa* showed significantly increased abundance in the OVH group. Five further species were specific to the OVH cohort while *Dilaster invisus* was present in the OVH and OSCC cohorts but not in the healthy control group.

Finally, Heng *et al.*, included 32 ‘pre-malignant lesions’ in their OSCC study described above. No clinical description was provided for these lesions of which 40.6% showed no dysplasia, 40.6% showed low-grade and 18.6% high grade, dysplasia on biopsy. It is not clear therefore how many of the 32 lesions included were true OPMDs. As in the OSCC study, the lesions themselves were not sampled, sampling instead consisting of buccal mucosa swabs, plaque and saliva. The ‘pre-malignant’ cohort showed increased abundance of Proteobacteria in the buccal swab and plaque samples and *Neisseria* in the buccal swab and saliva samples. Firmicutes were significantly decreased in all samples, *Streptococcus* significantly decreased in buccal swabs samples and *Veillonella* significantly decreased in plaque samples in the ‘pre-malignant’ cohort, compared with healthy controls.

The key significant differences reported in the literature between the oral microbiome in OLK/OPMDs and normal oral mucosa are summarised in **Table 4**. It is clearly evident that the study populations in Lee’s<sup>99</sup>, Decsi’s<sup>153</sup> and Heng’s<sup>103</sup> studies were diverse. Furthermore, Chen’s<sup>100</sup> study used non-standard terminology to describe their included white patches, while Lee *et al.* also provided no details on the proportions of the different histopathologies within their OPMD cohorts, which included hyperplasia and hyperkeratosis. Some forms of hyperkeratosis, such as frictional keratosis, carry no malignant potential whatsoever and

so cannot be considered ‘precursor’ lesions. Only those conditions identified as OPMDs by the WHO Collaborating Centre for Oral Cancer at the time of each study<sup>1,116</sup> should have been included in these studies.

#### 2.6.4 Systematic review

Shen *et al.*<sup>155</sup> recently carried out the only systematic review on the microbiome of oral epithelial dysplasia (OED), including 10 of the above studies<sup>61,62,79,81,96,99,100,103,153,154</sup> in their review. The authors noted significant heterogeneity of the included studies, particularly with regards to sampling methods, control groups and presence of confounding factors, and so an overall conclusion was not possible. However, despite the study differences, there was consensus that the diversity of the bacterial communities of OED differs from that of normal mucosa and lies between that of healthy tissue and that of OSCC. The most common findings in these studies, as noted previously, were significant increases in the phylum Bacteroidetes and the genus *Fusobacterium* and loss of genus *Streptococcus* in OED compared with healthy controls. Further studies are undoubtedly needed, and it is worth noting that the authors of this review advised caution in using saliva/oral rinse sampling due to the risk of introducing confounding factors that cannot be controlled for.

## 2.7 Models of microbial carcinogenesis

### 2.7.1 Epithelial change

Only one study to date has shown that bacteria may induce malignant transformation in oral mucosa, where *Porphyromonas gingivalis* and *Fusobacterium nucleatum* were shown to promote carcinogenesis in a chemically induced murine model of tongue OSCC<sup>156</sup>. The exact mechanisms of microbial carcinogenesis are still unclear, although recent research suggests a role for bacterial-induced epithelial-mesenchymal transition (EMT).

Abdalla *et al.*<sup>157</sup> demonstrated that the epithelial adhesion molecule E-cadherin, a marker of the epithelial state, is lost early in oral dysplasia and further showed that there is a progressive and linear loss from normal tissue to low-grade dysplasia to high-grade dysplasia. E-cadherin loss is considered an early event in EMT, a biological process whereby an epithelial cell undergoes multiple biochemical changes, altering it from an epithelial to a mesenchymal cell phenotype and therefore enabling enhanced migration, invasion, and resistance of apoptosis. EMT is completed when the underlying basement membrane is degraded, and the cell can migrate away from the epithelial layer<sup>158</sup>. EMT has therefore been associated with the initiation of carcinogenesis and facilitation of cancer spread. Along with the loss of E-cadherin, there is also upregulation of mesenchymal markers such as N-cadherin, vimentin and SNAIL in EMT<sup>158</sup>. Bacteria can induce EMT, with *Helicobacter pylori* a well-known example, inducing EMT via its cytotoxin-associated gene A (Cag-A)<sup>159</sup>. Long-term infection with *Porphyromonas gingivalis* has been associated with initial EMT changes in oral epithelial cells, including decreased E-cadherin expression and upregulation of vimentin<sup>115</sup>. *Porphyromonas gingivalis* has also been shown to induce proliferation of epithelial cells through upregulation of  $\beta$ -catenin and gingipain-dependent proteolytic degradation<sup>160</sup>.

Zhang *et al.*<sup>106</sup> recently showed that the *Fusobacterium nucleatum* adhesin, FadA, can facilitate EMT via binding, downregulation, and transfer of E-cadherin to the cytoplasm and upregulation of N-cadherin, vimentin and SNAIL. They further demonstrated that this pathway involved upregulation of the microRNA MIR4435-2HG. They concluded that *Fusobacterium nucleatum* was able to trigger EMT in both normal and OSCC epithelial cells through FadA and its effects on regulation of the MIR4435-2HG/miR-296-5p/Akt2/SNAIL signalling pathway, thus providing a possible mechanism for the role of *Fusobacterium nucleatum* in carcinogenesis. Rubenstein *et al.*<sup>161</sup> proposed a ‘two-hit’ model to explain the role of *Fusobacterium nucleatum* in colorectal carcinoma; describing a role for *Fusobacterium nucleatum* in the progression, but not initiation, of colorectal cancer. They proposed that an initial somatic mutation initiating carcinogenesis is followed by a second ‘hit’ by *Fusobacterium nucleatum*, stimulating the proliferation of cancer cells via upregulation of Annexin A1, a novel Wnt/ $\beta$ -catenin modulator expressed in proliferating colorectal carcinoma cells, by FadA through E-cadherin.

### 2.7.2 Acetaldehyde

Microbial production of acetaldehyde (ACH) is also likely to play a role in oral carcinogenesis. Alcohol consumption is a well-established risk factor for OSCC, although ethanol itself is not carcinogenic. Rather it is ethanol’s first metabolite, ACH, produced predominantly by the reduction of ethanol by host or microbial alcohol dehydrogenase (ADH)<sup>162</sup>, which has been classed as a Class 1 carcinogen by IARC<sup>163,164</sup>.

ACH is produced from ethanol by mucosal ADH and, at much higher levels, often > 100  $\mu$ M, by oral microbial ADH<sup>165–167</sup>. ACH is a genotoxic carcinogen capable of inducing DNA damage including point mutations, chromosomal alterations and DNA adducts. It can also induce secondary hyperproliferation of epithelium<sup>162,168–171</sup>. Mutagenic levels (>100  $\mu$ M) of ACH can be detected in saliva following ingestion of 0.5 g ethanol per kilogram body weight (approximately three standard drinks/units for a 60 kg person)<sup>172</sup>. Marttila *et al.*<sup>167</sup> demonstrated that bacteria cultured from smokers produces significantly higher levels of ACH, confirming an earlier finding by Homann *et al.*<sup>168</sup>, that smoking, along with poor dental status, significantly increased salivary ACH production to >100  $\mu$ M. In fact, active smoking resulted in an additional 200–400  $\mu$ M peak ACH concentration in saliva for the duration of the smoking<sup>168,172,173</sup>. Homann *et al.* further estimated that smoking 20 cigarettes per day increased salivary ACH production by 50–60%, while smoking and consumption of >40 g/ethanol per day could increase ACH production by 100%. The significant increase in ACH production associated with smoking and heavy alcohol consumption may be a mechanism for the synergistic effect of smoking and alcohol in OSCC causation. Interestingly, alcohol consumption was only associated with increased production of ACH when >40 g of ethanol per day was consumed, demonstrating that increased ACH production can occur in the absence of alcohol consumption. Of note, the amount of ACH produced was significantly reduced after three days of using a chlorhexidine mouthrinse<sup>172</sup>, supporting the role of bacteria in ACH production.

Several studies have identified *Neisseria* as potent microbial producers of ACH<sup>174–176</sup>, with Moritani *et al.* also demonstrating that all species of *Neisseria*, along with *Rothia mucilaginosa*, *Streptococcus mitis* and *Prevotella histicola*, could produce ACH from ethanol<sup>176</sup>. However, in contradiction to this, *Neisseria* spp. have been shown to be

decreased in smokers<sup>175</sup>, while the study of Yokoyama *et al.* on the production of ACH by the salivary microbiome in healthy individuals found an inverse relationship between the abundance of *Neisseria* spp. in saliva and ACH production, showing instead that individuals producing the highest levels of ACH had increased salivary levels of *Rothia mucilaginosa* and *Streptococcus salivarius*<sup>177</sup>. This association of *Rothia mucilaginosa* and ACH production is supported by a more recent study by Amer *et al.* which demonstrated that strains of *Rothia mucilaginosa* isolated from dysplastic OLKs were capable of producing mutagenic levels of ACH in the presence of alcohol<sup>178</sup>. Several studies have also found that certain strains of streptococci, particularly viridans group streptococci, have significant ACH-producing ability<sup>68,176,179</sup>, although as discussed, streptococci are typically associated with a healthy microbiome and reduced in OPMDs and OSCC.

Clearly more studies are needed to investigate the role of bacteria in ACH production and the various factors which influence its production; however, it is apparent that many bacteria within the normal oral microbiome are potent producers of ACH, raising questions about their role in oral carcinogenesis.

### 2.7.3 Inflammation

Inflammation, which has been recognised as a fundamental component of malignant processes<sup>180</sup>, is another potential link between the oral microbiome and OSCC. Chronic infection causes inflammation, and many cancers are known to be preceded by chronic inflammation at the cancer site, as demonstrated by the associations between hepatitis and hepatocellular carcinoma (HCC), inflammatory bowel disease (IBD) and colon cancer and *Helicobacter pylori*-induced gastritis and gastric cancer<sup>180</sup>. The role of infection-associated



inflammation in OSCC is more complex as the oral cavity harbours hundreds of bacteria and fungi as commensals without being in a constant state of inflammation. However, polymicrobial dysbiosis in the oral cavity does lead to inflammation, as demonstrated by periodontal disease, and so may also play a role in carcinogenesis. Bacterial pathogens can stimulate host immune cells, such as macrophages and neutrophils, and oral epithelial cells, via their surface receptors to activate the NF- $\kappa$ B signalling pathway, known to have a role in tumorigenesis through its roles in inflammation and enhanced cell proliferation and survival<sup>181</sup>. *Fusobacterium nucleatum* in particular has been shown to bind and enter host epithelial and endothelial cells and induce NF- $\kappa$ B and IL-6-mediated inflammation via its FadA adhesin molecule<sup>182,183</sup>. Finally, the periodontal pathogens, *Porphyromonas gingivalis*, *Prevotella* spp., and *Fusobacterium* spp. can produce volatile sulphur compounds, such as hydrogen sulphide which is both genotoxic and mutagenic, and methyl mercaptan which has been shown to induce chronic inflammation and cell proliferation<sup>155</sup>.

## 2.8 The role of *Candida*

*Candida* species form part of the normal oral microbiome. *C. albicans*, the most commonly isolated *Candida* species from the oral cavity and present in approximately 50% of the world's population, is an opportunistic polymorphic fungus. It exists in yeast form when commensal and in hyphal form when pathogenic, usually undergoing this morphological change in the presence of local or systemic immunosuppression<sup>184</sup>. The ability to switch from yeast to hyphal forms and grow hyphae is an essential virulence factor of *C. albicans* and a vital characteristic of its pathogenicity<sup>185,186</sup>. Hyphae are involved in adherence and cellular invasion, but hyphal formation alone does not cause cell damage or elicit a host response as the hyphae themselves simply mechanically invade the epithelial structure,

causing disruption to the original alignment but without causing cell damage<sup>185</sup>. Moyes *et al.* have established that it is the novel *C. albicans* toxin, candidalysin, coded by the *ECEL* gene and released from the hyphae, which is responsible for the epithelial cell damage and stimulation of the host response seen in candidal infections<sup>185</sup>. Candidalysin damages epithelial cells, leading to release of pro-inflammatory interleukins and activation of immune response pathways involved in cell growth and proliferation, survival, angiogenesis, and differentiation. These in turn activate several signalling pathways leading to activation of granulocyte-macrophage colony-stimulating factor (GM-CSF), an essential molecule in epithelial damage and in carcinogenesis<sup>186</sup>. Confirmation of this is provided by *C. albicans* strains with the *ECEL* gene deleted being less virulent<sup>185</sup>.

Further evidence of *C. albicans*' role in oncogenesis was reported recently by Vadovics *et al.*<sup>187</sup> who demonstrated that, *in vitro*, *C. albicans* increased migration and matrix metalloproteinase activity of OSCC cells and upregulated multiple genes associated with OSCC invasion while *in vivo*, using a mouse xenograft OSCC model, *C. albicans* influenced EMT by reducing expression of E-cadherin and increasing expression of vimentin and p63. *C. albicans* was also shown to upregulate several oncogenes in a non-malignant epithelial cell line. Finally, using another mouse model and the chemical carcinogen 4-nitroquinoline 1-oxide, the authors demonstrated that the addition of *C. albicans* was associated with the presence of significantly more dysplasia and higher grades of dysplasia.

*Candida* is also known to be associated with OLK, with numerous studies finding significantly more *Candida* in rinse and swab samples in patients with OLK compared with controls<sup>188–191</sup>. White patches that are chronically colonised by *Candida* and demonstrate hyphal infiltration of the epithelium on histological examination are termed chronic

hyperplastic candidiasis (CHC). They are typically found at the buccal commissures, particularly in smokers, and until recently were considered an OPMD<sup>116,192</sup>. CHC was removed from the most recent list of OPMDs as the evidence for malignant transformation of CHC was deemed unconvincing<sup>116,193</sup>. The most commonly isolated candidal species from CHC is *C. albicans*, although other species such as *C. dubliniensis*, *C. tropicalis*, *C. pintolopesii* and *C. glabrata* have also been identified<sup>189,192</sup>. The level of *Candida* colonisation of these leukoplakias has been associated with the degree of dysplasia found on biopsy, with *C. albicans* in particular being associated with severe dysplasia<sup>194–196</sup>. Candidal infection has also been associated with worsening of dysplasia over time with Barrett *et al.* showing that almost three times more fungal-infected dysplastic lesions demonstrated more severe degrees of dysplasia in subsequent biopsies than non-infected dysplastic lesions (21.9% compared with 7.6%)<sup>197</sup>. The fungi were not speciated but were presumed to be *Candida*.

Malignant transformation rates of CHC have been reported to be as high as 13%<sup>198</sup> suggesting that *Candida* may play a role in the development of OSCC. Significant levels of *Candida* colonisation have also been reported from swabs, rinses, and biopsies of OSCC<sup>77,86,189–191,195,196,199</sup> with Alnuaimi *et al.* demonstrating that even after adjusting for confounders such as alcohol, tobacco, and periodontal disease, *Candida* remained significantly associated with OSCC<sup>199</sup>. They further found that the presence of *C. albicans* and frequent alcohol consumption had a more than additive effect on the risk association. Alnuaimi *et al.* in a later study<sup>174</sup> also demonstrated significant associations between various candidal virulence factors and OSCC. They found that *Candida* isolated from patients with OSCC had significantly higher biofilm mass, biofilm metabolic activity and phospholipase and proteinase activity than their control group of non-oral cancers<sup>174</sup>.

*C. albicans* and other *Candida* species are known potent producers of ACH, with multiple studies demonstrating the yeasts' ability to produce mutagenic levels of ACH from ethanol and glucose *in vitro*. Greater production of ACH has also been associated with increased yeast density, particularly in smokers, and yeasts from smokers and alcohol consumers have been shown to produce more ACH than controls<sup>174,200–202</sup>. Marttila *et al.* investigated ACH production in OSCC and found that *Candida* species were cultured significantly more frequently from samples producing mutagenic levels (>100 µM) of ACH than from those that did not produce such high levels of ACH, although they noted that *Candida* were not isolated from all samples producing >100 µM of ACH<sup>167</sup>. Finally, the study of Alnuaimi *et al.* comparing candidal virulence factors in OSCC and non-oral cancer patients demonstrated that *Candida* producing high levels of ACH (defined as >40 µM) were significantly more prevalent in oral cancer than non-oral cancer patients<sup>174</sup>.

Perera *et al.*<sup>86</sup> were the first group to use NGS to identify ribosomal ITS2 to investigate the oral mycobiome in OSCC. Deep tissue samples were taken from 22 buccal mucosa or tongue OSCCs in Sri Lankan males and from 25 fibroepithelial polyps (FEP) as controls. There were significant differences between the cases and controls with regards to risk factors, namely significantly more use of areca nut, alcohol and tobacco and significantly poorer oral hygiene and more severe periodontal disease amongst cases. They found that *Candida*, *Hannaella*, and *Gibberella* were significantly more abundant in OSCC, while *Trametes* and *Alternaria* were strongly associated with FEP. At the species level, *C. albicans* and *C. etchellii* were significantly enriched in OSCC. *C. albicans* was identified in all samples (cases and controls), but significantly, the average relative abundance in OSCC was twice that in the controls (61.2% vs. 29.6%). They surmised that OSCC displays

a dysbiotic mycobiome characterised by lower species diversity and increased relative abundance of *C. albicans*<sup>86</sup>.

Although as mentioned, *Candida* has been associated with OLK, I chose not to analyse *Candida* colonisation in this study and to focus my investigations on the bacterial colonisation of OLK.

## 2.9 Conclusion

It is clear that there are changes in the microbiome and mycobiome associated with OSCC. There is also now growing evidence that these shifts can occur early, before the development of OSCC, and are associated with OPMDs, with several authors demonstrating trends of increasing abundance of anaerobic Gram-negative bacteria, including *Fusobacteria*, *Prevotella*, and *Alloprevotella*<sup>96,103</sup> and decreasing abundance of Firmicutes and *Streptococcus* from normal controls to OLK to OSCC<sup>81,94,103</sup>. However, what has not been considered to date is the role that the changing clinical appearance from normal mucosa to white patch (OLK) to the indurated ulcer with raised rolled borders typical of OSCC, plays in the shift in microbial communities observed.

It is also evident that external factors such as smoking, and alcohol alter the composition of the oral microbiome both directly and indirectly through their association with poor oral health and periodontal disease. These influences contribute to a more inflammatory and disease-state microbiome, which could create an environment for malignant transformation, possibly via induction of EMT, increased ACH production or triggering of chronic inflammation. These microbiome changes could be targets for diagnostic or

chemotherapeutic interventions, however large longitudinal studies are first required to definitively establish these links.

### 3 Aims and Objectives

The primary hypothesis of this study was that the microbial colonisation of OLK differs to that of contralateral normal sites.

The primary aim of the study therefore was to:

1. Investigate if OLK was associated with changes in the microbiome and whether these changes, if present, differed with differing degrees of dysplasia.

Secondary aims and objectives included to:

1. Examine if any clinical or patient factors are associated with the clinical characteristics of OLK.
2. Establish what clinical and patient factors affect the degree of dysplasia of OLK found on histological analysis.
3. Establish how other factors such as smoking and oral site affect the mucosal microbiome.
4. Examine if any changes in the microbiome associated with OLK can be used to discriminate sites at lower or higher risk of malignant transformation.
5. Investigate the predictive ability of the bacterial microbiome to discriminate high-risk from low-risk OLK.

## 4 Materials and Methods

### 4.1 Sample collection

Ethical approval was granted by the Joint Hospitals' Research Ethics Committee, Tallaght University Hospital, Dublin (Ref: 2017–11-Chairman's actions). Prior to sample collection, two power calculations were carried out; firstly, to estimate the number of samples needed to examine the association of individual risk factors with the presence and degree of dysplasia (four categories) and secondly to estimate the number of samples needed to investigate the association of bacterial clusters (identified in a previous study)<sup>61</sup> with the presence and degree of dysplasia. Using a simulated ordinal logistic regression based on at least 10 events per variable<sup>203</sup> it was estimated that with power set at 0.9 and an effect size of 0.3, sample sizes of 197 and 212 respectively would be needed. It was not possible to carry out a power calculation to determine the number of samples needed to examine the association of individual bacterial taxa with the presence and degree of dysplasia as no such analysis had been carried out prior to this study.

Following written informed consent (**Appendix C Patient information leaflet, Appendix D Patient consent form**), oral mucosal swabs were collected from patients attending Oral Medicine and Oral Mucosal Dysplasia clinics in the Dublin Dental University Hospital (DDUH) between June 2014 and December 2021 with white patches deemed by a consultant in oral medicine (Dr Sheila Galvin or Prof Claire Healy) to be clinically consistent with OLK. Swabs were taken of white patch(es) and contralateral normal site(s), if available, using sterile polyurethane foam swabs (BD BBL CultureSwab EZ, Copan Italia SpA, Brescia, Italy, distributed by Becton, Dickinson and Company, MD, USA) by gently rubbing the surface of the white patch/mucosa 15 times with rotation of the swab by Dr Sheila



Galvin, Prof Claire Healy or Ms Sviatlana Anishchuk (research dental hygienist). In the case of large white patches, swabs were collected from the proposed incisional biopsy site (or adjacent to the incisional biopsy site if the swab was collected after the biopsy was taken) which was marked on a clinical photograph. Samples were recovered at least one hour after eating, drinking, or any oral hygiene measure. If the contralateral site was clinically involved, another similarly keratinised or non-keratinised site was selected as the control site. Patients had either had, or were awaiting, a clinically indicated biopsy of the OLK (i.e., swab sampling could take place before or after biopsy). It has been shown that the oral microbiome is stable over time<sup>24,204</sup> and so participants were swabbed at a single time point. The median time between swab and biopsy sampling was one month with more than half (51.1%) of swab samples carried out within one month of the biopsy, and a further 14.6% carried out within one to three months of biopsy. Comparisons of the clinical appearance of the OLKs at the time of biopsy and at swab sampling were made with the aid of photographs to ensure there had been no significant changes in the OLK. In cases where significant changes in clinical appearance (e.g., in size or homogeneity) were noted at the time of biopsy, a repeat swab was taken and this one used for microbiome analysis, while the original swab was discarded. Data on patient demographics, lesional clinical features, smoking status (current [smoked within past year], former [quit smoking more than one year ago], never), alcohol consumption (units/week), denture wear, tooth number<sup>53</sup> and oral hygiene practices (toothbrushing frequency and mouthwash use) were recorded (**Appendix E Data recording form**). The presence and degree of dysplasia were established by histopathological analysis of biopsy specimens by consultant Oral and Maxillofacial Pathologists, Dr Esther O'Regan, Dr Mary Toner, Dr Róisín O'Connor, Dr Bijal Shah and Dr Peter Moloney in St James's Hospital, Dublin. Dysplasia was graded according to the three-tier WHO 2017 classification (mild, moderate, severe dysplasia/carcinoma *in situ*)<sup>114</sup>

(**Figure 4**, **Figure 5**, **Figure 6**). The final grading of all borderline samples (i.e., mild to moderate dysplasia or moderate to severe dysplasia) was decided with the consensus of at least two of these pathologists or following discussion at a monthly Clinicopathological Case Conference. Where more than one degree of dysplasia was given on the biopsy report, the higher grade was recorded.

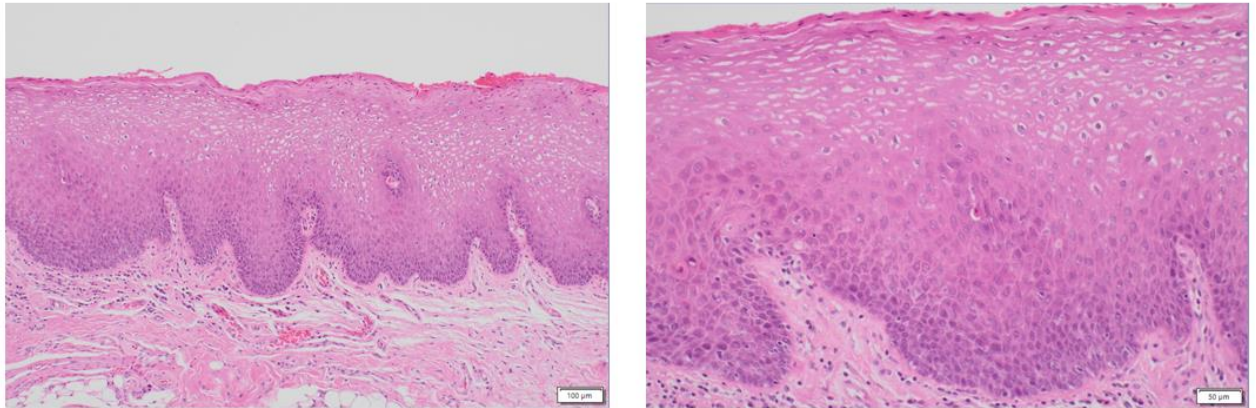


**Figure 4.** Example of H&E stains of mild dysplasia at 10x and 100x magnification. Photos courtesy of Dr Róisín O'Connor, St. James's Hospital, Dublin.

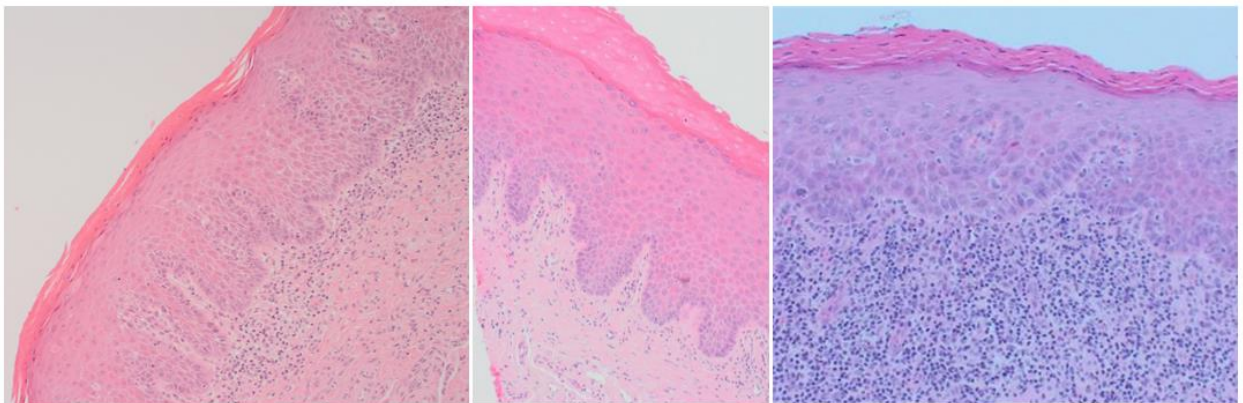


**Figure 5.** Examples of H&E stains of moderate dysplasia at 20x magnification. Photos courtesy of Dr Esther O'Regan, St. James's Hospital, Dublin.

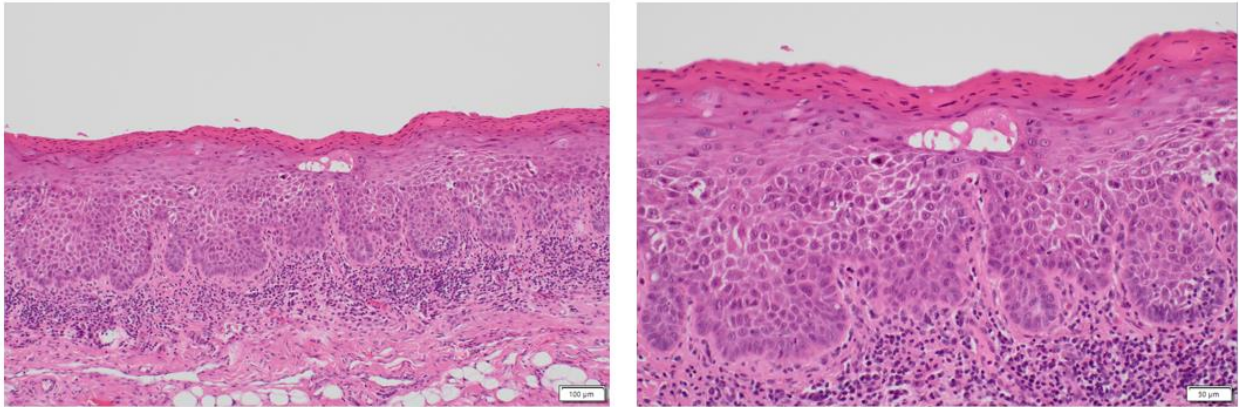


**Figure 6.** Example of H&E stains of severe dysplasia at 10x and 100x magnification. Photos courtesy of Dr Róisín O'Connor, St. James's Hospital, Dublin.

Oral hygiene was measured using the Greene and Vermillion Simplified Oral Hygiene Index (OHI-S) at the time of swab sampling by the person carrying out the swab sampling<sup>205</sup> (**Appendix F OHI-S**). Patients with diabetes mellitus, Crohn's disease, ulcerative colitis, previous gastrointestinal tract malignancy, dry mouth, or current upper respiratory tract infection were excluded, along with those who within the previous two months had taken systemic antibiotics or immunosuppressants, including corticosteroids, or who had used topical corticosteroids intra-orally. Patients were also excluded if there was no available histopathological report related to the OLK. Healthy controls were sourced from dentally and periodontally stable patients attending for routine reviews in periodontal and oral hygiene clinics in the DDUH and were subject to the same exclusion criteria. After obtaining written informed consent, control subjects were examined by a clinician to ensure they had no visible mucosal abnormality. Single swab samples were then collected from two of the following mucosal sites: buccal mucosa, lateral border of tongue, floor of mouth, soft and hard palate, from each control subject. All swabs were transported to the laboratory within two hours of collection and resuspended in 500 µl of TSE buffer (10mM Tris-HCL

[pH 7.8], 1mM EDTA, 100mM NaCl) in a screw-top tube. They were then stored at -80°C until DNA extraction was carried out.

## 4.2 DNA extraction

DNA extraction was carried out using the MasterPure DNA/RNA Purification Kit (Epicentre Biotechnologies, Madison, WI, United States). The manufacturer's protocol was supplemented with an additional incubation step with Ready-Lyse lysozyme (Epicentre Biotechnologies) and a bead disruption step after Proteinase K treatment as described by Amer *et al*<sup>61</sup>. The samples were first defrosted at room temperature. All reagents were certified DNA-free, and all consumables were either certified DNA-free or sterilised prior to use. 500 U of Ready-lyse lysozyme (Epicentre Biotechnologies, Madison, WI, United States) was added to each sample and the samples left to incubate at room temperature for 15 minutes with occasional mixing. Next, Proteinase K was diluted with 2x T and C lysis buffer (1 µl of Proteinase K to 150 µl of T and C lysis buffer) and 150 µl of this solution was added to each tube and the samples incubated at 65°C for 15 minutes. 250 µl of sterile glass beads (100 µm diameter) was then added to each tube and the samples disrupted in a Minibead beater (Biospec Products, Bartlesville, OK, United States) for 30 seconds. The samples were spun in a centrifuge for two minutes to remove the foam. The top layer was then removed and transferred to a new 1.5 ml microcentrifuge (Eppendorf) tube, 1 µl of RNase was added and the samples incubated at 37°C for 30 minutes. After this, the samples were placed on ice, 175 µl of MPC protein precipitate solution was added and the samples vortexed, then centrifuged at 4°C for 10 minutes after which the supernatant was removed and added to a new microcentrifuge tube. 500 µl of isopropanol was added, the samples inverted several times, and then centrifuged again at 4°C for 10 minutes. Next, the liquid

was removed, the resultant DNA pellets dried, resuspended in 35 µl Tris-EDTA buffer and then stored at -80°C. DNA concentrations were estimated using a NanoDrop™ 2000 Spectrophotometer (Thermo Scientific™, Waltham, MA, United States) prior to freezing.

### 4.3 DNA sequencing

Amplification of the V1-V3 region of the 16S rRNA gene was carried out using the primers 27F-YM and 519 R (27F-YM: 5'AGAGTTTGATYMTGGCTCAG; 519 R: 5'GWATTACCGCGGCKGCTG)<sup>206</sup> and was performed in duplicate using separate template dilutions (1:1 and 1:10). The V1–V3 region of the 16S rRNA gene (**Figure 7**) was chosen for analysis as it has been shown that this region is more divergent and so can provide more phylogenetic resolution, particularly for *Streptococcus* spp., than other regions<sup>207</sup>. It also provides greater phylotype evenness and richness for oral bacteria than the previously favoured V3-V4 region, and so gives a better assessment of population diversity and community<sup>208,209</sup>. Amplification and sequencing were carried out by Integrated Microbiome Resources (IMR; Dalhousie, Halifax, Canada) according to their standardised in-house protocols described by Comeau *et al.*<sup>210</sup> and detailed here <https://www.protocols.io/workspaces/integrated-microbiome-resource-imr/publications>.

The PCR products were verified visually on a high-throughput Hamilton Nimbus Select robot using Coastal Genomics Analytical Gels, then pooled, cleaned up and normalised using the high-throughput Charm Biotech Just-a-Plate 96-well Normalisation Kit. The pooled library was quantified fluorometrically before sequencing. Paired end sequencing was performed using the Illumina 600 cycle MiSeq V3 kit to generate 300 bp paired end reads.



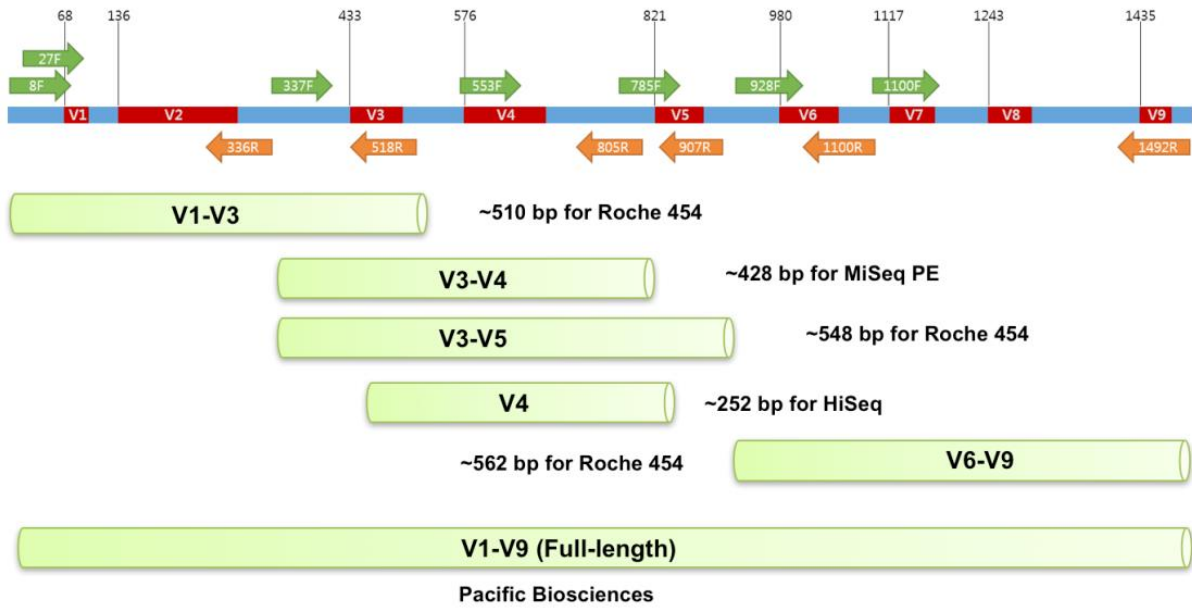


**Figure 7.** Graphic of the 16S rRNA gene with variable regions highlighted (Image @ Pacific Biosciences)

## 4.4 Data analysis

The bacterial 16S rRNA sequence data was imported to RStudio. Dada2<sup>109</sup> was used to filter and process the reads with the following parameters: maxN=0, maxEE=c(2,5) and truncQ=2<sup>111</sup>. Forward and reverse reads were first shortened to remove low quality bases; forward reads being shortened to 290bp and reverse to 270bp. Sequencing errors were then identified and removed using the Dada2 algorithm, after which paired reads were merged and chimeras removed following their identification with the removeBimeraDenovo command. Finally, any samples with less than 3000 reads remaining were excluded, leaving a total of 548 samples, 428 samples from 187 patients, consisting of 226 disease and 202 contralateral normal sites, and 120 samples from 61 healthy controls available for analysis. A table of abundances of amplicon sequence variants (ASVs) was created and taxonomy was assigned using the expanded Human Oral Microbiome Database (eHOMD)

classifications (eHOMD 16S rRNA RefSeq version 15.1)<sup>212</sup>. These tables, along with a sample metadata file, were imported to Phyloseq<sup>213</sup> (<https://bioconductor.org/packages/release/bioc/html/phyloseq.html>) and null values excluded.

The packages Phyloseq<sup>213</sup> and MicrobiotaProcess<sup>214</sup> were used to calculate alpha diversity indices for each sample and for comparisons of beta diversity (community structure). Further visualisations (biplots) and analysis of community structure, including PERMANOVA (Permutational Multivariate Analysis of Variance) testing, were conducted using the package MicroViz in RStudio<sup>215</sup>. Matrices of the relative abundances were also visualised using Principal Component Analysis (PCA). Bar plots demonstrating the most abundant phyla and genera associated with OLK and CLN sites were generated for qualitative visualisation using MicrobiotaProcess.

Multivariate analysis was conducted with MaAsLin2<sup>216</sup> using relative abundance data and the default linear models (LM) setting with log transformation of the data. ‘Healthy’ variables were set as the references (e.g., never smokers, 0 alcohol units/week, brushing >1 per day, good oral hygiene, no denture wear, mouthwash use, <5 missing teeth<sup>53</sup>). MaAsLin2 generates tabular outputs detailing a coefficient of variance from the reference, adjusted p values (q-values; Benjamini-Hochberg<sup>217</sup> false discovery rate (FDR)) and heatmaps showing the 50 most significant values (default padj <0.25) with the effect size calculated as  $-\log(\text{q-value}) * (\text{coefficient})$ . p values of <0.05 and adjusted p values of <0.1 were accepted as significant. DESeq2<sup>218</sup> in RStudio was also used to analyse the species level count data of various disease versus CLN sites as a second measure to ensure consistency of results.

Predictive modelling was carried out using the SIAMCAT<sup>219</sup> (Statistical Inference of Associations between Microbial Communities And host phenoTypes) package in RStudio using both LASSO and random forest (RF) models. SIAMCAT carries out statistical association testing and statistical modelling, and plots and analyses receiver operator characteristic (ROC) and precision-recall (PRC) curves. Prior to modelling, features were normalised using log transformation of the data and z-score standardisation. The significance level was set at 0.1 and the default linear model setting used. The data were split into five cross-validation folds with four fifths of the data used to train the model, and one fifth for testing the model, which was based on the best predictions from the training dataset. Modelling was performed with both RF and LASSO methods and barplots of feature weights and robustness, heatmaps displaying metagenomic features and metadata z-scores (i.e., number of standard deviations from the mean) and boxplots showing feature proportional weights, generated. To further refine the results, the z-scores of the top 50 microbial features generated by the RF modelling in SIAMCAT were used for additional predictive modelling analysis both with and without clinical and patient factors using the R packages rpart<sup>220</sup> and randomForest<sup>221</sup>. Up to 10,000 iterations of Classification trees using random subsets of 90% of cases were carried out to estimate sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and classification accuracy with corresponding confidence intervals (CI), along with ROC curves and Gini indices (measures of the impurity of the dataset). RFs were then run with 500 trees with 25 (for microbial data) or 30 (for combined microbial and clinical/patient data) random variables for each tree.

The clinical and patient demographic data was also analysed in RStudio. Pearson's Chi-squared, Fisher's exact and Fisher's exact with simulated p values based on 2000 replicates



tests were used as appropriate and a table of counts and percentages generated. Logistic regression was initially carried out with single variables to determine crude odds ratios and CIs using the R packages `epiR`<sup>222</sup> and `car`<sup>223</sup>. To generate adjusted odds ratios and CIs, variable selection was based on both Akaike information criterion (AIC) and Bayesian information criterion (BIC) and on clinical factors to determine the best model. Variance inflation factor (VIF) was used to test for multicollinearity. Multinomial regression was performed using the R package `nnet`<sup>224</sup>, variable selection again being based on clinical factors and AIC. Finally predictive modelling was carried out on the clinical and patient demographic data using the R packages `rpart`<sup>220</sup> and `randomForest`<sup>221</sup>. Text variables were converted to factors and minimum values of 20 observations per node before a new split, and 10 observations per terminal node, along with a maximum depth of five were set. Up to 10,000 iterations of Classification trees using random subsets of 90% of samples for each iteration were carried out. Sensitivities, specificities, PPV, NPV and model accuracies with CIs, along with ROC curves and Gini indices, were generated for each Classification tree.

## 5 Demographics of patient cohort and clinical characteristics and associations of OLK

### 5.1 Introduction

The aims of this chapter were to examine if any clinical or patient factors were associated with the clinical characteristics of OLK and to establish which clinical and patient factors, if any, affected the degree of dysplasia of OLK found on histological analysis. Any such associations are clearly important from a clinical perspective with regards to treating, monitoring and risk factor management of patients with OLK. However, it was also important to establish any clinical/patient factor associations with OLK which could impact on microbial colonisation and would therefore need to be adjusted for in later analyses.

### 5.2 Methods

A total of 268 OLKs from 203 patients were sampled as described in **section 4.1**. Prior to DNA analysis, 33 samples were excluded post-collection for reasons including resolution prior to biopsy (18), oral lichen planus on biopsy (5), not meeting inclusion criteria (5) and sampling error (5). Two hundred and thirty five disease site samples from 192 patients remained. Ten of these OLKs were found to be OSCC on biopsy and so were excluded from the analysis of characteristics, associations and risk factors of OLK. Four of the patients with OSCC also had a separate concurrent OLK and so these patients and OLKs were included in the analysis leaving a total of 186 patients with 225 OLK (**Figure 8, Figure 9, Figure 10**). Thirty one patients had more than one OLK.

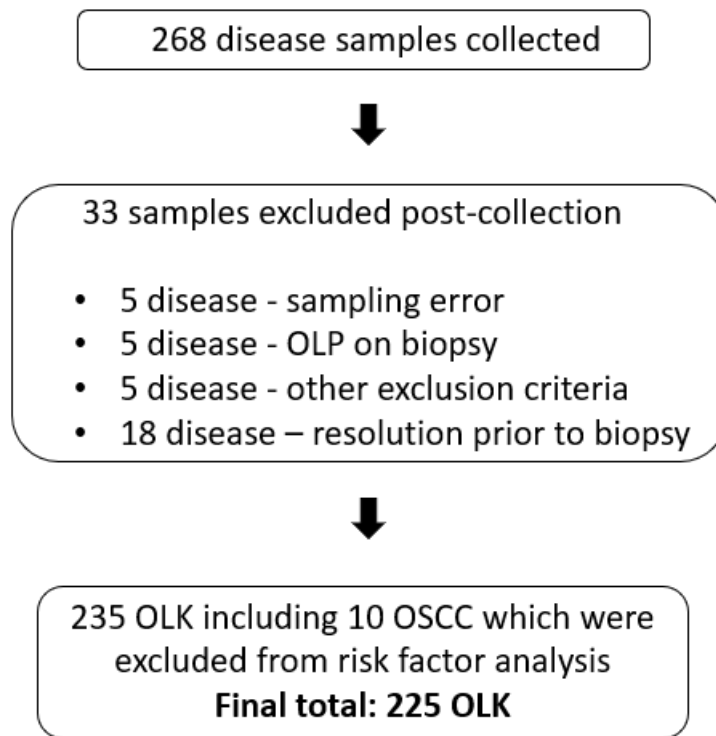


**Figure 8.** Flowchart of patient cohort

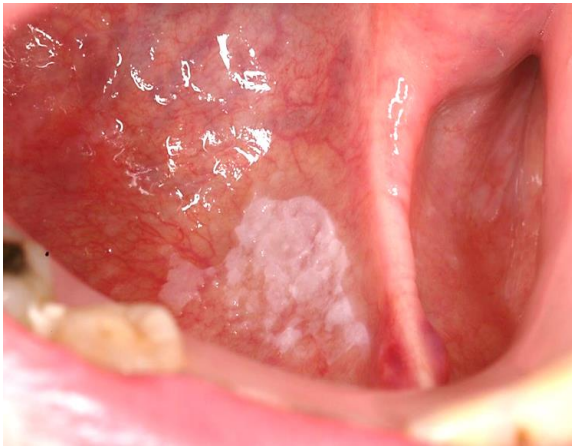
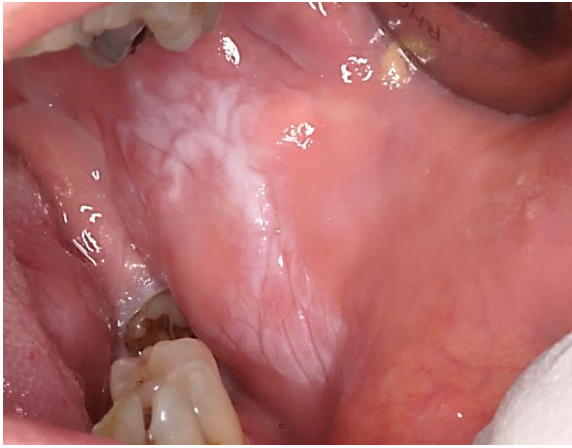


**Figure 9:** Selection of included homogenous OLKs. Photographs courtesy of Dr. S Galvin



**Figure 10.** Selection of included non-homogenous (including ulcerated) OLKs. Photos courtest of Dr. S Galvin.

## 5.3 Results

### 5.3.1 Smoking and alcohol consumption

The sample population comprised 186 patients (98 males and 88 females) with a male to female ratio of 1.1:1. The average age was 58.3 years (median 58 years) with a range of 23 to 89 years. Present or past tobacco use was reported by 81.2% (n=151) of patients, (current smokers n=98 (52.7%), and former smokers, defined as having stopped smoking more than one year previously, n=53 (28.5%)), while 18.8% (n=35) had never smoked (**Table 5**).

**Table 5.** Demographic data of patient cohort

| <b>Demographic</b>                          | <b>Patients</b> |
|---------------------------------------------|-----------------|
| Total                                       | 186             |
|                                             |                 |
| <b>Gender</b>                               |                 |
| Males                                       | 98              |
| Females                                     | 88              |
|                                             |                 |
| <b>Age range</b>                            |                 |
| <40                                         | 18              |
| 40-60                                       | 87              |
| >60                                         | 81              |
|                                             |                 |
| <b>Smoking status</b>                       |                 |
| Current smoker                              | 98              |
| Former smoker                               | 53              |
| Never smoker                                | 35              |
|                                             |                 |
| <b>Alcohol units/week</b>                   |                 |
| 0                                           | 39              |
| 1 to 10                                     | 96              |
| 11 to 20                                    | 25              |
| More than 20                                | 26              |
|                                             |                 |
| <b>Alcohol &gt;20 units/week and smoker</b> |                 |
| Yes                                         | 27              |
| No                                          | 159             |
|                                             |                 |
| <b>Missing teeth</b>                        |                 |
| <5                                          | 25              |
| 5 to 15                                     | 76              |
| More than 15                                | 35              |
| Not recorded                                | 50              |
|                                             |                 |
| <b>OHI Score</b>                            |                 |
| Good (OHI-S < 0.5)                          | 39              |
| Fair (OHI-S 0.5-1)                          | 60              |
| Poor (OHI-S > 1)                            | 38              |
| Edentulous                                  | 7               |
| Not recorded                                | 42              |
|                                             |                 |
| <b>Brushing times/day</b>                   |                 |
| <1                                          | 13              |
| 1                                           | 27              |
| >1                                          | 125             |
| Not recorded                                | 21              |
|                                             |                 |
| <b>Denture</b>                              |                 |
| Complete (Full upper and full lower)        | 8               |
| Partial                                     | 36              |
| None                                        | 142             |
|                                             |                 |
| <b>Mouthwash use</b>                        |                 |
| Yes                                         | 87              |
| No                                          | 99              |

There was no difference between the smoking habits of females and males ( $p=0.43$ ). Smoking appeared to be associated with age ( $p=0.00003$ ) (**Table 6**) although the association was not significant on an adjusted multinomial regression model.

**Table 6.** Association of age with smoking habit

| Age   | Current smokers | % Current smokers | Former smokers | % Former smokers | Never smokers | % Never smokers | Total | p value                            |
|-------|-----------------|-------------------|----------------|------------------|---------------|-----------------|-------|------------------------------------|
| <40   | 11              | 61.1              | 4              | 22.2             | 3             | 16.7            | 18    |                                    |
| 40-60 | 61              | 70.1              | 15             | 17.2             | 11            | 12.6            | 87    |                                    |
| >60   | 26              | 32.1              | 34             | 42               | 21            | 25.9            | 81    | <b>0.00003</b><br>(Fisher's exact) |

However, smoking was significantly associated with the site of the OLK ( $p=0.001$ ) with 35.7% of OLKs in current smokers being located on the floor of the mouth or ventral tongue (**Table 7**) (see **Table 13** for distribution of OLK sites). Multinomial regression confirmed this association, with OLKs in current smokers showing an increased relative risk ratio (RRR) of 3.89 (95%CI: 1.13 - 13.42) of being located on the floor of mouth/ventral tongue compared with the buccal mucosa. After adjusting for age, gender, oral hygiene, alcohol consumption and missing teeth the adjusted RRR for OLKs in current smokers being sited on the floor of mouth/ventral tongue was 6.71 (adj 95%CI: 1.3 – 34.55). Former smokers also presented a significantly increased risk of floor of mouth/ventral tongue OLKs compared with buccal mucosa OLKs in the adjusted model (adj RRR 9.31; adj 95%CI: 1.63 – 53.19) (**Table 8**).



**Table 7.** Association of smoking status and site of OLK (% in brackets) (Alv/Gin=alveolus/gingiva, BM=buccal mucosa, DT=dorsum tongue, FOM/VT=floor of mouth/ventral tongue, HP=hard palate, LBT=lateral border tongue, SP=soft palate)

| Smoking status | Alv/Gin (%) | BM (%)    | DT (%)  | FOM/VT (%) | HP (%)  | LBT (%)   | SP (%)  | Total | p value                                                 |
|----------------|-------------|-----------|---------|------------|---------|-----------|---------|-------|---------------------------------------------------------|
| Current        | 12 (10.4)   | 29 (25.2) | 2 (1.7) | 41 (35.7)  | 1 (0.9) | 15 (13)   | 15 (13) | 115   |                                                         |
| Former         | 9 (13.4)    | 13 (19.4) | 5 (7.5) | 19 (26.9)  | 5 (7.5) | 14 (20.9) | 3 (4.5) | 67    |                                                         |
| Never          | 10 (23.3)   | 11 (25.6) | 1 (2.3) | 5 (9.3)    | 3 (7)   | 14 (30.2) | 1 (2.3) | 43    | <b>0.001</b><br>(Fisher's exact with simulated p value) |

**Table 8.** Results of multinomial regression investigating the association of smoking with site of OLK. Reference is never smokers and BM (buccal mucosa). Significant results in bold. CIs in brackets. (\*DT combined with LBT due to small number (n=8) of DT OLKs, HP combined with SP due to small number (n=9) of HP OLKs). The adjusted model was adjusted for age, gender, oral hygiene, alcohol consumption and missing teeth.

| Site    | Current smoker RRR (CI)    | Adjusted model Current smoker RRR (CI) | Former smoker RRR (CI) | Adjusted model former smoker RRR (CI) |
|---------|----------------------------|----------------------------------------|------------------------|---------------------------------------|
| BM      |                            |                                        |                        |                                       |
| Alv/Gin | 0.46 (0.15 – 1.35)         | 1.34 (0.17 – 10.22)                    | 0.76 (0.23 – 2.54)     | 5.22 (0.57 – 47.42)                   |
| DT/LBT* | 0.46 (0.17 – 1.24)         | 0.28 (0.06 – 1.28)                     | 1.15 (0.4 – 3.3)       | 2.37 (0.53 – 10.7)                    |
| FOM/VT  | <b>3.89 (1.13 - 13.32)</b> | <b>6.71 (1.3 – 34.55)</b>              | 3.81 (0.99 – 14.67)    | <b>9.31 (1.63 – 53.19)</b>            |
| HP/SP*  | 1.52 (0.41 – 5.55)         | 1.54 (0.25 – 9.54)                     | 1.69 (0.4 – 7.17)      | 2.85 (0.39 – 21.06)                   |

BM=buccal mucosa, Alv/Gin=Alveolus/gingiva, DT/LBT=dorsum tongue/lateral border tongue, FOM/VT=floor of mouth/ventral tongue, HP/SP=hard palate/soft palate

There were no significant differences in degree of dysplasia ( $p=0.21$ ), clinical appearance of the OLK ( $p=0.7$ ), brushing habits ( $p=0.87$ ), missing teeth ( $p=0.31$ ), denture wear ( $p=0.32$ ) or mouthwash use ( $p=0.42$ ) between the different smoking cohorts (data not shown).

Seventy-nine percent of the cohort consumed alcohol (range 1 to 140 units, mean 10.7 units, median 5 units per week) with 27.4% consuming more than 10 units per week, and 14% consuming more than 20 units per week. Alcohol consumption was significantly associated with gender, with significantly more males both consuming alcohol ( $p=0.002$ ), and consuming higher volumes of alcohol per week ( $p=0.005$ ), and females making up almost

two thirds (64.1%) of the non-consumers. Thirty seven point eight percent of males and 15.9% of females consumed more than 10 units of alcohol per week, while 21.4% of males but only 5.7% of females consumed more than 20 units of alcohol per week (**Table 9**). Regression analysis confirmed this association and showed that males were 3.17 times more likely to consume 11 to 20 units of alcohol per week, and 7.5 times more likely to consume more than 20 units of alcohol per week, than females. An adjusted model found that males presented a more than 15 times increased odds of heavy alcohol consumption (more than 20 units per week) than females. (**Table 10**).

**Table 9.** Association of alcohol consumption with gender

| Alcohol (units/week) | Male | % Male | Female | % Female | Total | p value                       |
|----------------------|------|--------|--------|----------|-------|-------------------------------|
| 0                    | 14   | 35.9   | 25     | 64.1     | 39    |                               |
| 1 to 10              | 47   | 48.5   | 49     | 51.5     | 96    |                               |
| 11 to 20             | 16   | 61.5   | 9      | 38.5     | 25    |                               |
| >20                  | 21   | 77.8   | 5      | 22.2     | 26    | <b>0.002</b><br>(Chi-squared) |
| Total consumers      | 84   |        | 63     |          |       |                               |

**Table 10.** Logistic regression analysis showing the odds of higher alcohol consumption in males. The reference category was female gender and 0 units of alcohol per week. Significant results in bold. The adjusted model was adjusted for smoking habit, age, oral hygiene, missing teeth and degree of dysplasia of OLK.

| Alcohol u/week | Odds ratio (OR) | p value       | 95% Confidence interval (CI) | Adjusted OR  | Adjusted p value | Adjusted 95% CI     |
|----------------|-----------------|---------------|------------------------------|--------------|------------------|---------------------|
| 0              |                 |               |                              |              |                  |                     |
| 1 to 10        | 1.71            | 0.17          | 0.8 – 3.76                   | 1.38         | 0.56             | 0.47 – 4.25         |
| 11 to 20       | <b>3.17</b>     | <b>0.03</b>   | <b>1.14 – 9.35</b>           | 2.95         | 0.16             | 0.68 – 13.95        |
| > 20           | <b>7.5</b>      | <b>0.0008</b> | <b>2.46 – 26.61</b>          | <b>15.66</b> | <b>0.005</b>     | <b>2.69 – 138.6</b> |

Alcohol was also unsurprisingly significantly associated with smoking ( $p=0.006$ ), with 34.7% of current smokers consuming more than 10 units of alcohol per week compared with 14.3% of never smokers. Twelve point two percent of current smokers did not consume

alcohol, while 45.7% of never smokers also did not consume alcohol (**Table 11**). This association was confirmed by multinomial regression which showed that current smokers had 4.95, 4.67 and 26.67 times increased RRRs of consuming one to 10, 11 to 20 and more than 20 units of alcohol per week respectively than never smokers. These associations were still present, and the risks higher, after adjusting for age, gender, oral hygiene, and missing teeth (best fit model, AIC: 228.7), although the confidence intervals were wider. Former smokers also showed an increased RR of consuming one to 10 units of alcohol per week than never smokers, although this association was not significant after adjustment (**Table 12**). Interestingly on this model, current smokers were also significantly more likely to be missing five to 15 teeth than never smokers (RR 6.73; 95%CI: 1.57-28.84, results not shown).

**Table 11.** Association of smoking and alcohol consumption

| Alcohol units/week | Current smokers | % Current smokers | Former smokers | % Former smokers | Never smokers | % Never smokers | Total | p value                                                     |
|--------------------|-----------------|-------------------|----------------|------------------|---------------|-----------------|-------|-------------------------------------------------------------|
| <b>0</b>           | 12              | 30.8              | 11             | 28.2             | 16            | 41              | 39    |                                                             |
| <b>1 to 10</b>     | 52              | 54.2              | 30             | 31.2             | 14            | 14.6            | 96    |                                                             |
| <b>11 to 20</b>    | 14              | 56.0              | 7              | 28               | 4             | 16              | 25    |                                                             |
| <b>&gt;20</b>      | 20              | 76.9              | 5              | 19.2             | 1             | 3.8             | 26    | <b>0.006</b><br>Fisher's exact<br>with simulated<br>p value |

**Table 12.** Results of multinomial regression investigating the association of smoking with alcohol consumption. The reference category is never smokers and no alcohol. Significant results in bold. CIs in brackets.

| Alcohol units/week | Current smoker RRR (CI)      | Adjusted model Current smoker RRR (CI) | Former smoker RRR (CI)    | Adjusted model former smoker RRR (CI) |
|--------------------|------------------------------|----------------------------------------|---------------------------|---------------------------------------|
| <b>0</b>           |                              |                                        |                           |                                       |
| <b>1 to 10</b>     | <b>4.95 (1.91 – 12.85)</b>   | <b>8.14 (1.95 – 34)</b>                | <b>3.11 (1.15 – 8.44)</b> | 2.63 (0.68 – 10.2)                    |
| <b>11 to 20</b>    | <b>4.67 (1.22 – 17.82)</b>   | <b>10.19 (1.26 – 82.5)</b>             | 2.55 (0.6 – 10.84)        | 3.57 (0.39 – 32.6)                    |
| <b>&gt;20</b>      | <b>26.67 (3.13 – 227.44)</b> | <b>61.13 (4.22 – 884.95)</b>           | 7.27 (0.74 – 71.13)       | 8.76 (0.56 – 136.35)                  |

Finally, alcohol consumption was associated with mouthwash use ( $p=0.02$ ) with all categories of alcohol consumption presenting increased odds of mouthwash use. However, this association was not significant after adjusting for other variables. There was no association between alcohol consumption and the site ( $p=0.98$ ), clinical appearance ( $p=0.62$ ) or degree of dysplasia of the OLK ( $p=0.37$ ) or with the oral hygiene index score ( $p=0.43$ ), brushing habit ( $p=0.39$ ), number of missing teeth ( $p=0.14$ ) or denture wear ( $p=0.91$ ) (data not shown).

### 5.3.2 Disease sites

The most common disease site was the floor of mouth/ventral tongue (63 cases, 28%), followed by the buccal mucosa (53 cases, 23.6%) and lateral border of the tongue (42 cases, 18.7%) (**Table 13**).

**Table 13.** Disease sites, histological findings and clinical appearance of OLK

| OLK data                           | Total             |
|------------------------------------|-------------------|
| <b>Disease sites (225)</b>         |                   |
| Floor of mouth/ventral tongue      | 63 (28%)          |
| Buccal mucosa/commissure           | 53 (23.6%)        |
| Lateral border tongue              | 42 (18.7%)        |
| Alveolus/Gingiva                   | 31 (13.8%)        |
| Soft palate                        | 19 (8.4%)         |
| Hard palate                        | 9 (4%)            |
| Dorsum tongue                      | 8 (3.5%)          |
|                                    |                   |
| <b>Histological findings (225)</b> |                   |
| No dysplasia                       | 11 (4.9%)         |
| Mild dysplasia                     | 83 (36.9%)        |
| Moderate dysplasia                 | 81 (36%)          |
| Severe dysplasia                   | 50 (22.2%)        |
|                                    |                   |
| <b>Clinical appearance (225)</b>   |                   |
| Homogenous                         | 186 (82.7%)       |
| Non-homogenous                     | 39 (17.3%)        |
| - of which ulcerated               | 7 (3.1% of total) |

### 5.3.3 Oral hygiene

The oral hygiene index values ranged from 0 to 4.2, with a mean of 0.96 (median 1). The mean number of missing teeth in the cohort was 11 (median 8, range 0 to 32). Brushing habits ranged from 0 to three times per day (data not shown). Some form of denture was worn by 23.7% (n=44) of the cohort.

### 5.3.4 Degree of dysplasia

All degrees of dysplasia were represented in the sample cohort. Eleven (4.9%) samples showed no dysplasia on histopathological examination while 83 (36.9%), 81 (36%) and 50 (22.2%) had mild, moderate and severe dysplasia, respectively (**Table 13**). Ten OSCCs, which presented clinically as OLK (all non-homogenous), were also sampled and were excluded from the analysis of risk factors. Analysis was also carried out with dysplasia classified into low-risk and high-risk dysplasia (low-risk=no/mild dysplasia and high-risk=moderate/severe dysplasia)<sup>124,125,225</sup>, in an attempt to classify risk more clearly, but also due to small sample numbers in some subsets preventing analysis.

Site of OLK ( $p=0.0005$ ), clinical appearance of OLK ( $p=0.0005$ ) and presence of ulceration ( $p=0.003$ ) were all significantly associated with the degree of dysplasia (**Table 14, Table 15**).

**Table 14.** Distribution of clinical parameters by three-tier grade of dysplasia \*Fisher's exact test with simulated p value for 2000 replicates (n= in brackets, significant results in bold)

| DYSPLASIA                            | NONE        | MILD (%)     | MODERATE (%) | SEVERE (%)   | P VALUE*       |
|--------------------------------------|-------------|--------------|--------------|--------------|----------------|
| PARAMETER (n=OLK)                    | 4.9% (n=11) | 36.9% (n=83) | 36% (n=81)   | 22.2% (n=50) |                |
| <b>GENDER</b>                        |             |              |              |              |                |
| Male (n=117)                         | 3.4 (4)     | 41 (48)      | 35 (41)      | 20.5 (24)    |                |
| Female (n=108)                       | 6.5 (7)     | 32.4 (35)    | 37 (40)      | 24.2 (26)    | 0.47           |
| <b>AGE</b>                           |             |              |              |              |                |
| <40 years (n=20)                     | 15 (3)      | 55 (11)      | 20 (4)       | 10 (2)       |                |
| 40-60 years (n=106)                  | 4.7 (5)     | 38.7 (41)    | 34 (36)      | 22.6 (24)    |                |
| >60 years (n=99)                     | 3 (3)       | 31.8 (31)    | 41.4 (41)    | 24.4 (24)    | 0.09*          |
| <b>SMOKING STATUS</b>                |             |              |              |              |                |
| Current (n=115)                      | 3.5 (4)     | 42.6 (49)    | 35.7 (41)    | 18.3 (21)    |                |
| Former (n=67)                        | 6 (4)       | 35.8 (24)    | 37.3 (25)    | 20.9 (14)    |                |
| Never (n=43)                         | 7 (3)       | 23.3 (10)    | 34.9 (15)    | 34.9 (15)    | 0.21*          |
| <b>ALCOHOL UNITS/WEEK</b>            |             |              |              |              |                |
| 0 (n=44)                             | 6.8 (3)     | 27.3 (12)    | 34.1 (15)    | 31.8 (14)    |                |
| 1 to 10 (n=115)                      | 7 (8)       | 40 (46)      | 34.8 (40)    | 18.3 (21)    |                |
| 11 to 20 (n=30)                      | 0 (0)       | 30 (9)       | 46.7 (14)    | 23.3 (7)     |                |
| >20 (n=36)                           | 0 (0)       | 44.4 (16)    | 33.3 (12)    | 22.2 (8)     | 0.37*          |
| <b>CLINICAL</b>                      |             |              |              |              |                |
| NHOLK (n=39)                         | 0 (0)       | 2.6 (1)      | 30.8 (12)    | 66.7 (26)    |                |
| OLK (n=186)                          | 5.9 (11)    | 44.1 (82)    | 37.1 (69)    | 12.9 (24)    | <b>0.0005</b>  |
| <b>ULCERATION</b>                    |             |              |              |              |                |
| No (n=218)                           | 4.8 (11)    | 38.1 (83)    | 36.7 (80)    | 20.2 (44)    |                |
| Yes (n=7)                            | 0 (0)       | 0 (0)        | 14.3 (1)     | 85.7 (6)     | <b>0.004</b>   |
| <b>SITE</b>                          |             |              |              |              |                |
| Alveolus/Gingiva (n=31)              | 9.4 (3)     | 48.4 (15)    | 35.5 (11)    | 6.5 (2)      |                |
| Buccal mucosa (n=53)                 | 5.6 (3)     | 47.2 (25)    | 41.5 (22)    | 5.7 (3)      |                |
| Dorsum tongue (n=8)                  | 0 (0)       | 50 (4)       | 25 (2)       | 25 (2)       |                |
| Floor of mouth/ventral tongue (n=63) | 1.6 (1)     | 31.7 (20)    | 36.5 (23)    | 30.2 (19)    |                |
| Hard palate (n=9)                    | 11.1 (1)    | 55.6 (5)     | 33.3 (3)     | 0 (0)        |                |
| Lateral tongue (n=42)                | 4.7 (2)     | 16.7 (7)     | 28.6 (12)    | 50 (21)      |                |
| Soft palate (n=19)                   | 5.3 (1)     | 36.8 (7)     | 42.1 (8)     | 15.8 (3)     | <b>0.0005*</b> |
| <b>MISSING TEETH</b>                 |             |              |              |              |                |
| <5 (n=29)                            | 10.3 (3)    | 31 (9)       | 27.6 (8)     | 31 (9)       |                |
| 5 to 15 (n=100)                      | 2 (2)       | 37 (37)      | 41 (41)      | 20 (20)      |                |
| >15 (n=42)                           | 2.4 (1)     | 19 (8)       | 33.3 (14)    | 45.2 (19)    | <b>0.02*</b>   |
| <b>DENTURE WEAR</b>                  |             |              |              |              |                |
| None (n=175)                         | 5.7 (10)    | 39.4 (69)    | 34.9 (61)    | 20 (35)      |                |
| Partial (n=41)                       | 2.4 (1)     | 34.1 (14)    | 34.1 (14)    | 29.3 (12)    |                |
| Complete (n=9)                       | 0 (0)       | 0 (0)        | 66.7 (6)     | 33.3 (3)     | 0.12           |
| <b>OHI SCORE</b>                     |             |              |              |              |                |
| Poor (n=46)                          | 4.3 (2)     | 37 (17)      | 37 (17)      | 21.7 (10)    |                |
| Fair (n=72)                          | 5.6 (4)     | 45.8 (33)    | 26.4 (19)    | 22.2 (16)    |                |
| Good (n=49)                          | 10.2 (5)    | 34.7 (17)    | 34.7 (17)    | 20.4 (10)    | 0.73           |

|                  |          |           |           |           |       |
|------------------|----------|-----------|-----------|-----------|-------|
| <b>BRUSHING</b>  |          |           |           |           |       |
| <1 (n=22)        | 0 (0)    | 31.8 (7)  | 36.4 (8)  | 31.8 (7)  |       |
| 1 (n=36)         | 0 (0)    | 33.3 (12) | 41.7 (15) | 25 (9)    |       |
| >1 (n=151)       | 7.3 (11) | 39.1 (59) | 31.8 (48) | 21.9 (33) | 0.46* |
| <b>MOUTHWASH</b> |          |           |           |           |       |
| No (n=125)       | 5.6 (7)  | 34.4 (43) | 38.4 (48) | 21.6 (27) |       |
| Yes (n=100)      | 4 (4)    | 40 (40)   | 33 (33)   | 23 (23)   | 0.75  |

**Table 15.** Distribution of clinical parameters by binary category of dysplasia \*Fisher's exact test with simulated p value for 2000 replicates (n= in brackets, bold=p<0.05)

| DYSPLASIA                            | NONE/MILD    | MODERATE/SEVERE | P VALUE            |
|--------------------------------------|--------------|-----------------|--------------------|
| PARAMETER (n=OLK)                    | 41.8% (n=94) | 58.2% (n=131)   |                    |
| <b>GENDER</b>                        |              |                 |                    |
| Male (n=117)                         | 44.4 (52)    | 56.6 (65)       |                    |
| Female (n=108)                       | 38.9 (42)    | 61.1 (66)       | 0.4                |
| <b>AGE</b>                           |              |                 |                    |
| <40 years (n=20)                     | 70 (14)      | 30 (6)          |                    |
| 40-60 years (n=106)                  | 43.3 (46)    | 56.6 (60)       |                    |
| >60 years (n=99)                     | 34.3 (34)    | 65.7 (65)       | <b>0.01</b>        |
| <b>SMOKING STATUS</b>                |              |                 |                    |
| Current (n=115)                      | 46.4 (53)    | 53.9 (62)       |                    |
| Former (n=67)                        | 41.8 (28)    | 58.2 (39)       |                    |
| Never (n=43)                         | 30.2 (13)    | 69.8 (30)       | 0.2                |
| <b>ALCOHOL UNITS/WEEK</b>            |              |                 |                    |
| 0 (n=44)                             | 34.1 (15)    | 55.9 (29)       |                    |
| 1 to 10 (n=115)                      | 47 (54)      | 53 (61)         |                    |
| 11 to 20 (n=30)                      | 30 (9)       | 70 (21)         |                    |
| >20 (n=36)                           | 44.4 (16)    | 55.6 (20)       | 0.21               |
| <b>CLINICAL</b>                      |              |                 |                    |
| NHOLK (n=39)                         | 2.6 (1)      | 97.4 (38)       |                    |
| OLK (n=186)                          | 50 (93)      | 50 (93)         | <b>0.000000002</b> |
| <b>ULCERATION</b>                    |              |                 |                    |
| No (n=218)                           | 43.1 (94)    | 56.9 (124)      |                    |
| Yes (n=7)                            | 0 (0)        | 100 (7)         | <b>0.043</b>       |
| <b>SITE</b>                          |              |                 |                    |
| Alveolus/Gingiva (n=31)              | 58.1 (18)    | 41.9 (13)       |                    |
| Buccal mucosa (n=53)                 | 52.8 (28)    | 47.1 (25)       |                    |
| Dorsum tongue (n=8)                  | 50 (4)       | 50 (4)          |                    |
| Floor of mouth/ventral tongue (n=63) | 33.3 (21)    | 66.7 (42)       |                    |
| Hard palate (n=9)                    | 66.7 (6)     | 33.3 (3)        |                    |
| Lateral tongue (n=42)                | 21.4 (9)     | 78.6 (33)       |                    |
| Soft palate (n=19)                   | 42.1 (8)     | 57.9 (11)       | <b>0.004</b>       |
| <b>MISSING TEETH</b>                 |              |                 |                    |
| <5 (n=29)                            | 41.4 (12)    | 58.6 (17)       |                    |
| 5 to 15 (n=100)                      | 39 (39)      | 61 (61)         |                    |
| >15 (n=42)                           | 21.4 (9)     | 78.6 (33)       |                    |
| <b>DENTURE WEAR</b>                  |              |                 |                    |
| None (n=175)                         | 45.1 (79)    | 54.9 (96)       |                    |
| Partial (n=41)                       | 36.6 (15)    | 63.4 (26)       |                    |
| Complete (n=9)                       | 0 (0)        | 100 (9)         | <b>0.02</b>        |
| <b>OHI SCORE</b>                     |              |                 |                    |
| Poor (n=46)                          | 41.3 (19)    | 58.7 (27)       |                    |
| Fair (n=72)                          | 58.7 (37)    | 48.6 (35)       |                    |
| Good (n=49)                          | 44.9 (22)    | 55.1 (27)       | 0.54               |



|                  |           |           |      |
|------------------|-----------|-----------|------|
| <b>BRUSHING</b>  |           |           |      |
| <1 (n=22)        | 31.8 (7)  | 68.2 (15) |      |
| 1 (n=36)         | 33.3 (12) | 66.7 (24) |      |
| >1 (n=151)       | 46.4 (70) | 53.6 (81) | 0.2  |
| <b>MOUTHWASH</b> |           |           |      |
| No (n=125)       | 40 (50)   | 60 (75)   |      |
| Yes (n=100)      | 44 (44)   | 56 (56)   | 0.55 |

#### 5.3.4.1 Site

Site was significantly associated with the degree of dysplasia ( $p=0.0005$ ) (**Table 14**). 84% of the OLKs with severe dysplasia were located on the tongue or floor of mouth (42% on lateral border tongue (LBT), 38% on floor of mouth/ventral tongue (FOM/VT), 4% on dorsum tongue (DT)), compared with just 6% on the buccal mucosa (BM), 4% on the alveolus/gingiva (Alv/Gin), and 6% on the palate (PAL) (**Figure 11**) .

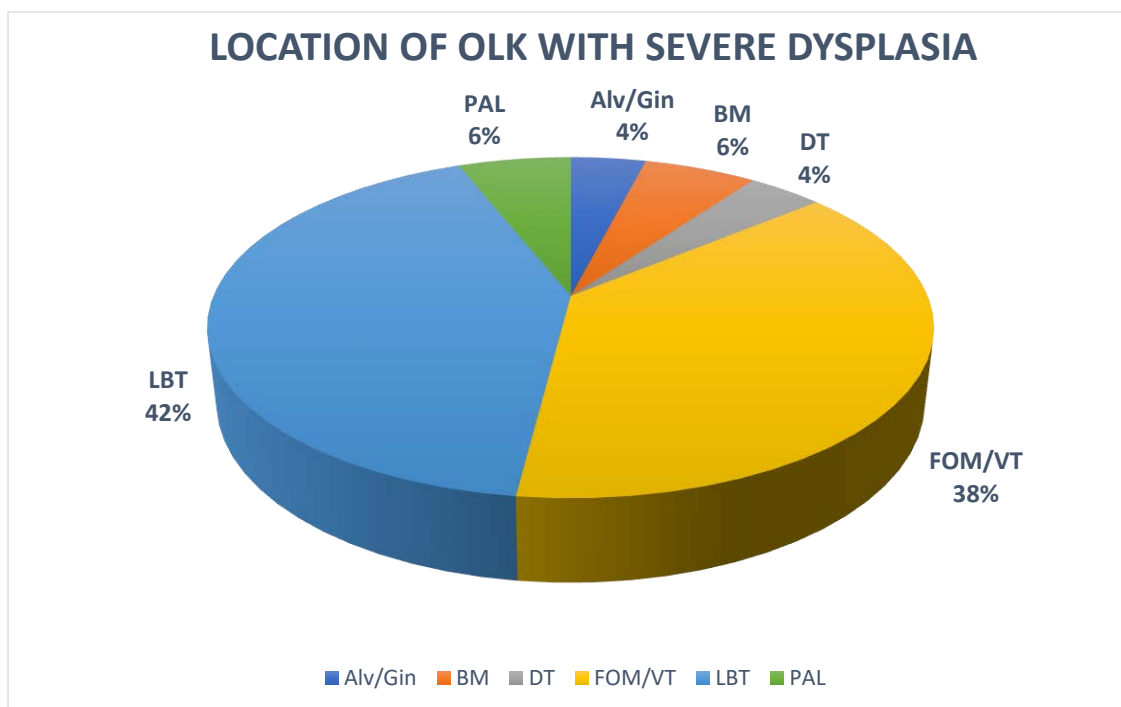


**Figure 11.** Location of OLKs with severe dysplasia. LBT=lateral border tongue, PAL=palate, Alv/Gin=Alveolus/Gingiva, BM=Buccal mucosa, DT=Dorsum tongue, FOMVT=Floor of mouth/ventral tongue

Logistic regression was carried out with dysplasia classified as low-risk and high-risk and dorsum and lateral border of tongue sites combined (due to the small number of dorsum tongue OLKs [n=8]). After adjusting for age, gender, smoking status, oral hygiene, missing teeth and clinical appearance (best fit model AIC: 172.18), OLKs on the dorsum/lateral tongue (DT/LBT), floor of mouth/ventral tongue, and soft palate (SP), were found to have significantly increased odds of having high-risk dysplasia compared with those on the buccal mucosa (adjusted odd ratios of 5.59 (95%CI:1.72-29.48), 3.59 (95%CI:1.18-11.7) and 6.63 respectively (95%CI:1.42-34.87)) (**Table 16**).

**Table 16.** Regression analysis showing odds of high-risk dysplasia by site (no/mild dysplasia and buccal mucosa used as reference category). Significant results in bold. (BM=Buccal mucosa, Alv/Gin=Alveolus/Gingiva, DT/LBT=Dorsum/Lateral tongue, FOM/VT=Floor of mouth/Ventral tongue, HP=Hard palate, SP=soft palate).

| Site    | Odds ratios (OR) | p value      | 95%CI              | Adjusted OR | Adjusted p value | Adjusted 95%CI      |
|---------|------------------|--------------|--------------------|-------------|------------------|---------------------|
| BM      |                  |              |                    |             |                  |                     |
| Alv/Gin | 0.81             | 0.68         | 0.33 - 1.97        | 1.65        | 0.51             | 0.35 - 7.27         |
| DT/LBT  | <b>3.19</b>      | <b>0.006</b> | <b>1.42 - 4.49</b> | <b>5.59</b> | <b>0.009</b>     | <b>1.72 - 29.48</b> |
| FOM/VT  | <b>2.24</b>      | <b>0.04</b>  | <b>1.06 - 4.8</b>  | <b>3.59</b> | <b>0.03</b>      | <b>1.18 - 11.7</b>  |
| HP      | 0.56             | 0.44         | 0.11 - 2.36        | 2.1         | 0.55             | 0.19 - 23.75        |
| SP      | 1.54             | 0.42         | 0.54 - 4.57        | <b>6.63</b> | <b>0.02</b>      | <b>1.42 - 34.87</b> |

Multinomial regression was then carried out with dysplasia classified as none, mild, moderate and severe and hard and soft palate sites combined. After adjusting for age, gender, smoking status and oral hygiene (best fit model AIC: 422.55), DT/LBT and FOM/VT sites presented 38.64 and 70.4 times, respectively increased risks of being severely dysplastic on biopsy. Although the best fit model was used, the CIs were wide due to small numbers in some subsets (**Table 17**). Therefore, dorsal and lateral tongue sites were combined with floor of mouth/ventral tongue sites, and the analysis repeated using the same

model (AIC: 418.24). Using these groupings FOM/tongue sites showed a 51.32 times increased risk of having severe dysplasia (RR 14, 95%CI:3.45 – 762.62) (**Table 18**).

**Table 17.** Multinomial regression investigating the effect of site on the degree of dysplasia. The reference category is no dysplasia and buccal mucosa. Significant results in bold. CIs in brackets.

| Site    | Mild dysplasia RRR (CI) | Adjusted model mild dysplasia RRR (CI) | Moderate dysplasia RRR (CI) | Adjusted model moderate dysplasia RRR (CI) | Severe dysplasia RRR (CI)            | Adjusted model severe dysplasia RRR (CI) |
|---------|-------------------------|----------------------------------------|-----------------------------|--------------------------------------------|--------------------------------------|------------------------------------------|
| BM      |                         |                                        |                             |                                            |                                      |                                          |
| DT/LBT  | 0.66<br>(0.1 - 4.52)    | 0.75<br>(0.09 - 6.69)                  | 0.95<br>(0.14 - 6.45)       | 1.17<br>(0.13 - 10.5)                      | <b>11.5</b><br><b>(1.33 - 99.32)</b> | <b>38.64</b><br><b>(2.18 - 685.68)</b>   |
| FOM/VT  | 2.4<br>(0.23 - 24.87)   | 2.45<br>(0.21 - 29.16)                 | 3.14<br>(0.3 - 32.47)       | 3.75<br>(0.31 - 45.68)                     | <b>19</b><br><b>(1.45 - 248.22)</b>  | <b>70.4</b><br><b>(2.93 - 1693.43)</b>   |
| Alv/Gin | 0.6<br>(0.11 - 3.36)    | 0.74<br>(0.11 - 5.02)                  | 0.5<br>(0.09 - 2.9)         | 0.31<br>(0.04 - 2.52)                      | 0.67<br>(0.06 - 7.35)                | 1.23<br>(0.04 - 34.32)                   |
| PAL     | 0.72<br>(0.11 - 4.9)    | 0.67<br>(0.08 - 5.65)                  | 0.75<br>(0.11 - 5.17)       | 1.3<br>(0.15 - 11.18)                      | 1.5<br>(0.14 - 16.54)                | 5.45<br>(0.24 - 124.31)                  |

**Table 18.** Multinomial regression investigating the effect of site on the degree of dysplasia with dorsum, lateral and ventral tongue and floor of mouth (FOM) sites combined as FOM/tongue. The reference category is no dysplasia and buccal mucosa. Significant results in bold. CIs in brackets.

| Degree of dysplasia | FOM/tongue RRR (CI)                 | Adjusted model FOM/tongue RRR (CI)     | Alveolus/ Gingiva RRR (CI) | Adjusted model Alveolus/gingiva RRR (CI) | Palate RRR (CI)       | Adjusted model Palate RRR (CI) |
|---------------------|-------------------------------------|----------------------------------------|----------------------------|------------------------------------------|-----------------------|--------------------------------|
| None                |                                     |                                        |                            |                                          |                       |                                |
| Mild                | 1.24<br>(0.23 – 6.69)               | 1.35<br>(0.21 – 8.68)                  | 0.6<br>(0.11 – 3.36)       | 0.75<br>(0.11 – 5.05)                    | 0.72<br>(0.11 – 4.89) | 0.67<br>(0.08 – 5.8)           |
| Moderate            | 1.68<br>(0.31 – 9.07)               | 2.07<br>(0.32 – 13.58)                 | 0.5<br>(0.09 – 2.89)       | 0.31<br>(0.04 – 2.53)                    | 0.75<br>(0.11 – 5.17) | 1.32<br>(0.15 – 11.42)         |
| Severe              | <b>14</b><br><b>(1.93 – 101.71)</b> | <b>51.32</b><br><b>(3.45 – 762.62)</b> | 0.7<br>(0.06 – 7.35)       | 1.12<br>(0.04 – 34.37)                   | 1.5<br>(0.14 – 16.54) | 5.78<br>(0.25 – 132.68)        |

#### 5.3.4.2 Clinical appearance (Homogeneity and ulceration)

The degree of dysplasia was significantly associated with the clinical appearance of OLK (p=0.0005) with 66.7% of non-homogenous OLKs (NHOLK) showing severe dysplasia on biopsy compared with 12.9% of homogenous OLKs, while only 2.6% of NHOLKs but 44.1% of homogenous OLKs showed mild dysplasia (**Table 19**). Logistic regression

confirmed this association with NHOLKs presenting significantly increased odds (adj OR=25.41) of having high-risk dysplasia, although the confidence interval was wide, likely due to the small number of NHOLKs in the sample (**Table 19**). Multinomial regression was not possible for the same reason.

**Table 19.** Regression analysis showing odds of high-risk dysplasia by clinical appearance

| Parameter          | Odds ratio (OR) | p value | 95%CI         | Adjusted OR | Adjusted p value | Adjusted 95%CI |
|--------------------|-----------------|---------|---------------|-------------|------------------|----------------|
| Non-homogenous OLK | 38              | 0.0004  | 7.98 – 681.48 | 25.41       | 0.004            | 4.13 – 519.38  |

Presence of ulceration was also significantly associated with the degree of dysplasia (p=0.003) with 85.7% of ulcerated OLKs showing severe dysplasia, compared with 20.2% of non-ulcerated OLKs (**Table 14**). Regression analysis was not possible on this subset due to the small number of ulcerated OLKs (n=7).

#### 5.3.4.3 Age

Age was significantly associated with the degree of dysplasia with dysplasia classified in a binary manner (p=0.01) (**Table 15**). Multinomial regression confirmed this association with age over 60 years presenting an increased risk of mild, moderate and severe dysplasia. After adjusting for site, gender, smoking status and oral hygiene (best model AIC: 418.24), age over 60 years presented an increased risk of having moderate dysplasia (adj RRR 17.57), but not severe dysplasia (**Table 20**).

**Table 20.** Results of multinomial regression showing association of age with the degree of dysplasia. The reference category is no dysplasia and age <40. Significant results in bold. CIs in brackets.

| Degree of dysplasia | Age 40-60 years<br>RRR (CI) | Adjusted model<br>Age 40-60 years<br>RRR (CI) | Age >60 years<br>RRR (CI)             | Adjusted model<br>Age >60 years<br>RRR (CI) |
|---------------------|-----------------------------|-----------------------------------------------|---------------------------------------|---------------------------------------------|
| None                |                             |                                               |                                       |                                             |
| Mild                | 2.24<br>(0.46 – 10.84)      | 2.09<br>(0.35 – 12.55)                        | <b>2.82</b><br><b>(4.49 – 16.09)</b>  | 4.77<br>(0.62 – 36.86)                      |
| Moderate            | 5.39<br>(0.92 – 31.55)      | 7.45<br>(0.86 – 64.68)                        | <b>10.25</b><br><b>(1.53 – 68.62)</b> | <b>17.57</b><br><b>(1.66 – 186.12)</b>      |
| Severe              | 7.19<br>(0.94 – 54.94)      | 4.96<br>(0.47 – 52.46)                        | <b>12</b><br><b>(1.39 – 103.48)</b>   | 7.92<br>(0.63 – 99.95)                      |

#### 5.3.4.4 Other parameters

Finally, there were significant associations between the number of missing teeth ( $p=0.02$ ) when dysplasia was classified as none, mild, moderate and severe (**Table 14**), and with denture wear ( $p=0.02$ ) with dysplasia classified in a binary manner (**Table 15**). However, these associations with missing teeth and denture wear were not confirmed on regression analysis. Gender, smoking status, alcohol consumption, oral hygiene score, brushing habit and mouthwash use were not associated with the degree of dysplasia on biopsy (**Table 14**).

### 5.3.5 Clinical appearance

Of the 225 OLKs sampled, 186 were homogenous and 39 were non-homogenous (**Figure 12**). Seven of the 39 non-homogenous OLKs were ulcerated. The clinical appearance was significantly associated with the degree of dysplasia as mentioned above.



**Figure 12.** Non-homogenous leukoplakia right lateral border of the tongue. Photograph courtesy Dr. S. Galvin.

Clinical appearance was also significantly associated with the site of the OLK, with 76.9% of the NHOLKs located either on the FOM/VT or LBT (41% on the FOM/VT, 35.9% on the LBT) ( $p=0.01$ ). Regression analysis confirmed this association with FOM/VT OLKs showing 3.27 times increased odds, and LBT OLKs 4.8 times increased odds, of being non-homogenous (**Table 21**).

**Table 21.** Logistic regression analysis showing odds of NHOLK at various sites (significant results in bold)

| Parameter             | Odds ratio (OR) | p value      | 95%CI               | Adjusted OR | Adjusted p value | Adjusted 95%CI      |
|-----------------------|-----------------|--------------|---------------------|-------------|------------------|---------------------|
| Dorsum tongue         | 1.37            | 0.79         | 0.07 – 10.35        | 1.3         | 0.82             | 0.06 – 10.42        |
| FOM/ventral tongue    | <b>3.27</b>     | <b>0.03</b>  | <b>1.17 – 10.64</b> | <b>3.18</b> | <b>0.04</b>      | <b>1.12 – 10.59</b> |
| Alveolus/Gingiva      | 0.32            | 0.31         | 0.02 – 2.11         | 0.3         | 0.8              | 0.02 – 2.09         |
| Lateral border tongue | <b>4.8</b>      | <b>0.006</b> | <b>1.65 – 16.17</b> | <b>4.82</b> | <b>0.008</b>     | <b>1.6 – 16.73</b>  |
| Hard/soft palate      | 0.74            | 0.73         | 0.1 – 3.7           | 5.55        | 0.78             | 0.1 – 3.98          |

## 5.4 Follow up

Twenty three OLKs had a repeat biopsy over the study period due to a change in the clinical appearance. Two had progressed to OSCC, five had an increase in severity of the degree of dysplasia, six had a reduction in the degree of dysplasia, while the degree of dysplasia was unchanged in the remaining ten. No significant associations between any clinical or demographic characteristic and progression or regression of OLK were found, although this is likely due to the small numbers of OLKs that progressed or regressed over the study period.

The two OLKs that progressed to OSCC over the timeframe of the study were in two male former smokers, both aged over 75. Both patients consumed alcohol, one more than 10 units, the other more than 20 units, per week. Both OLKs showed moderate dysplasia on initial biopsy, one on alveolus/gingiva, one on lateral tongue. The time from initial biopsy to progression to OSCC was 22 and 32 months respectively.

## 5.5 Discussion

It is well established that smoking and alcohol consumption are the principal risk factors for OLK<sup>117–121,129,226</sup> and this is also true for my patient cohort with 81.2% reporting current (52.7%) or past (28.5%) tobacco use, and 79% consuming alcohol, including almost 14% who were heavy alcohol consumers (more than 20 units per week). There was no difference between the smoking habits of males and females. However, alcohol consumption was significantly associated with gender; males being more than 15 times more likely to drink heavily than females. Alcohol consumption was also significantly associated with smoking, current smokers having increased relative risk ratios of 8.14, 10.19 and 61.13 for consuming one to 10, 11 to 20 and more than 20 units per week respectively, compared with never smokers. With tobacco and alcohol known to act as synergistic oral carcinogens<sup>5,28–30</sup>, it is concerning that smokers are more likely to consume alcohol, particularly heavily.

The floor of mouth/ventral tongue was the most common site for OLK, with 28.1% of OLKs located here, followed by the buccal mucosa (23%) and lateral border of tongue (19.1%), somewhat disagreeing with current literature which reports the tongue and buccal mucosa as the most common sites for OLK<sup>3,129,140</sup>. However, in agreement with the literature<sup>129,227,228</sup>, the site of the OLK was significantly associated with smoking habit, with several authors demonstrating that the floor of the mouth is the site of predilection for OLK in smokers. Schepman *et al.*<sup>227</sup>, in a retrospective review of 145 OLKs, found OLKs in male smokers almost 8.5 times, and in female smokers 18 times, more likely to occur in the floor of the mouth than in other oral sites. Interestingly, male smokers were also found to be 7.5 times more likely to have buccal mucosa OLKs than non-smokers. Pentenero *et al.*<sup>228</sup> also found smoking habit to be associated with OLK site. Dysplastic and non-dysplastic floor of mouth OLKs were seven and nine times respectively more likely to occur in smokers (defined as



ever smokers) than non-smokers, while dysplastic buccal mucosa OLKs were 10 times more likely to occur in smokers than non-smokers. Both authors defined smokers as ever smokers and used logistic regression; comparing the individual OLK sites with other OLK sites combined.

In this study, I analysed the data slightly differently, using multivariate analysis to assess the relative risk of OLK location compared to location on the buccal mucosa in the various smoking categories as the buccal mucosa is considered a low-risk site for malignant transformation. In agreement with the literature above<sup>227,228</sup>, OLKs presented a six times increased relative risk in current smokers, and a more than nine times increased relative risk in former smokers, of being located on the floor of mouth/ventral tongue compared with the buccal mucosa. This association also explains the predominance of floor of mouth/ventral tongue OLKs in this study as 81.2% of the patient cohort were either current or past smokers.

All degrees of dysplasia were represented in the sample cohort. There were eleven (4.9%) OLKs with no dysplasia, 83 (36.9%) with mild dysplasia, 81 (36%) with moderate dysplasia and 50 (22.2%) with severe dysplasia. Site of OLK ( $p=0.0005$ ), clinical appearance of OLK ( $p=0.0005$ ) and presence of ulceration ( $p=0.003$ ) were all significantly associated with the degree of dysplasia. Eighty four percent of the OLKs with severe dysplasia were located on the tongue and floor of mouth, compared with just 6% on the buccal mucosa, 4% on the alveolus/gingiva, and 6% on the palate. Logistic regression, with dysplasia classified as low-risk (no/mild dysplasia) and high-risk (moderate/severe dysplasia) and dorsum and lateral border of tongue sites combined (due to the small number of dorsum tongue OLKs) showed that OLKs on the dorsum/lateral tongue, floor of mouth/ventral tongue, and soft palate, had significantly increased odds of having high-risk dysplasia compared with those

on the buccal mucosa. Multinomial regression with dysplasia classified as none, mild, moderate and severe found that DT/LBT and FOM/VT OLKs presented significantly higher relative risks of having severe dysplasia. After adjustment, DT/LBT OLKs were 38.64 times more likely, and FOM/VT OLKs 70.4 times more likely, to be severely dysplastic than OLKs on the buccal mucosa. Soft and hard palate sites could not be assessed separately in the multinomial regression model due to the small number of soft (n=19) and hard (n=9) palate OLKs. There is agreement in the literature that OLKs on the tongue and floor of mouth are at higher risk of malignant transformation than OLKs at other sites<sup>3,40,129,130,133–136,139,140</sup>. The largest systematic review examining the malignant transformation of OLK<sup>3</sup> included 16,604 patients with OLK and found that 56.3% of the OLKs that underwent malignant transformation were located on the tongue, 98% of which were on the ventral or lateral surfaces. Further systematic reviews by Guan *et al.*, Paglioni *et al.*, and Warnakulsuriya *et al.*<sup>128–130</sup> also found that OLKs on the tongue, along with the floor of mouth, were at increased risk of malignant transformation; Guan *et al.* reported that tongue OLKs had a 6.38 times, and Paglioni *et al.* that tongue/floor of mouth OLKs had a 4.48 times, increased risk of malignant transformation while Warnakulasuriya *et al.* reported that 24.2% and 14.9% of the OLKs that transformed to cancer in their review were located on the tongue or tongue/floor of mouth respectively. My findings that OLKs on the dorsum/lateral tongue and floor of mouth/ventral tongue were almost 40 and 70 times, respectively more likely to show severe dysplasia support this, as it is well established that the degree of dysplasia on biopsy is the most accurate predictor of malignant transformation risk, severe dysplasia having the highest risk<sup>3,127,128,131,134,136</sup>. The meta-analysis of Aguirre-Urizar *et al.*<sup>3</sup> found that severe dysplasia had a 5.21 times increased risk of malignant transformation compared with mild/moderate dysplasia, while the analysis of Guan *et al.*<sup>128</sup> identified a 7.13 times increased risk of malignant transformation for severe dysplasia

compared with no dysplasia. Finally, Warnakulasuriya *et al.* showed that the 10-year malignant transformation rate for severe dysplasia was 26.7% compared with 4.8% for mild dysplasia.

I also found that NHOLKs were 25.41 times more likely to show moderate or severe dysplasia, while 85.7% of the ulcerated OLKs were severely dysplastic, agreeing with the results of a retrospective study by Embaló *et al.*<sup>229</sup> who reported that non-homogenous OLKs were 9.25 times more likely than homogenous OLKs to show high-risk dysplasia, and 10.5 times more likely to have severe, rather than mild, dysplasia. Non-homogeneity is widely accepted as a risk factor for malignant transformation, with systematic reviews reporting a 4.9 to 6.5 times increased risk of malignant transformation for NHOLK compared with homogenous OLK<sup>3,128,130</sup>. My finding that NHOLKs were significantly more likely to show high-risk dysplasia supports this association. Interestingly non-homogeneity was also associated with OLK site; lateral tongue OLKs having 4.8 times, and floor of mouth/ventral tongue OLKs 3.27 times, increased odds of being non-homogenous, confirming again that LBT and FOM/VT OLKs should be considered high risk for severe dysplasia and therefore for malignant transformation.

The only other variable associated with degree of dysplasia was age; age over 60 years being associated with an increased risk of moderate, but not severe dysplasia (adj RR 17.57). This agrees with current literature where some studies have found an association between increasing age and increased risk of malignant transformation. Aguirre-Urizar *et al.*<sup>3</sup> reported that 82.5% of the malignant transformations of OLK included in their systematic review were in those aged over 50 years, while Warnakulasuriya *et al.*<sup>134</sup> and Liu *et al.*<sup>136</sup>

found that the malignant transformation of OLK was significantly associated with age over 65 years and age over 60 years, respectively.

The degree of dysplasia was not found to be significantly associated with gender, smoking, alcohol consumption, oral hygiene score, brushing habit, tooth number, mouthwash use or denture wear in this study.

## 5.6 Conclusion

In conclusion, my findings demonstrate that oral site, clinical appearance and presence of ulceration are significantly associated with the degree of dysplasia of OLK on biopsy, with OLKs on the floor of the mouth and tongue, and those showing non-homogeneity and ulceration being significantly more likely to show severe dysplasia. The degree of dysplasia on biopsy is currently the most accurate predictor of malignant transformation of OLK and so OLKs on the floor of the mouth and tongue, non-homogenous OLKs, and OLKs with ulceration should be considered higher risk, in agreement with the literature. With regards to risk factors, smoking and alcohol consumption were unsurprisingly prevalent in our patient cohort. Smokers and former smokers were also more likely to consume alcohol than never smokers. Finally, smokers were significantly more likely to have OLKs located on the floor of the mouth/ventral tongue, a site known to be higher risk for malignant transformation.

In summary with regards to the aims of this chapter:

- Smoking was associated with OLK location on the floor of the mouth.
- Site, non-homogeneity and ulceration were associated with the degree of dysplasia OLK.

## 6 The microbiome of healthy mucosal surfaces and the influence of various factors on microbial colonisation

### 6.1 Introduction

The aim of this chapter was to establish how factors such as smoking and oral site affected the microbiome of clinically healthy mucosa. It was important to establish any influencing variables prior to investigating if the microbial colonisation pattern of OLK differed to that of contralateral normal sites, so that these factors could be adequately adjusted for in later analyses.

### 6.2 Methods

Healthy mucosal surfaces (contralateral normal [CLN] sites from dysplasia clinic patients and healthy control [HC] sites from healthy control subjects) were analysed together to investigate the influence of smoking, alcohol consumption, brushing habit, oral hygiene, denture wear, mouthwash use, tooth number, age, gender and mucosal site on the microbiome of clinically normal mucosal surfaces. A total of 202 CLN samples from 180 patients and 120 HC samples from 61 healthy controls were available for analysis (**Figure 13**). The methods as outlined in **Chapter 4** were followed. MaAsLin2 was used for multivariate analysis to assess the influence of smoking, alcohol consumption, tooth brushing, oral hygiene, denture wear, mouthwash use, tooth number and site on the mucosal microbiome. Significant similarities in microbiome composition were noted between dorsum tongue (DT) and lateral border tongue (LBT) sites, and between alveolus/gingiva

(Alv/Gin) and buccal mucosa (BM) sites on an initial analysis of CLN sites and so these sites were combined as DT/LBT and Buc/Alv/Gin respectively for the microbial site analysis.

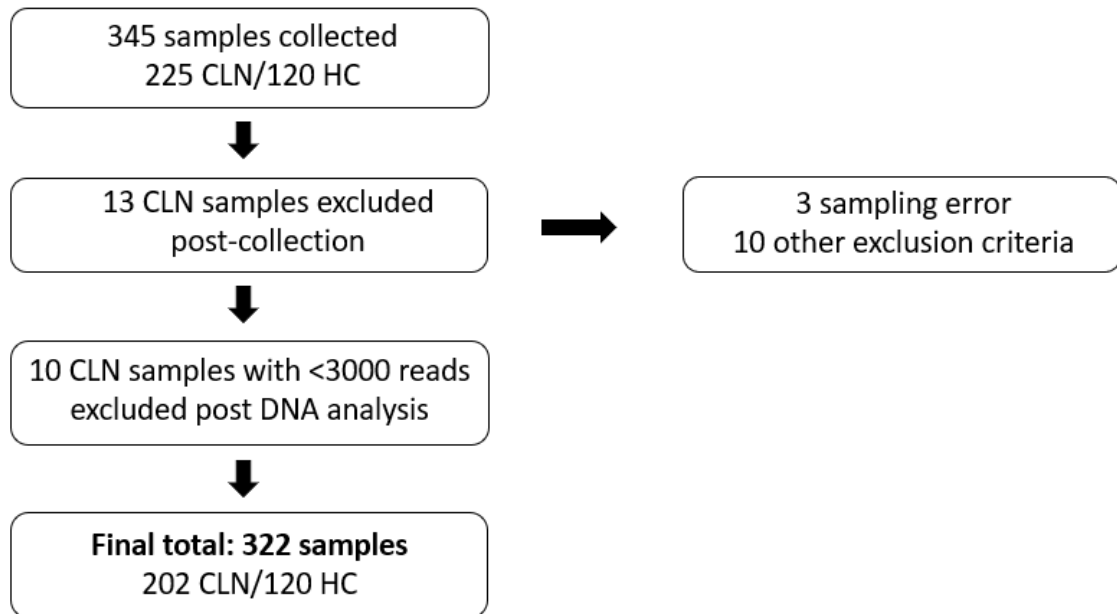


**Figure 13.** Flowchart of sample cohort of contralateral normal and healthy control sites

## 6.3 Results

There were significant differences between the patient and control cohorts for all demographic parameters other than denture wear and mouthwash use (**Table 22**). This is not unexpected given that the healthy control cohort was drawn from motivated, dentally healthy patients attending for routine oral hygiene treatment. However, these differences also ensured that there were sufficient cohorts of patients within each variable to facilitate statistical analysis.

**Table 22.** Demographics of control and contralateral normal (patient) cohorts. Significant differences in demographics highlighted in bold.

| Demographic                                 | Controls | Patients | p value         |
|---------------------------------------------|----------|----------|-----------------|
| Total                                       | 61       | 180      |                 |
| <b>Gender</b>                               |          |          |                 |
| Males                                       | 21       | 94       |                 |
| Females                                     | 40       | 86       | <b>0.02</b>     |
| <b>Age range</b>                            |          |          |                 |
| <40                                         | 17       | 17       |                 |
| 40-60                                       | 30       | 87       |                 |
| >60                                         | 13       | 76       |                 |
| Not recorded                                | 1        | 0        | <b>0.0003</b>   |
| <b>Smoking status</b>                       |          |          |                 |
| Current smoker                              | 15       | 96       |                 |
| Former smoker                               | 15       | 49       |                 |
| Never smoker                                | 31       | 35       | <b>0.000004</b> |
| <b>Alcohol units/week</b>                   |          |          |                 |
| 0                                           | 13       | 39       |                 |
| 1 to 10                                     | 41       | 90       |                 |
| 11 to 20                                    | 2        | 23       |                 |
| More than 20                                | 5        | 28       | <b>0.0001</b>   |
| <b>Alcohol &gt;20 units/week and smoker</b> |          |          |                 |
| Yes                                         | 1        | 29       |                 |
| No                                          | 60       | 151      | <b>0.001</b>    |
| <b>Missing teeth</b>                        |          |          |                 |
| <5                                          | 26       | 23       |                 |
| 5 to 15                                     | 26       | 76       |                 |
| More than 15                                | 7        | 39       |                 |
| Not recorded                                | 2        | 42       | <b>0.0001</b>   |
| <b>OHI Score</b>                            |          |          |                 |
| Good (OHI-S < 0.5)                          | 37       | 38       |                 |
| Fair (OHI-S 0.5-1)                          | 15       | 61       |                 |
| Poor (OHI-S > 1)                            | 7        | 37       |                 |
| Edentulous                                  | 0        | 9        |                 |
| Not recorded                                | 2        | 35       | <b>0.00002</b>  |
| <b>Brushing times/day</b>                   |          |          |                 |
| <1                                          | 0        | 19       |                 |
| 1                                           | 5        | 30       |                 |
| >1                                          | 56       | 121      |                 |
| Not recorded                                | 0        | 10       | <b>0.0008</b>   |
| <b>Denture</b>                              |          |          |                 |
| Complete (F/F)                              | 1        | 8        |                 |
| Partial                                     | 7        | 33       |                 |
| None                                        | 53       | 139      | 0.3             |
| <b>Mouthwash use</b>                        |          |          |                 |
| No                                          | 26       | 151      |                 |
| Yes                                         | 35       | 29       | 0.22            |

The sites sampled are listed in **Table 23**. Dorsum tongue (DT) and alveolus/gingiva (Alv/Gin) sites were not sampled in the healthy control cohort.

**Table 23.** Sites sampled (percentages in brackets)

| Sites sampled                          | Number (%) |
|----------------------------------------|------------|
| <b>Contralateral normal (202)</b>      |            |
| Floor of mouth/ventral tongue (FOM/VT) | 26 (12.9%) |
| Buccal mucosa/commissure (BM)          | 71 (35.1%) |
| Lateral border tongue (LBT)            | 67 (33.2%) |
| Alveolus/Gingiva (Alv/Gin)             | 14 (6.9%)  |
| Palate (PAL)                           | 19 (9.4%)  |
| Dorsum tongue (DT)                     | 5 (2.5%)   |
| <b>Healthy controls (120)</b>          |            |
| Floor of mouth/ventral tongue          | 17 (14.2%) |
| Buccal mucosa                          | 43 (35.9%) |
| Lateral border tongue                  | 43 (35.9%) |
| Palate                                 | 17 (14.2%) |

### 6.3.1 Smoking

The alpha diversity refers to the overall structure of an ecological community with regards to its richness (number of different taxonomic groups) and/or evenness (distribution of abundances between taxonomic groups)<sup>230</sup>. ACE, Chao1, Observe, Shannon and Simpson indices were used to calculate the alpha diversities of CLN and control samples. The ACE, Chao1 and Observe indices place more emphasis on richness while the Shannon and Simpson indices take both the richness and evenness into account<sup>231</sup> (<https://docs.onecodex.com/en/articles/4136553-alpha-diversity>). No significant differences in alpha diversity were noted between the different smoking categories (current, former and never).



The beta-diversity refers to the differences between samples or environments in the cohort being examined<sup>232</sup> (<https://docs.onecodex.com/en/articles/4150649-beta-diversity>). The Bray-Curtis dissimilarity index<sup>233</sup>, which examines both the abundances of species shared between two samples and the number of species found in each sample, was used to calculate the beta-diversity and compare the bacterial community structure of each patient sample. Matrices of the beta-diversity values were visualised using non-metric multidimensional scaling (NMDS) and the R package MicroViz was then used to carry out Permutational Multivariate Analysis of Variance Using Distance Matrices (PERMANOVA) tests to determine which communities exhibited significant differences in structure. Multivariate analysis was carried out using MaAsLin2 in R Studio with all variables first analysed separately to assess their individual influence on the microbiome before combining various factors into different multivariate models.

Unsurprisingly, smoking was found to have a significant influence on the microbial community of mucosal surfaces. Visualisation of the Bray-Curtis plots showed separation of current, former and never smokers' samples, particularly on axis 3 (**Figure 14, Figure 15**), and the PERMANOVA test confirmed that this separation was significant ( $p_{adj}=0.001$ ). Pairwise PERMANOVA comparisons showed that there were significant differences between the microbial communities of all three categories of smokers (**Table 24**).

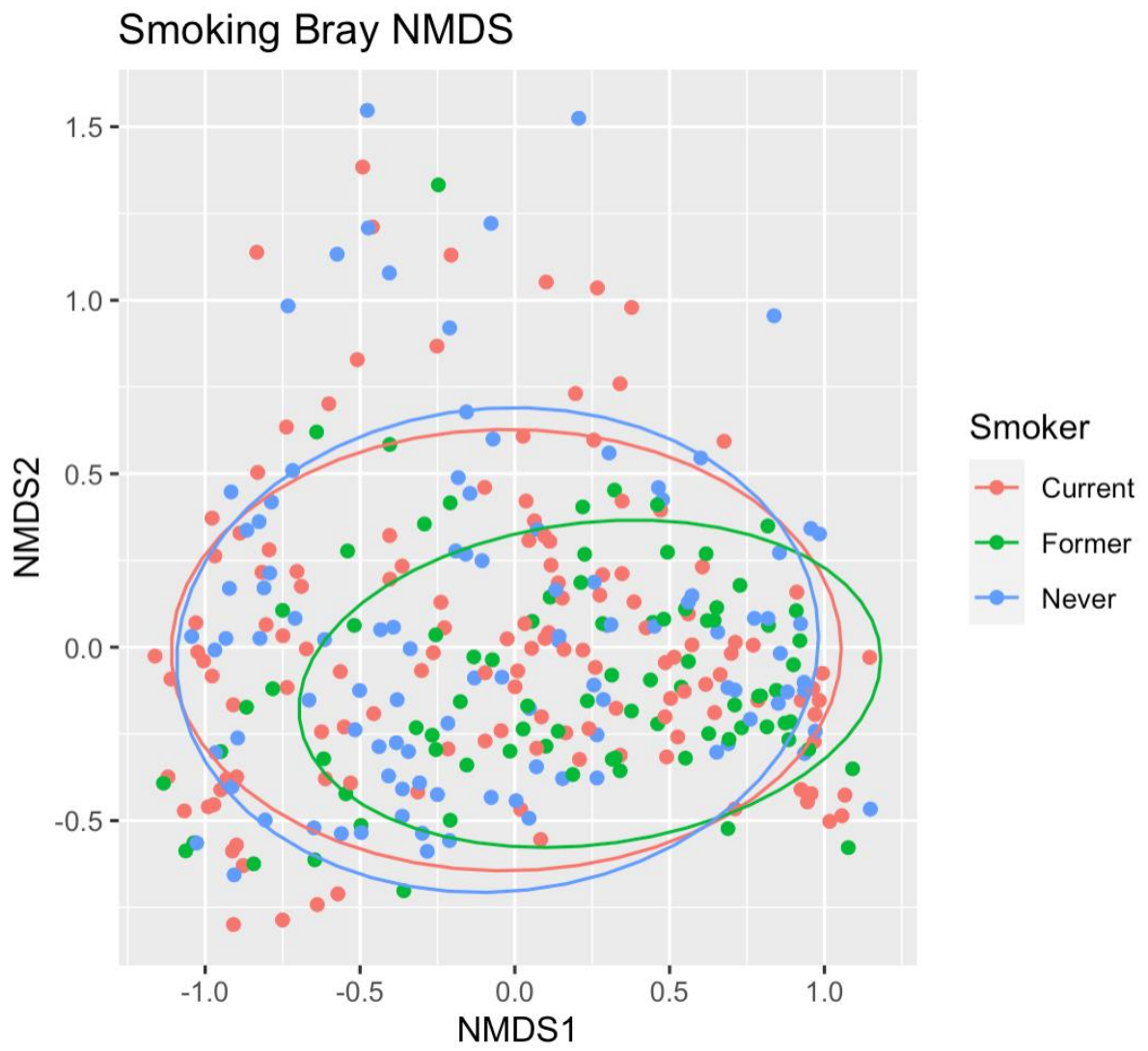


**Figure 14.** Bray Curtis NMDS plot of smoking status on axes 1 and 2 (padj=0.001)

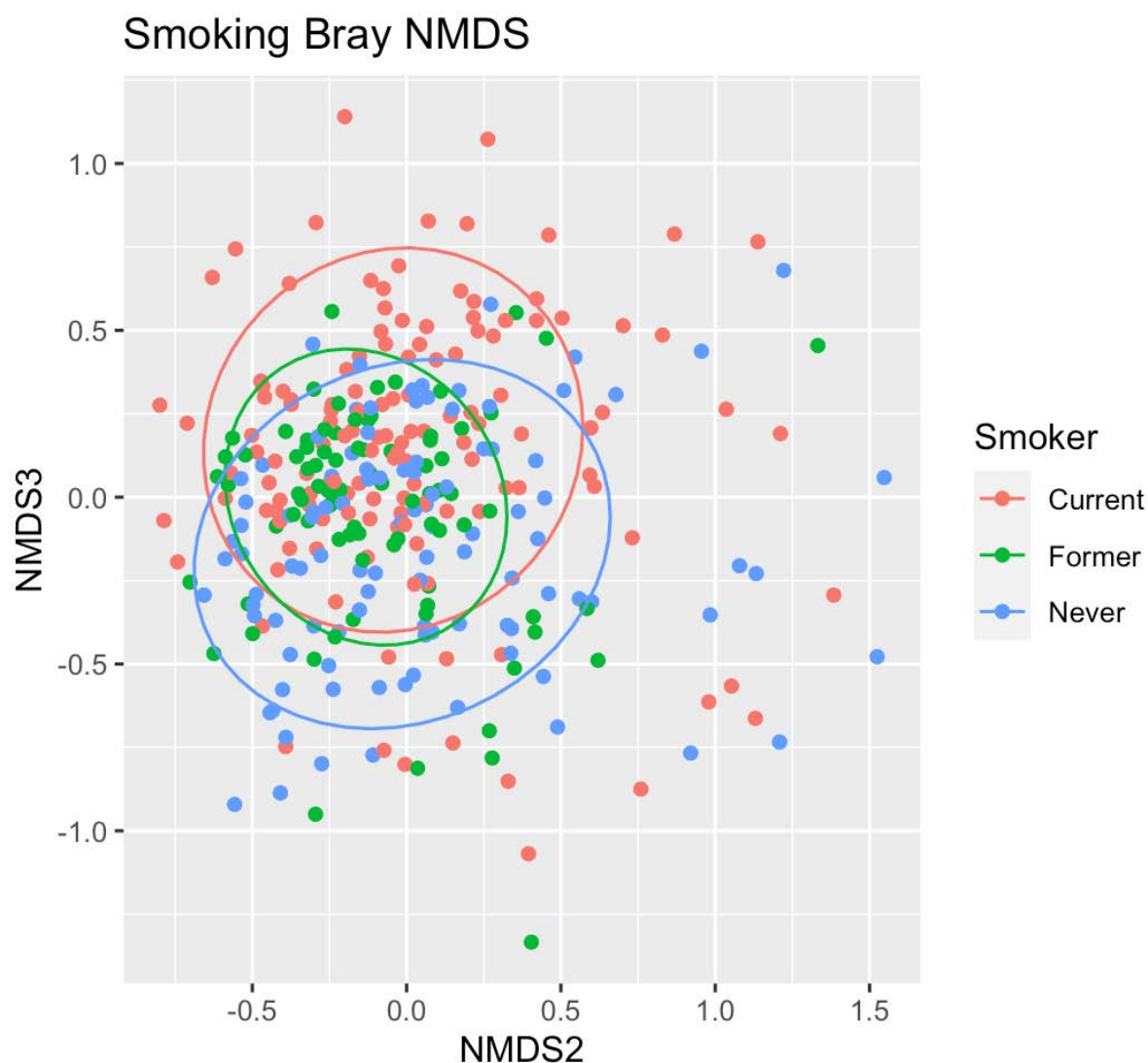


**Figure 15.** Bray Curtis NMDS plot of smoking status on axes 2 and 3 (padj=0.001)

**Table 24.** Pairwise PERMANOVA results for smoking status.

| Smoking status comparison | p value | Adjusted p value |
|---------------------------|---------|------------------|
| Never vs current          | 0.0001  | 0.0003           |
| Never vs former           | 0.002   | 0.002            |
| Current vs former         | 0.0002  | 0.0003           |

MaAsLin2 was used for multivariate analysis to identify which species were responsible for these differences. Thirty taxa showed significantly decreased, and 34 taxa significantly increased (padj<0.1) abundances associated with current smoking. Current smoking was

associated with loss of species typically associated with healthy mucosa, including several species of genera *Neisseria* and *Streptococcus*, and *Lautropia mirabilis*, *Gemella morbillorum*, *Rothia aeria*, *Selenomonas artemidis* and *Abiotrophia defectiva*. Several Gram-negative species from genera *Aggregatibacter*, *Campylobacter*, *Kingella*, *Bergeyella*, *Fusobacterium*, *Capnocytophaga* and a single species from *Haemophilus* (*Haemophilus sputorum*) also showed significant decreases in abundance in association with current smoking (**Table 25, Figure 16**). The species showing increased abundance associated with smoking were, as expected, predominantly periodontal pathogens including species of *Prevotella*, *Actinomyces*, *Porphyromonas*, *Fusobacterium*, *Tannerella*, *Parvimonas*, *Filifactor*, *Clostridiales*, *Veillonella* and *Peptostreptococcus* genera. Several streptococci were also significantly increased. Previous smoking was associated with reduced abundance of five *Streptococcus* spp., two *Actinomyces* spp., *Selenomonas artemidis*, *Cardiobacterium hominis*, *Rothia aeria* and increased abundance of *Ruminococcaceae* G2 bacterium HMT085 and one unspecified streptococcal species.

**Table 25.** Taxa with significant changes in abundance associated with smoking on MaAsLin2 analysis (negative coefficient values denote reduced abundance).

| Taxa                                         | Variable       | Coefficient | p value  | Adjusted p value |
|----------------------------------------------|----------------|-------------|----------|------------------|
| <i>Neisseria perflava flavescens</i>         | Current smoker | -1.23       | 4.81E-11 | 1.31E-08         |
| <i>Streptococcus</i> sp. HMT066 sp. HMT074   | Current smoker | -0.73       | 2.74E-07 | 0.00002          |
| <i>Lautropia mirabilis</i>                   | Current smoker | -0.72       | 7.02E-11 | 1.31E-08         |
| <i>Haemophilus parainfluenzae</i>            | Current smoker | -0.72       | 6.17E-09 | 5.77E-07         |
| <i>Rothia aeria</i>                          | Current smoker | -0.68       | 4.03E-09 | 5.03E-07         |
| <i>Streptococcus australis</i>               | Current smoker | -0.67       | 6.45E-08 | 4.82E-06         |
| <i>Kingella oralis</i>                       | Current smoker | -0.64       | 1.45E-06 | 0.00005          |
| <i>Streptococcus sanguinis</i>               | Current smoker | -0.52       | 9.61E-06 | 0.0002           |
| <i>Neisseria macacae flava mucosa sicca</i>  | Current smoker | -0.50       | 0.0003   | 0.004            |
| <i>Streptococcus</i> sp. HMT066              | Current smoker | -0.50       | 0.0004   | 0.005            |
| <i>Fusobacterium periodonticum hwasookii</i> | Current smoker | -0.48       | 0.0001   | 0.002            |
| <i>Actinomyces</i> sp. HMT169                | Current smoker | -0.47       | 0.00002  | 0.0004           |
| <i>Stomatobaculum</i> sp. HMT097             | Current smoker | -0.41       | 8.53E-07 | 0.00004          |
| <i>Capnocytophaga sputigena</i>              | Current smoker | -0.41       | 0.00002  | 0.0004           |
| <i>Bergeyella</i> sp. HMT322                 | Current smoker | -0.38       | 5.01E-06 | 0.0001           |

|                                                                                 |                |       |          |         |
|---------------------------------------------------------------------------------|----------------|-------|----------|---------|
| <i>Abiotrophia defectiva</i>                                                    | Current smoker | -0.32 | 4.08E-06 | 0.0001  |
| <i>Neisseria oralis</i>                                                         | Current smoker | -0.31 | 0.00002  | 0.0004  |
| <i>Gemella morbillorum</i>                                                      | Current smoker | -0.31 | 0.005    | 0.03    |
| <i>Streptococcus intermedius</i>                                                | Current smoker | -0.31 | 0.0005   | 0.006   |
| <i>Campylobacter concisus</i>                                                   | Current smoker | -0.30 | 0.007    | 0.04    |
| <i>Cardiobacterium hominis</i>                                                  | Current smoker | -0.28 | 0.00001  | 0.0003  |
| <i>Selenomonas artemidis</i>                                                    | Current smoker | -0.24 | 0.004    | 0.03    |
| <i>Neisseria elongata</i>                                                       | Current smoker | -0.21 | 0.008    | 0.05    |
| <i>Capnocytophaga gingivalis</i>                                                | Current smoker | -0.21 | 0.004    | 0.03    |
| <i>Aggregatibacter</i> sp. HMT512                                               | Current smoker | -0.20 | 0.003    | 0.02    |
| <i>Aggregatibacter</i> sp. HMT458                                               | Current smoker | -0.19 | 0.006    | 0.04    |
| <i>Lachnoanaerobaculum umeaense</i>                                             | Current smoker | -0.19 | 0.005    | 0.03    |
| <i>Porphyromonas catoniae</i>                                                   | Current smoker | -0.12 | 0.005    | 0.03    |
| <i>Haemophilus sputorum</i>                                                     | Current smoker | -0.21 | 0.01     | 0.07    |
| <i>Streptococcus</i> sp. HMT064 <i>oralis</i> subsp. <i>dentisani</i> clade 398 | Current smoker | -0.19 | 0.02     | 0.09    |
| <i>Prevotella</i> sp. HMT526                                                    | Current smoker | 0.19  | 0.004    | 0.03    |
| <i>Treponema</i> sp. HMT237                                                     | Current smoker | 0.20  | 0.004    | 0.03    |
| <i>Fusobacterium</i> sp. HMT203                                                 | Current smoker | 0.21  | 0.004    | 0.03    |
| <i>Peptostreptococcaceae</i> XI G6 <i>nodatum</i>                               | Current smoker | 0.22  | 0.006    | 0.04    |
| <i>Streptococcus</i> sp. HMT064 sp. HMT423 <i>mitis</i>                         | Current smoker | 0.23  | 0.006    | 0.04    |
| <i>Streptococcus thermophilus salivarius</i>                                    | Current smoker | 0.23  | 0.0005   | 0.006   |
| <i>Clostridiales</i> F1 G1 <i>bacterium</i> HMT093                              | Current smoker | 0.23  | 0.00001  | 0.0003  |
| <i>Peptostreptococcaceae</i> XI G4 <i>bacterium</i> HMT369                      | Current smoker | 0.27  | 0.0008   | 0.009   |
| <i>Bacteroidales</i> G2 <i>bacterium</i> HMT274                                 | Current smoker | 0.27  | 0.003    | 0.02    |
| <i>Stomatobaculum longum</i>                                                    | Current smoker | 0.32  | 0.0002   | 0.003   |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 070                   | Current smoker | 0.33  | 0.005    | 0.03    |
| <i>Veillonella denticariosi</i>                                                 | Current smoker | 0.33  | 0.005    | 0.03    |
| <i>Fusobacterium nucleatum</i> subsp. <i>vincentii</i>                          | Current smoker | 0.34  | 0.007    | 0.04    |
| <i>Mycoplasma faucium</i>                                                       | Current smoker | 0.34  | 0.0001   | 0.002   |
| <i>Mogibacterium timidum</i>                                                    | Current smoker | 0.34  | 0.0001   | 0.002   |
| <i>Filifactor alocis</i>                                                        | Current smoker | 0.36  | 0.004    | 0.03    |
| <i>Alloprevotella tannerae</i>                                                  | Current smoker | 0.37  | 0.009    | 0.05    |
| <i>Parvimonas micra</i>                                                         | Current smoker | 0.38  | 0.004    | 0.03    |
| <i>Actinomyces lingnae</i> Not Validly Published                                | Current smoker | 0.38  | 0.0006   | 0.007   |
| <i>Prevotella</i> sp. HMT313                                                    | Current smoker | 0.41  | 0.002    | 0.02    |
| <i>Streptococcus</i> sp. HMT057                                                 | Current smoker | 0.41  | 0.002    | 0.02    |
| <i>Porphyromonas endodontalis</i>                                               | Current smoker | 0.43  | 0.0007   | 0.007   |
| <i>Peptostreptococcaceae</i> XI G5 <i>saphenum</i>                              | Current smoker | 0.49  | 4.59E-06 | 0.0001  |
| <i>Actinomyces graevenitzi</i>                                                  | Current smoker | 0.49  | 0.0007   | 0.008   |
| <i>Prevotella pleuritidis</i>                                                   | Current smoker | 0.51  | 1.32E-06 | 0.00005 |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 071                   | Current smoker | 0.52  | 6.79E-06 | 0.0002  |
| <i>Prevotella histicola</i>                                                     | Current smoker | 0.59  | 0.0006   | 0.007   |
| <i>Atopobium parvulum</i>                                                       | Current smoker | 0.60  | 0.00002  | 0.0004  |
| <i>Streptococcus parasanguinis</i> clade 411 <i>parasanguinis</i> clade 721     | Current smoker | 0.74  | 5.75E-07 | 0.00003 |
| <i>Tannerella forsythia</i>                                                     | Current smoker | 0.23  | 0.009    | 0.05    |

|                                                                                                                                                                                  |                |       |         |        |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|-------|---------|--------|
| <i>Veillonella rogosae dispar parvula atypica</i>                                                                                                                                | Current smoker | 0.22  | 0.01    | 0.05   |
| <i>Streptococcus vestibularis salivarius</i>                                                                                                                                     | Current smoker | 0.28  | 0.01    | 0.07   |
| <i>Streptococcus parasanguinis</i> clade 411                                                                                                                                     | Current smoker | 0.36  | 0.02    | 0.09   |
| <i>Peptostreptococcaceae</i> XI G9 <i>brachy</i>                                                                                                                                 | Current smoker | 0.27  | 0.02    | 0.09   |
| <i>Streptococcus gordonii</i>                                                                                                                                                    | Former smoker  | -0.60 | 0.0009  | 0.009  |
| <i>Actinomyces</i> sp. HMT169                                                                                                                                                    | Former smoker  | -0.45 | 0.0003  | 0.004  |
| <i>Streptococcus intermedius</i>                                                                                                                                                 | Former smoker  | -0.40 | 0.00006 | 0.0010 |
| <i>Actinomyces</i> sp. HMT175                                                                                                                                                    | Former smoker  | -0.32 | 0.006   | 0.04   |
| <i>Selenomonas artemidis</i>                                                                                                                                                     | Former smoker  | -0.31 | 0.001   | 0.01   |
| <i>Streptococcus</i> sp. HMT064 <i>oralis</i> subsp. <i>dentisani</i> clade 398                                                                                                  | Former smoker  | -0.24 | 0.01    | 0.05   |
| <i>Cardiobacterium hominis</i>                                                                                                                                                   | Former smoker  | -0.22 | 0.002   | 0.02   |
| <i>Streptococcus australis</i>                                                                                                                                                   | Former smoker  | -0.33 | 0.01    | 0.07   |
| <i>Rothia aeria</i>                                                                                                                                                              | Former smoker  | -0.3  | 0.02    | 0.09   |
| <i>Streptococcus</i> sp. HMT064 <i>australis oralis</i> subsp. <i>dentisani</i> clade 398 sp. HMT423 <i>infantis</i> clade 431 <i>infantis</i> clade 638 <i>mitis pneumoniae</i> | Former smoker  | 0.28  | 0.003   | 0.02   |
| <i>Ruminococcaceae</i> G2 <i>bacterium</i> HMT085                                                                                                                                | Former smoker  | 0.23  | 0.009   | 0.05   |



**Figure 16.** Heatmap showing taxa with significant changes in abundance associated with smoking with effect size calculated as  $-\log(q\text{-value}) \times (\text{coefficient})$ . Colour intensity relates to effect size where increased abundance is indicated in red and reduced abundance in blue (compared to never smokers).

### 6.3.2 Alcohol consumption

Alcohol consumption was not associated with any significant shifts in either alpha or beta diversity ( $p=0.38$ ) (**Figure 17**) and MaAsLin2 analysis confirmed that alcohol consumption alone had no influence on the microbial colonisation of normal mucosal surfaces.

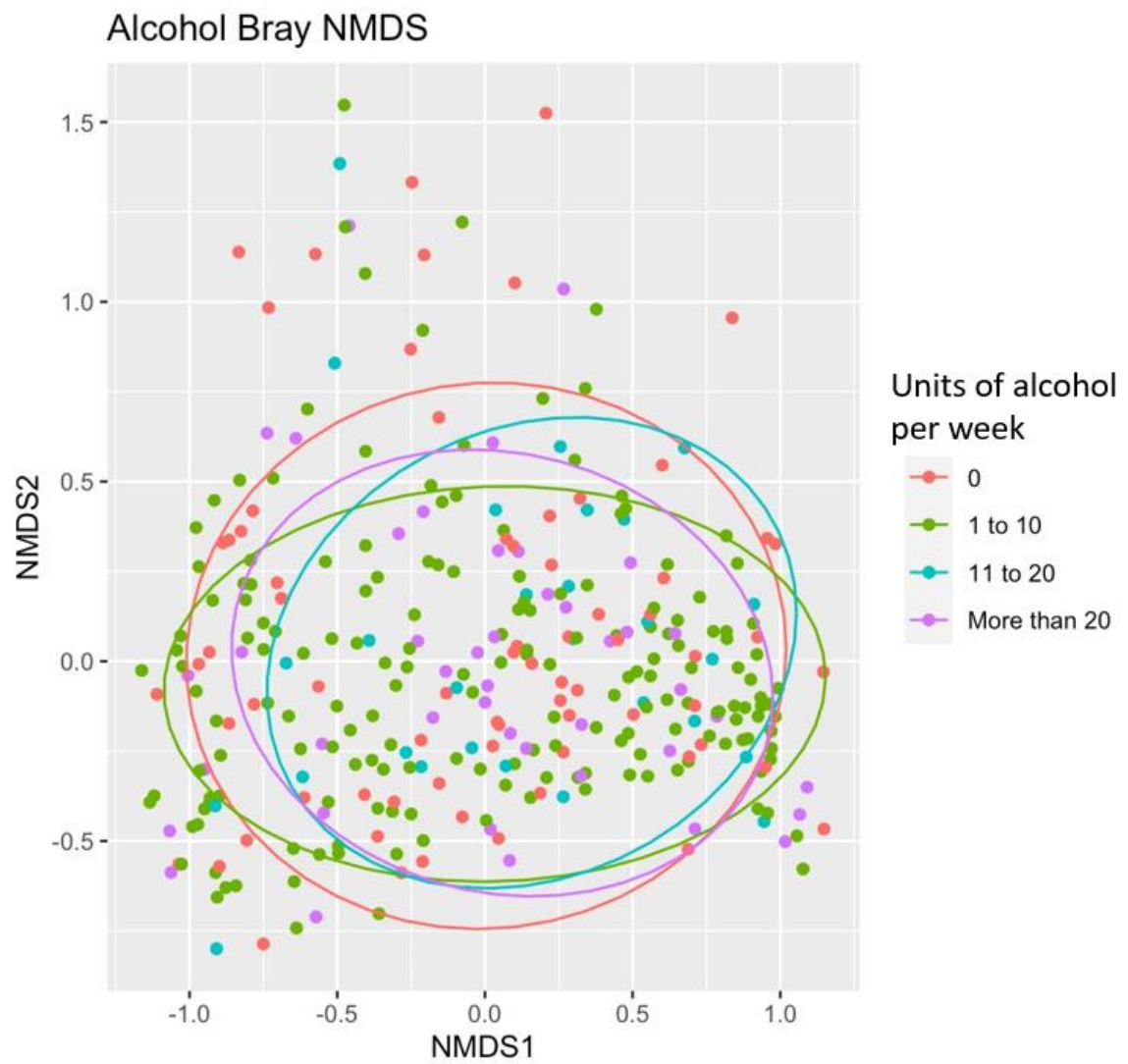


**Figure 17.** Bray Curtis NMDS plot of categories of alcohol consumption (Alcohol bin=units/week) ( $p=0.38$ )



### 6.3.3 Tooth brushing

Toothbrushing habits were found to influence the alpha diversity with brushing less than once per day associated with reduced biodiversity indicated by significantly reduced Simpson and Shannon index values (**Figure 17**).

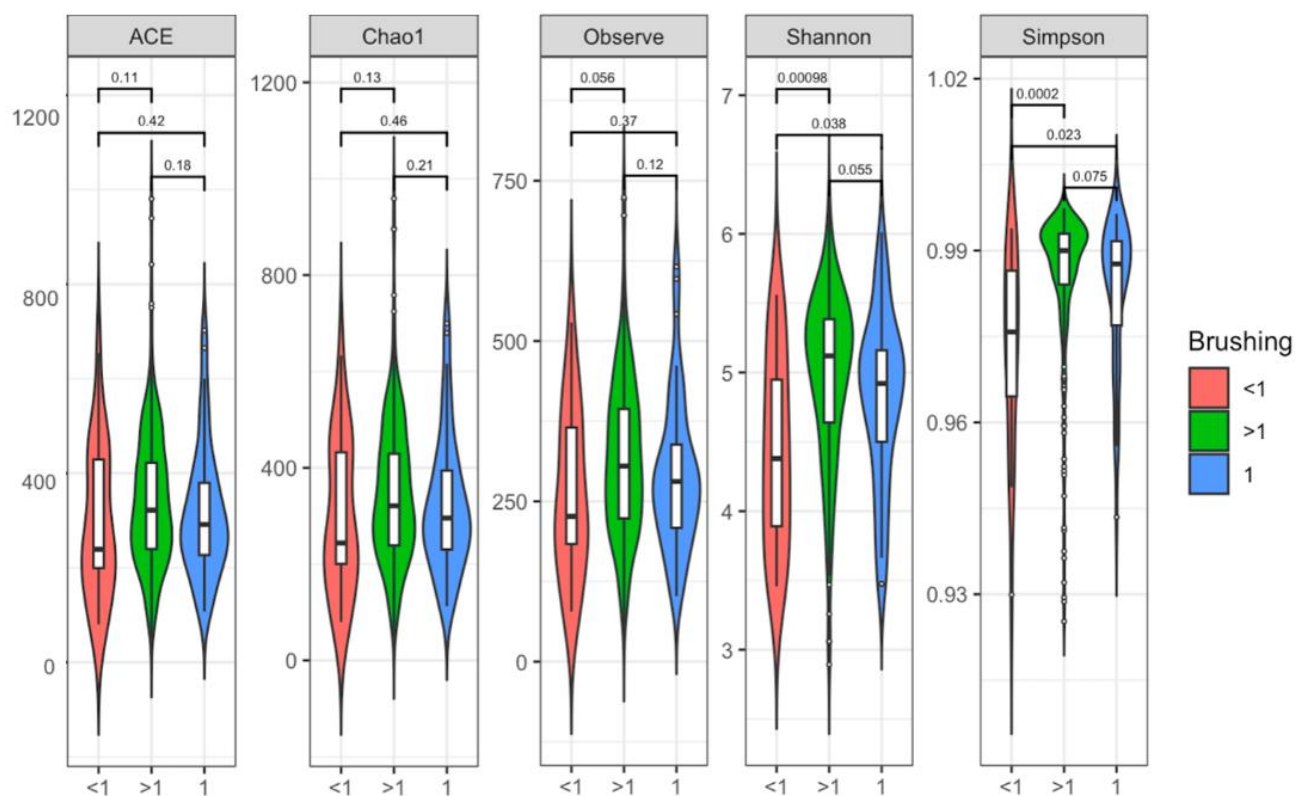


**Figure 18.** Alpha diversity plots by daily tooth brushing, p values denoted by numbers above square brackets. (<1=Brushing less than once per day, 1=Brushing once per day, >1=Brushing more than once per day)

Visualisation of the Bray-Curtis NMDS plot of tooth brushing habits demonstrated separation of samples where brushing occurred more than once per day from those where brushing occurred once or less than once per day, with PERMANOVA testing confirming this separation was significant ( $p=0.03$ ) (**Figure 19**). Pairwise PERMANOVA tests showed significant differences in the microbial communities of those that brushed once compared

with more than once per day ( $p=0.046$ ), and more than once compared with less than once per day ( $p=0.03$ ), although the adjusted  $p$  values were not significant for any of the comparisons (**Table 27**).

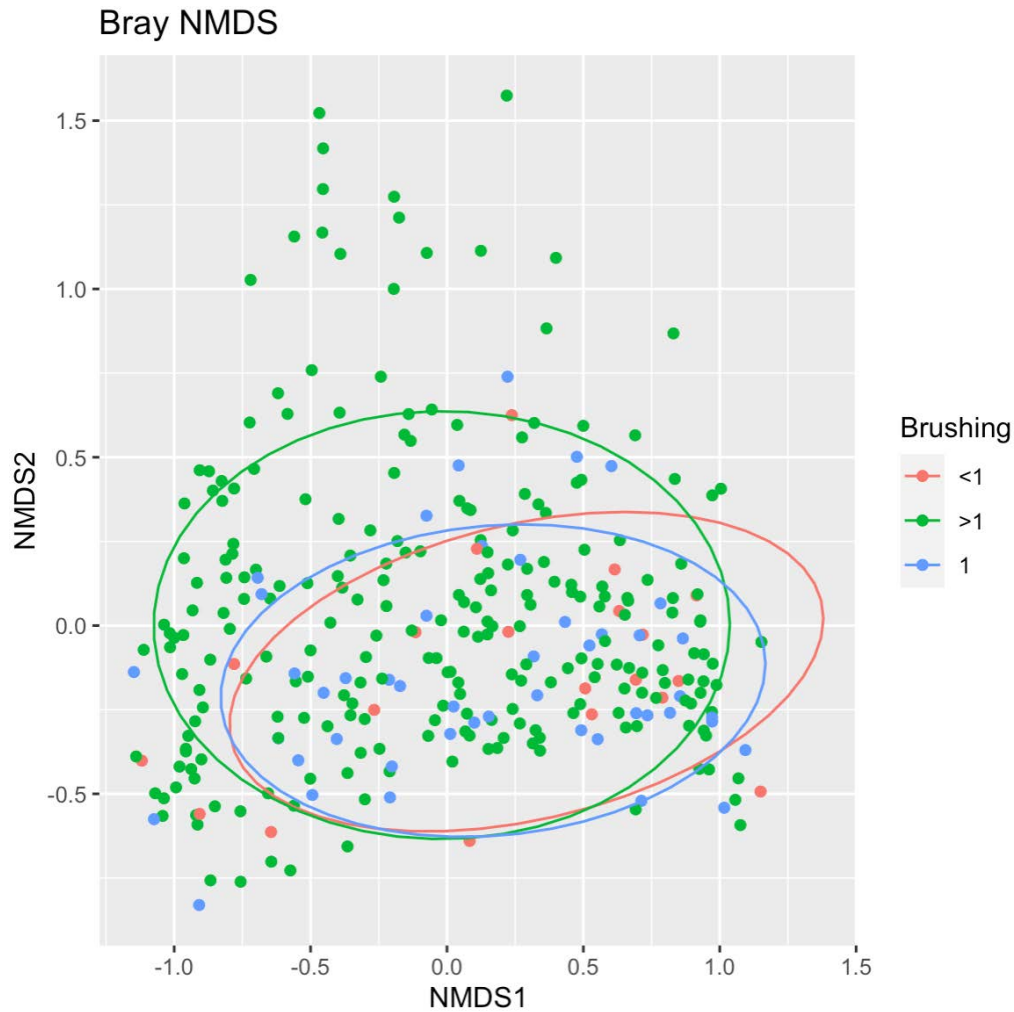


**Figure 19.** Bray-Curtis NMDS plot of brushing habits: <1=less than once per day, 1=once per day, >1=more than once per day ( $p=0.03$ )

**Table 27.** Pairwise PERMANOVA results for toothbrushing ( $*p<0.05$ )

| Comparison                    | p value | Adjusted p value |
|-------------------------------|---------|------------------|
| Brushing once vs > once/day   | 0.046*  | 0.14             |
| Brushing once vs < once/day   | 0.34    | 0.5              |
| Brushing > once vs < once/day | 0.03*   | 0.14             |

The brushing categories of less than once per day and once per day were combined for MaAsLin2 analysis due to the small number of samples where brushing was recorded as occurring less than once per day (n=19). Brushing once or less than once per day was associated with significantly reduced abundance of *Kingella oralis*, *Oribacterium sinus*, *Streptococcus* sp. HMT066 and *Rothia dentocariosa* and significantly increased abundance of *Granulicatella elegans* (all padj<0.09) (Table 28).

**Table 28.** Taxa significantly associated with brushing once or less per day on MaAsLin2 analysis (negative values denote reduced abundance)

| Taxa                            | Brushing/day | Coefficient | p value | Adjusted p value |
|---------------------------------|--------------|-------------|---------|------------------|
| <i>Kingella oralis</i>          | Once/less    | -0.51       | 0.0003  | 0.03             |
| <i>Oribacterium sinus</i>       | Once/less    | -0.45       | 0.0003  | 0.03             |
| <i>Streptococcus</i> sp. HMT066 | Once/less    | -0.51       | 0.0007  | 0.04             |
| <i>Rothia dentocariosa</i>      | Once/less    | -0.40       | 0.002   | 0.08             |
| <i>Granulicatella elegans</i>   | Once/less    | 0.33        | 0.001   | 0.06             |

### 6.3.4 Oral hygiene

There were no differences in alpha diversity measures between the different oral hygiene categories. It appeared that there may be some shifts in population structure related to oral hygiene on the NMDS plot although the differences were not significant on pairwise PERMANOVA testing (**Figure 20, Table 29**).

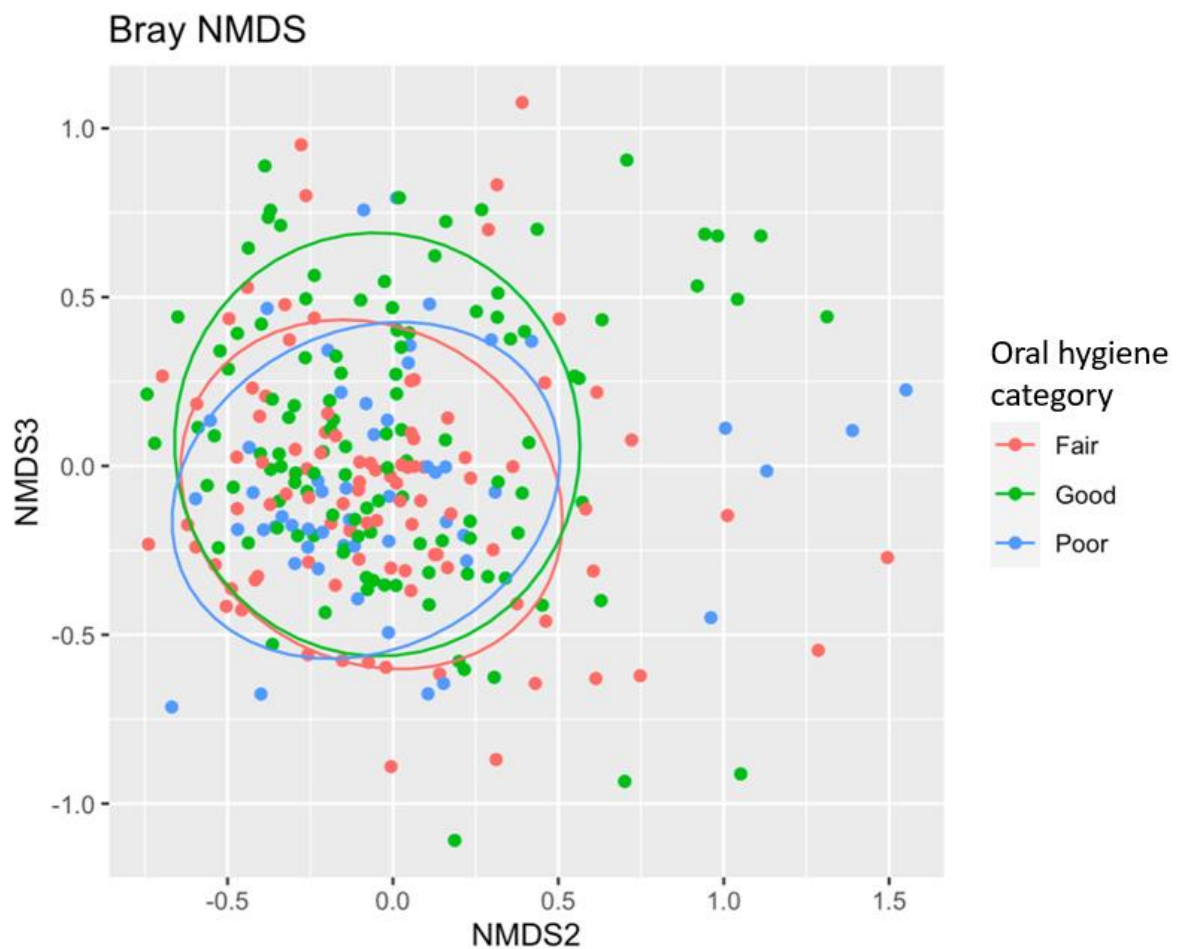


**Figure 20.** Bray-Curtis NMDS plot of oral hygiene category (Good=oral hygiene index score (OHI-S)<0.5, Fair=OHI-S 0.5-1, Poor=OHI-S>1) (p=0.3)

**Table 29.** Results of pairwise PERMANOVA for oral hygiene categories

| Comparison                | p value | Adjusted p value |
|---------------------------|---------|------------------|
| Fair vs Good oral hygiene | 0.19    | 0.36             |
| Fair vs Poor oral hygiene | 0.48    | 0.48             |
| Good vs Poor oral hygiene | 0.13    | 0.36             |

However, MaAsLin2 analysis demonstrated differences in abundance of several taxa between the oral hygiene categories. Poor oral hygiene was associated with significantly increased abundance of several plaque and periodontal disease associated pathogens including *Prevotella denticola*, *Prevotella oulorum*, another *Prevotella* species, and a single species each of genera *Leptotrichia*, *Saccharibacteria*, *Lachnospiraceae*, *Tannerella* and *Streptococcus*, along with reduced abundance of *Streptococcus sanguinis* and *Haemophilus parainfluenzae*, although none of these results were significant after adjustment (all  $p_{adj} > 0.1$ ). Fair oral hygiene was associated with increased abundance of three *Streptococcus* spp. and reduced abundance of *Kingella oralis* and a single species each of the genera *Actinomyces*, *Haemophilus* and *Stomatobaculum*, although again, the adjusted p values were not significant (Table 30, Figure 21).

**Table 30.** Taxa with significant ( $p < 0.05$ ) changes in abundance associated with oral hygiene on MaAsLin2 analysis (negative values denote reduced abundance)

| Taxa                                                            | Oral hygiene category | Coefficient | p value | Adjusted p value |
|-----------------------------------------------------------------|-----------------------|-------------|---------|------------------|
| <i>Streptococcus thermophilus salivarius</i>                    | Fair                  | 0.20        | 0.002   | 0.17             |
| <i>Streptococcus australis infantis</i> clade 431               | Fair                  | 0.35        | 0.004   | 0.18             |
| <i>Streptococcus cristatus</i> clade 578                        | Fair                  | 0.35        | 0.006   | 0.19             |
| <i>Kingella oralis</i>                                          | Fair                  | -0.42       | 0.004   | 0.17             |
| <i>Actinomyces</i> sp. HMT169                                   | Fair                  | -0.38       | 0.002   | 0.17             |
| <i>Haemophilus paraphrohaemolyticus</i> <i>parahaemolyticus</i> | Fair                  | -0.19       | 0.006   | 0.19             |
| <i>Stomatobaculum</i> sp. HMT097                                | Fair                  | -0.25       | 0.007   | 0.20             |
| <i>Prevotella</i> sp. HMT313                                    | Poor                  | 0.50        | 0.003   | 0.17             |
| <i>Prevotella denticola</i>                                     | Poor                  | 0.48        | 0.002   | 0.17             |
| <i>Leptotrichia</i> sp. HMT498                                  | Poor                  | 0.35        | 0.003   | 0.17             |
| <i>Saccharibacteria</i> TM7 G 1 <i>bacterium</i> HMT346         | Poor                  | 0.31        | 0.002   | 0.17             |
| <i>Lachnospiraceae</i> G3 <i>bacterium</i> HMT100               | Poor                  | 0.33        | 0.007   | 0.19             |
| <i>Tannerella</i> sp. HMT286                                    | Poor                  | 0.33        | 0.008   | 0.20             |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 071   | Poor                  | 0.37        | 0.01    | 0.24             |
| <i>Prevotella oulorum</i>                                       | Poor                  | 0.31        | 0.01    | 0.24             |
| <i>Streptococcus sanguinis</i>                                  | Poor                  | -0.45       | 0.002   | 0.17             |
| <i>Haemophilus parainfluenzae</i>                               | Poor                  | -0.44       | 0.006   | 0.19             |

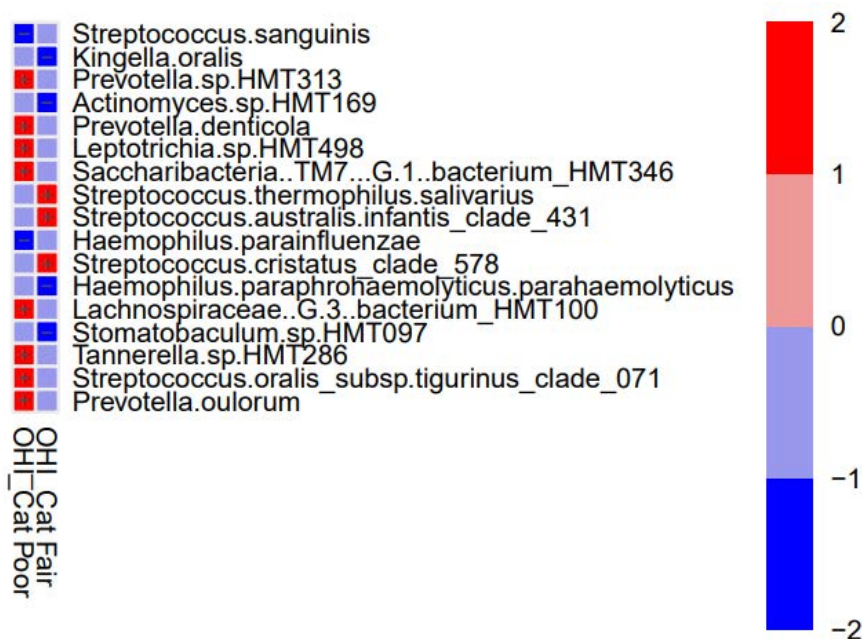


**Figure 21.** Heatmap showing taxa with significant changes in abundance associated with poor (OHI\_Cat Poor) and fair (OHI\_Cat Fair) oral hygiene. (Effect size calculated as  $-\log(q\text{-value}) * (\text{coefficient})$ . Red=increased, blue=reduced relative to good oral hygiene (OHI-S<0.5))

### 6.3.5 Denture wear

No differences in alpha diversity were noted between no, partial or complete denture wear on ACE or Chao1 indices, reflecting their similar richness. However, no denture wear differed significantly from partial denture wear on Observe, Shannon and Simpson indices ( $p=0.03$ ,  $p=0.0003$  and  $p=9.2E-0.5$  respectively) and from complete denture wear on Simpson and Shannon indices ( $p=0.03$  and  $p=0.008$  respectively). Visualisation of the Bray-Curtis NMDS plot of denture wear demonstrated significant shifts in population structure ( $p=0.007$ ) (**Figure 22**). Pairwise PERMANOVA demonstrated that there were significant differences between the oral mucosal microbial communities of partial denture wearers compared to those who did not wear dentures ( $p=0.005$ ,  $\text{padj}<0.01$ ) (**Table 31**). There were

only nine complete denture wearers in the sample cohort which may explain why there appeared to be visual separation of this group on the NMDS plot but there was no significant difference on pairwise PERMANOVA testing.

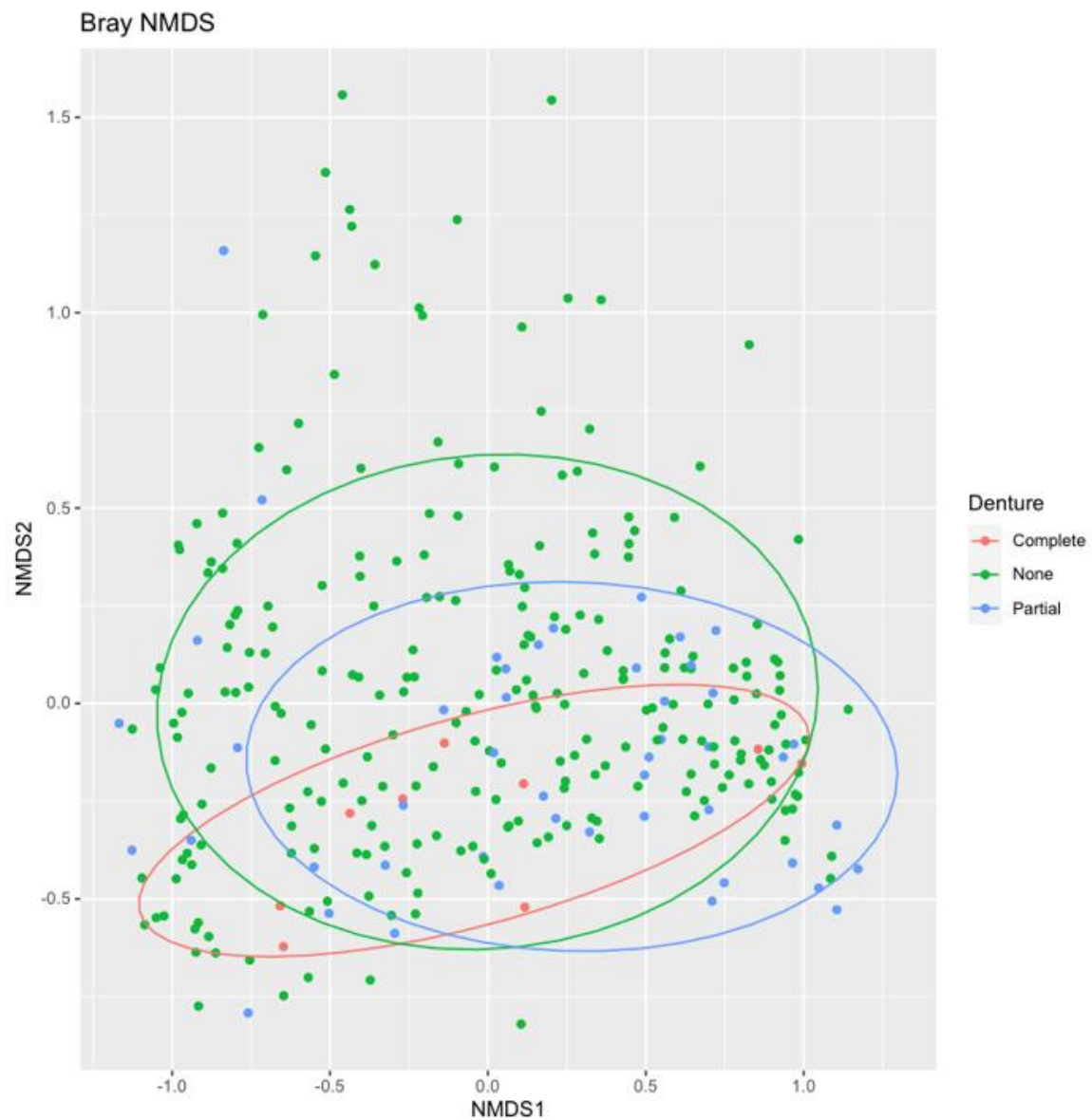


**Figure 22.** Bray-Curtis NMDS plot of denture wear ( $p=0.007$ )



**Table 31.** Pairwise PERMANOVA results for denture wear (\*p<0.05, padj<0.1)

| Comparison                  | p value | Adjusted p value |
|-----------------------------|---------|------------------|
| No vs Partial denture       | 0.005*  | 0.01*            |
| No vs Complete denture      | 0.11    | 0.16             |
| Partial vs Complete denture | 0.17    | 0.17             |

MaAsLin2 analysis demonstrated that complete denture wear was associated with reduced abundance of *Rothia dentocariosa* (p=0.00005, padj=0.02) and several other species of *Streptococcus*, *Parvimonas*, *Fusobacterium* and *Alloprevotella* genera, although only the adjusted p value for *Rothia dentocariosa* reached significance. Wearing a partial denture was associated with reduced abundance of *Streptococcus sanguinis*, *Lautropia mirabilis*, *Haemophilus parainfluenzae*, *Corynebacterium durum*, *Rothia aeria*, *Filifactor alocis* and an *Actinomyces* species and increased abundance of *Streptococcus infantis clade 431 mitis*, although the adjusted p values were not significant (**Table 32, Figure 23**).

**Table 32.** Results of MaAsLin2 analysis showing taxa with significant changes in abundance associated with denture wear (negative values denote reduced abundance) (\*=padj<0.1)

| Taxa                                                             | Denture  | Coefficient | p value | Adjusted p value |
|------------------------------------------------------------------|----------|-------------|---------|------------------|
| <i>Rothia dentocariosa</i>                                       | Complete | -1.23       | 0.00005 | 0.02*            |
| <i>Streptococcus</i> sp. HMT064 sp. HMT423                       | Complete | 0.39        | 0.007   | 0.23             |
| <i>Streptococcus sanguinis</i>                                   | Complete | -0.89       | 0.002   | 0.19             |
| <i>Parvimonas micra</i>                                          | Complete | -0.95       | 0.004   | 0.19             |
| <i>Fusobacterium nucleatum</i> subsp. <i>vincentii naviforme</i> | Complete | -0.96       | 0.004   | 0.19             |
| <i>Streptococcus cristatus clade 578</i>                         | Complete | -0.80       | 0.007   | 0.23             |
| <i>Alloprevotella tanneriae</i>                                  | Complete | -0.94       | 0.007   | 0.23             |
| <i>Streptococcus infantis clade 431 mitis</i>                    | Partial  | 0.29        | 0.003   | 0.19             |
| <i>Streptococcus sanguinis</i>                                   | Partial  | -0.47       | 0.001   | 0.10             |
| <i>Actinomyces</i> sp. HMT169                                    | Partial  | -0.39       | 0.004   | 0.19             |
| <i>Lautropia mirabilis</i>                                       | Partial  | -0.38       | 0.005   | 0.22             |
| <i>Haemophilus parainfluenzae</i>                                | Partial  | -0.41       | 0.007   | 0.23             |
| <i>Filifactor alocis</i>                                         | Partial  | -0.39       | 0.009   | 0.23             |
| <i>Corynebacterium durum</i>                                     | Partial  | -0.21       | 0.008   | 0.23             |
| <i>Rothia aeria</i>                                              | Partial  | -0.36       | 0.01    | 0.24             |



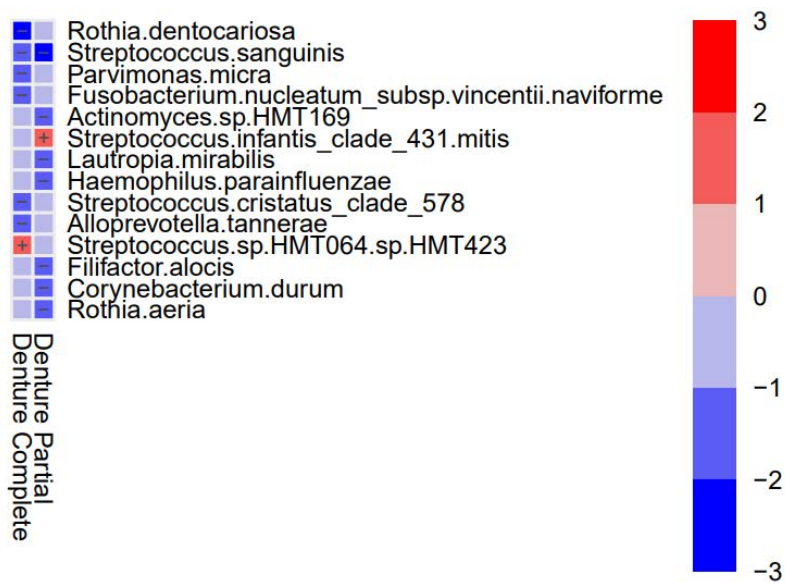


**Figure 23.** Heatmap showing taxa with significant changes in abundance associated with denture wear. (Effect size calculated as  $-\log(q\text{-value}) * (\text{coefficient})$ ). Red=increased, blue=reduced relative to no denture wear)

### 6.3.6 Mouthwash use

Alpha diversity measures did not differ between mouthwash users and non-users. However, significant differences in beta-diversities, measured using Bray-Curtis dissimilarity values, were noted ( $p=0.03$ ) (**Figure 23**).

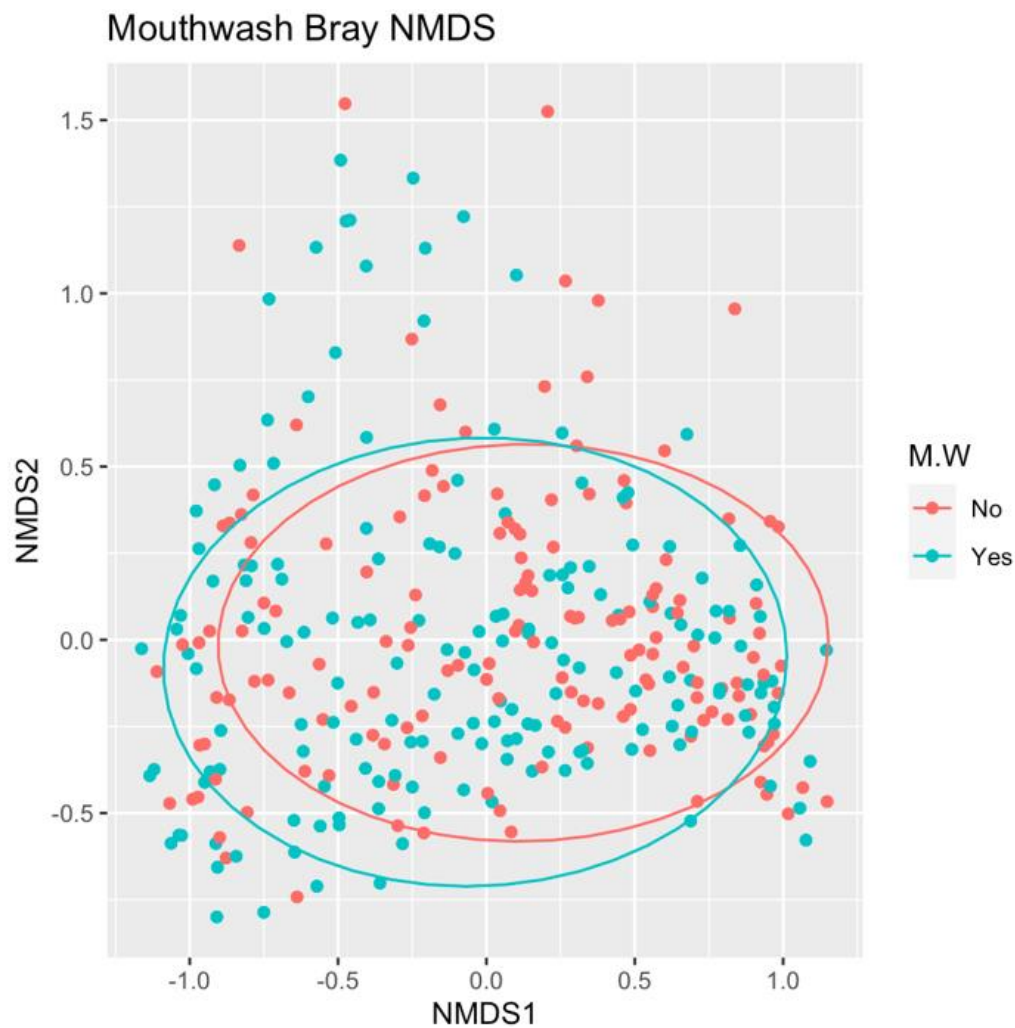


**Figure 24.** Bray-Curtis NMDS plot of mouthwash use ( $p=0.03$ ) (M.W=mouthwash)

MaAsLin2 analysis showed that this difference was largely due to reduced abundance *Streptococcus mitis pneumoniae* ( $p=0.0001$ ,  $\text{padj}=0.002$ ) in mouthwash users. *Granulicatella adiacens* ( $p=0.002$ ,  $\text{padj}=0.2$ ) and *Rothia mucilaginosa* ( $p=0.0004$ ,  $\text{padj}=0.23$ )

appeared to show increased abundance in mouthwash users, although the adjusted p values were not significant.

### 6.3.7 Missing teeth

Tooth number was found to have a significant influence on the microbial population of mucosal surfaces. Communities missing less than five teeth and missing five to 15 teeth differed significantly from those missing more than 15 teeth on all five alpha diversity indices, reflecting differences in both richness and biodiversity between these groups. Richness was lowest where more than 15 teeth were missing, reflecting loss of species (**Table 33**)

**Table 33.** Table showing p values from comparisons of alpha diversities of communities missing 5, 5 to 15 and more than 15 teeth. Significant results in bold.

| Comparison                   | ACE          | Chao1        | Observe      | Simpson       | Shannon       |
|------------------------------|--------------|--------------|--------------|---------------|---------------|
| <5 v 5 to 15 teeth missing   | 0.76         | 0.79         | 0.76         | 0.55          | 0.45          |
| <5 v > 15 teeth missing      | <b>0.01</b>  | <b>0.01</b>  | <b>0.01</b>  | <b>0.01</b>   | <b>0.02</b>   |
| 5 to 15 v > 15 teeth missing | <b>0.003</b> | <b>0.003</b> | <b>0.002</b> | <b>0.0006</b> | <b>0.0007</b> |

These differences were also apparent on beta-diversity analysis with an NMDS plot showing visible separation of those missing more than 15 teeth from those missing less than this number. PERMANOVA testing confirmed that this separation was significant (p=0.02). Pairwise PERMANOVA testing demonstrated significant differences between all categories of missing teeth (**Figure 25, Table 34**).

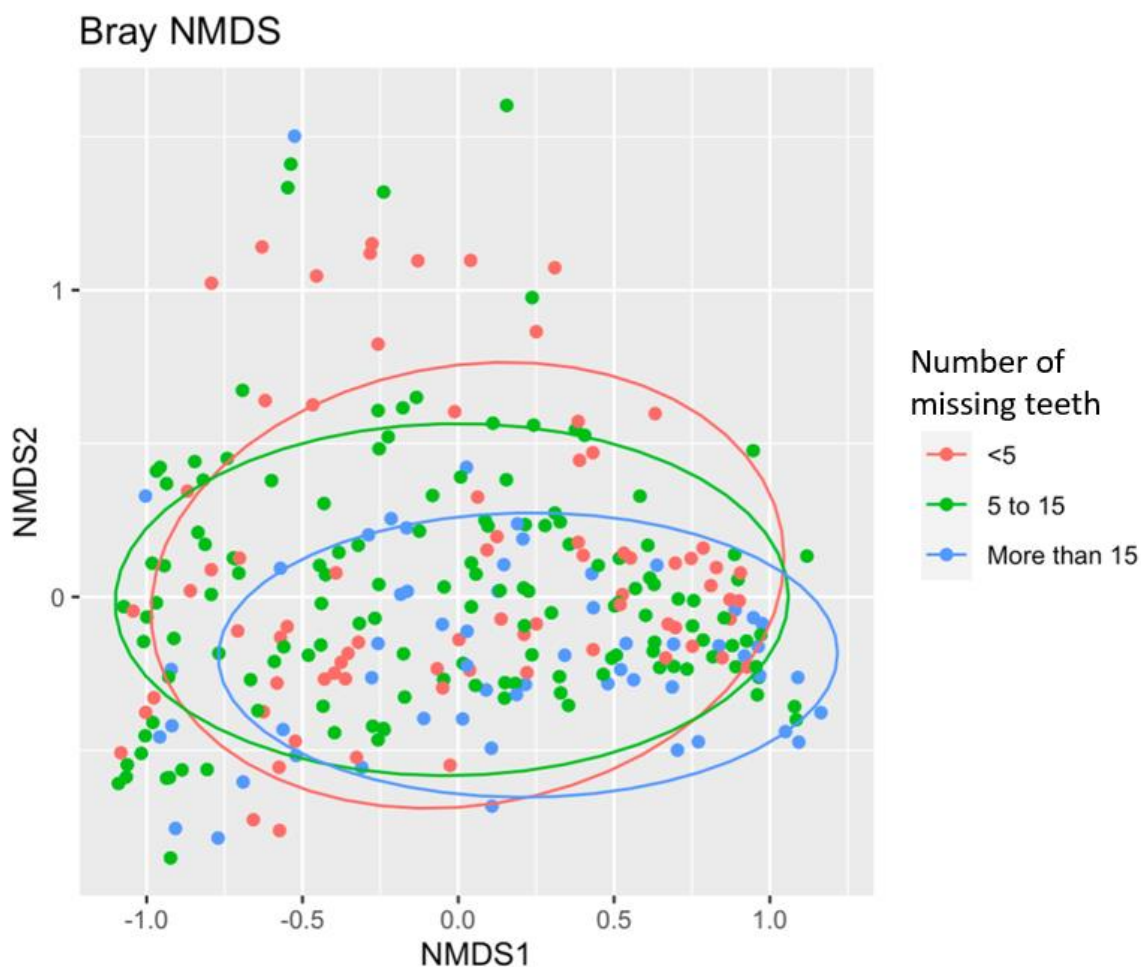


**Figure 25.** Bray-Curtis NMDS plot of population structure by number of missing teeth (Missing bin=number of missing teeth) ( $p=0.02$ )

**Table 34.** Pairwise PERMANOVA results for missing teeth.

| Comparison                   | p value | Adjusted p value |
|------------------------------|---------|------------------|
| <5 vs 5 - 15 teeth missing   | 0.04    | 0.05             |
| <5 vs >15 teeth missing      | 0.01    | 0.03             |
| 5 - 15 vs > 15 teeth missing | 0.04    | 0.05             |

Missing more than 15 teeth was associated with significantly reduced abundance of *Haemophilus sputorum*, *Streptococcus sanguinis*, two *Actinomyces* spp., *Rothia aeria*, *Rothia dentocariosa*, a *Fusobacterium nucleatum* subsp., *Haemophilus parainfluenzae*, an *Alloprevotella* sp., three streptococcal species, *Corynebacterium durum*, *Lautropia mirabilis* and *Peptostreptococcus stomatis* (all  $p_{adj}<0.1$ ) and significantly increased abundance of two

streptococcal species (both  $p_{adj} < 0.1$ ). Missing five to 15 teeth was significantly associated with increased abundance of three streptococcal species, a *Saccharibacteria* sp., *Eikenella corrodens*, a *Fretibacterium* sp., *Parvimonas micra* and a *Treponema* species, and reduced abundance of *Streptococcus australis*, *Rothia aeria*, *Haemophilus haemolyticus*, another *Haemophilus* sp. and an *Alloprevotella* sp. (all  $p_{adj} < 0.1$ ) (**Table 35**).

**Table 35.** Results of MaAsLin2 analysis showing taxa with significant changes in abundance associated with number of missing teeth (negative coefficient values denote reduced abundance).

| Taxa                                                                        | Missing teeth | Coefficient | p value  | Adjusted p value |
|-----------------------------------------------------------------------------|---------------|-------------|----------|------------------|
| <i>Streptococcus</i> sp. HMT064 sp. HMT423                                  | >15           | 0.24        | 0.003    | 0.07             |
| <i>Streptococcus infantis</i> clade 431 mitis                               | >15           | 0.32        | 0.006    | 0.09             |
| <i>Haemophilus sputorum</i>                                                 | >15           | -0.53       | 3.47E-06 | 0.001            |
| <i>Streptococcus sanguinis</i>                                              | >15           | -0.64       | 7.29E-05 | 0.01             |
| <i>Actinomyces</i> sp. HMT169                                               | >15           | -0.53       | 0.001    | 0.04             |
| <i>Rothia aeria</i>                                                         | >15           | -0.52       | 0.002    | 0.05             |
| <i>Fusobacterium nucleatum</i> subsp. <i>vincentii</i> naviforme            | >15           | -0.57       | 0.001    | 0.05             |
| <i>Alloprevotella</i> sp. HMT914                                            | >15           | -0.27       | 0.002    | 0.05             |
| <i>Haemophilus parainfluenzae</i>                                           | >15           | -0.51       | 0.002    | 0.06             |
| <i>Streptococcus</i> sp. HMT061                                             | >15           | -0.48       | 0.002    | 0.06             |
| <i>Streptococcus</i> sp. HMT066 sp. HMT074                                  | >15           | -0.59       | 0.003    | 0.07             |
| <i>Rothia dentocariosa</i>                                                  | >15           | -0.48       | 0.005    | 0.08             |
| <i>Peptostreptococcus stomatis</i>                                          | >15           | -0.38       | 0.005    | 0.08             |
| <i>Corynebacterium durum</i>                                                | >15           | -0.26       | 0.005    | 0.08             |
| <i>Streptococcus australis</i>                                              | >15           | -0.46       | 0.007    | 0.09             |
| <i>Lautropia mirabilis</i>                                                  | >15           | -0.43       | 0.007    | 0.09             |
| <i>Actinomyces</i> sp. HMT175                                               | >15           | -0.40       | 0.007    | 0.09             |
| <i>Streptococcus constellatus intermedius</i>                               | 5 to 15       | 0.63        | 8.35E-06 | 0.001            |
| <i>Streptococcus parasanguinis</i> clade 411 <i>parasanguinis</i> clade 721 | 5 to 15       | 0.59        | 0.0002   | 0.02             |
| <i>Streptococcus parasanguinis</i> clade 721                                | 5 to 15       | 0.51        | 0.0005   | 0.03             |
| <i>Saccharibacteria TM7 G1 bacterium</i> HMT349                             | 5 to 15       | 0.26        | 0.0008   | 0.04             |
| <i>Eikenella corrodens</i>                                                  | 5 to 15       | 0.27        | 0.003    | 0.07             |
| <i>Fretibacterium</i> sp. HMT359                                            | 5 to 15       | 0.21        | 0.005    | 0.08             |
| <i>Parvimonas micra</i>                                                     | 5 to 15       | 0.38        | 0.005    | 0.09             |
| <i>Haemophilus sputorum</i>                                                 | 5 to 15       | -0.40       | 6.53E-06 | 0.001            |
| <i>Streptococcus australis</i>                                              | 5 to 15       | -0.44       | 0.0009   | 0.04             |
| <i>Haemophilus paraphrohaemolyticus parahaemolyticus</i>                    | 5 to 15       | -0.21       | 0.004    | 0.07             |
| <i>Rothia aeria</i>                                                         | 5 to 15       | -0.36       | 0.004    | 0.08             |
| <i>Haemophilus haemolyticus</i>                                             | 5 to 15       | -0.28       | 0.007    | 0.09             |
| <i>Treponema</i> sp. HMT237                                                 | 5 to 15       | 0.19        | 0.006    | 0.09             |
| <i>Alloprevotella</i> sp. HMT914                                            | 5 to 15       | -0.18       | 0.007    | 0.09             |

### 6.3.8 Site

Mucosal site was found to have a profound influence on oral mucosal microbial communities.

Significant differences were found between the alpha diversity measures of Buc/Alv/Gin and DT/LBT sites on ACE, Chao1 and Observe indices reflecting differences in the bacterial richness between these sites. Significant differences were also noted between FOM/VT and PAL and DT/LBT sites on all five indices used, demonstrating significant differences in both the number of different bacteria at each site and their relative abundance. Buc/Alv/Gin differed from FOM/VT sites on only the Simpson index and no differences in alpha diversity were noted between Buc/Alv/Gin and PAL or between PAL and DT/LBT sites (**Figure 26, Figure 27**).

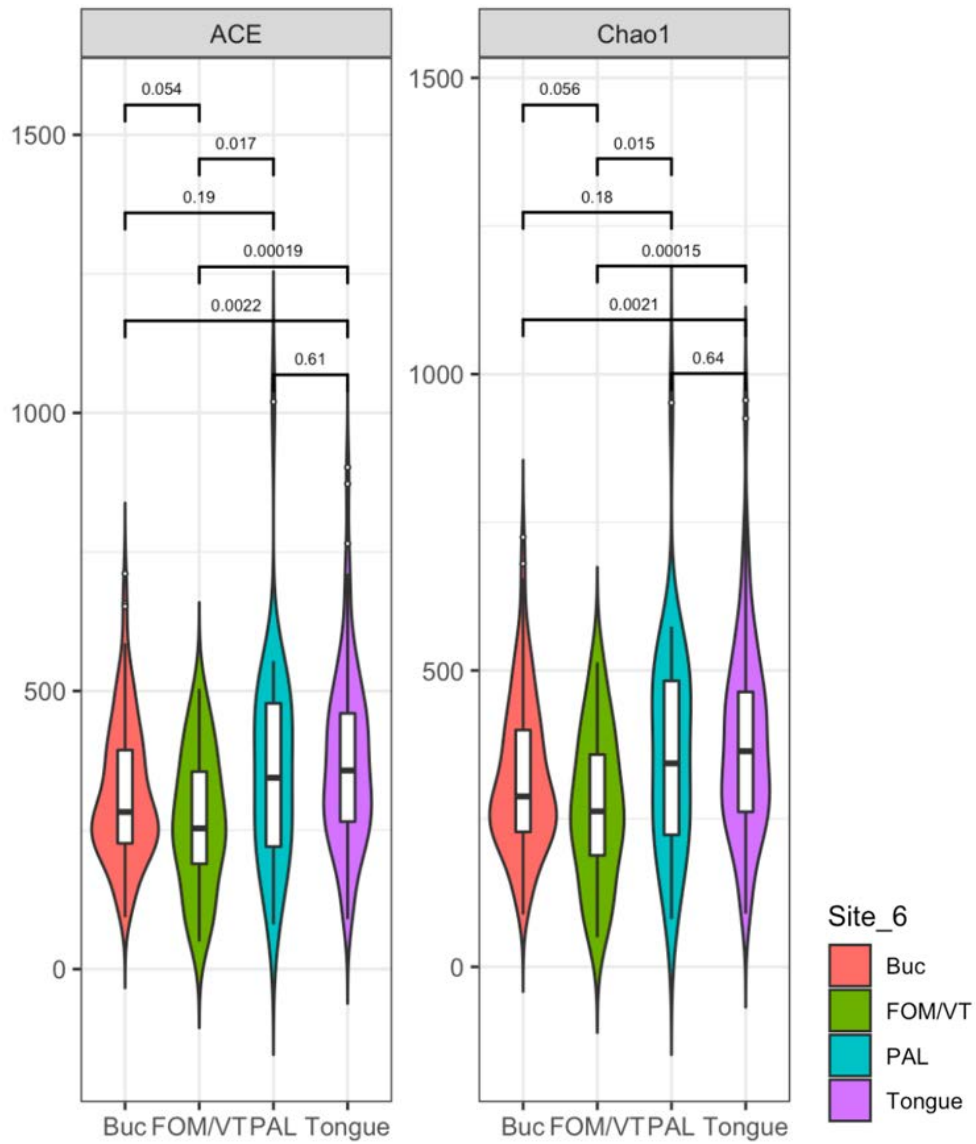


**Figure 26.** Alpha diversity plots by site using ACE and Chao1 indices, p value denoted by number above square brackets. (Buc=Buccal mucosa/alveolus/gingiva, FOM/VT=Floor of mouth/ventral tongue, PAL=Palate, Tongue=DT/LBT)



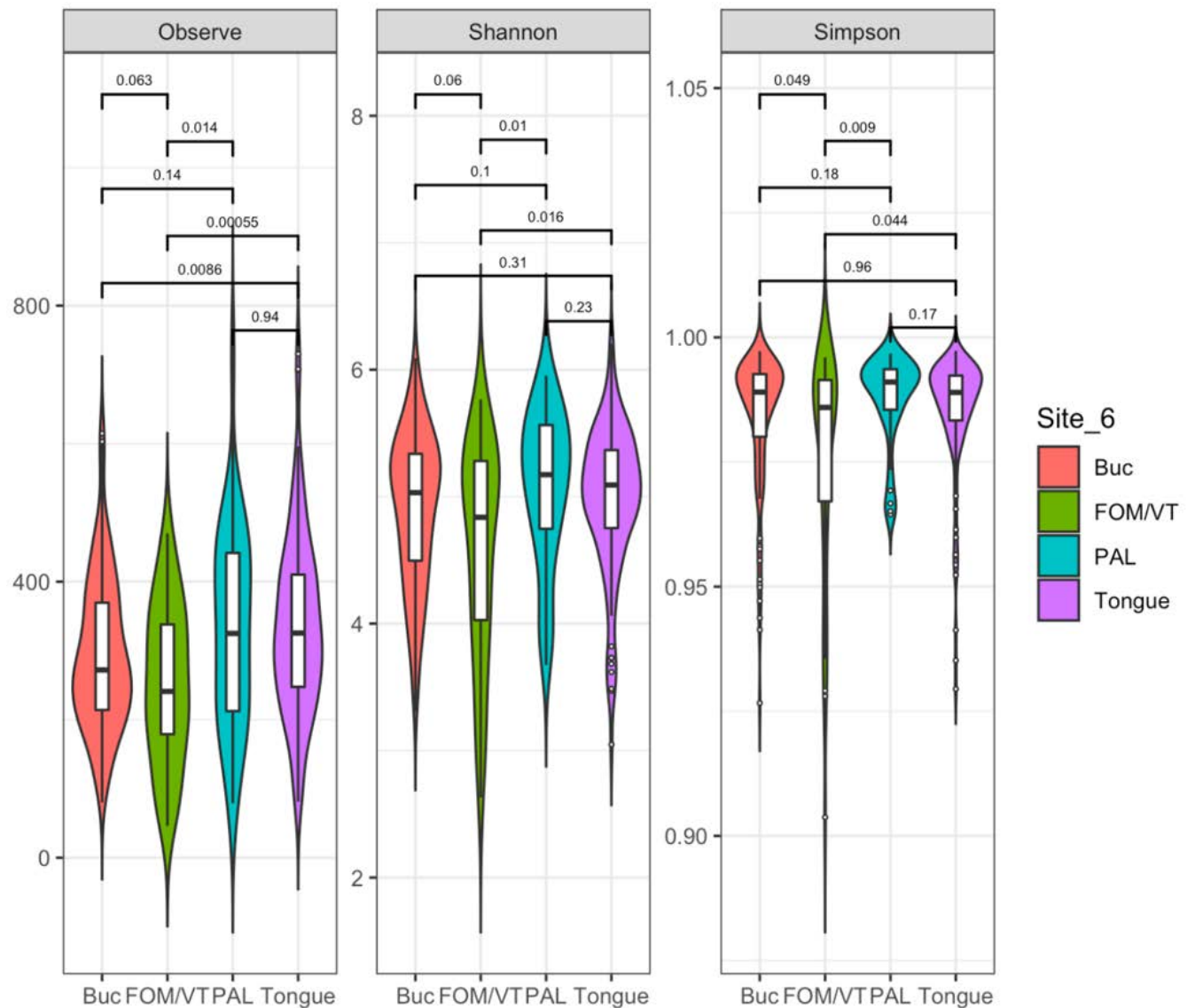


**Figure 27.** Alpha diversity plots by site using Observe, Shannon and Simpson indices, p value denoted by number above square brackets. (Buc=Buccal mucosa/alveolus/gingiva, FOM/VT=Floor of mouth/ventral tongue, PAL=Palate, Tongue=DT/LBT)

PCA (principal component analysis) plots of the relative abundance data by site, with genera responsible for populations shifts highlighted, were generated (**Figure 28, Figure 29**). Visualisation of the plots showed clear separation of DT/LBT and PAL sites while Buc/Alv/Gin and FOM/VT sites appeared to have similar colonisation patterns. Separation of DT/LBT sites appeared to be due to increased abundance of *Rothia* and *Neisseria* species, of PAL sites to increased abundance of *Prevotella* species, and of Buc/Alv/Gin and

FOM/VT sites by increased abundance of streptococcal species. Inter-site comparisons using pairwise PERMANOVA confirmed these differences, with significant differences found between the mucosal microbial communities of the tongue (DT/LBT) and (i) PAL ( $p=0.004$ ,  $p_{adj}=0.004$ ), (ii) Buc/Alv/Gin ( $p=0.0001$ ,  $p_{adj}=0.0002$ ) and (iii) FOM/VT ( $p=0.001$ ,  $p_{adj}=0.002$ ). Palatal communities also differed significantly from Buc/Alv/Gin ( $p=0.001$ ,  $p_{adj}=0.002$ ) and from FOM/VT ( $p=0.002$ ,  $p_{adj}=0.002$ ). No significant differences were found between the beta diversities of Buc/Alv/Gin and FOM/VT sites ( $p=0.13$ ,  $p_{adj}=0.13$ ) agreeing with the visualisation of the PCA plots (**Table 35**).

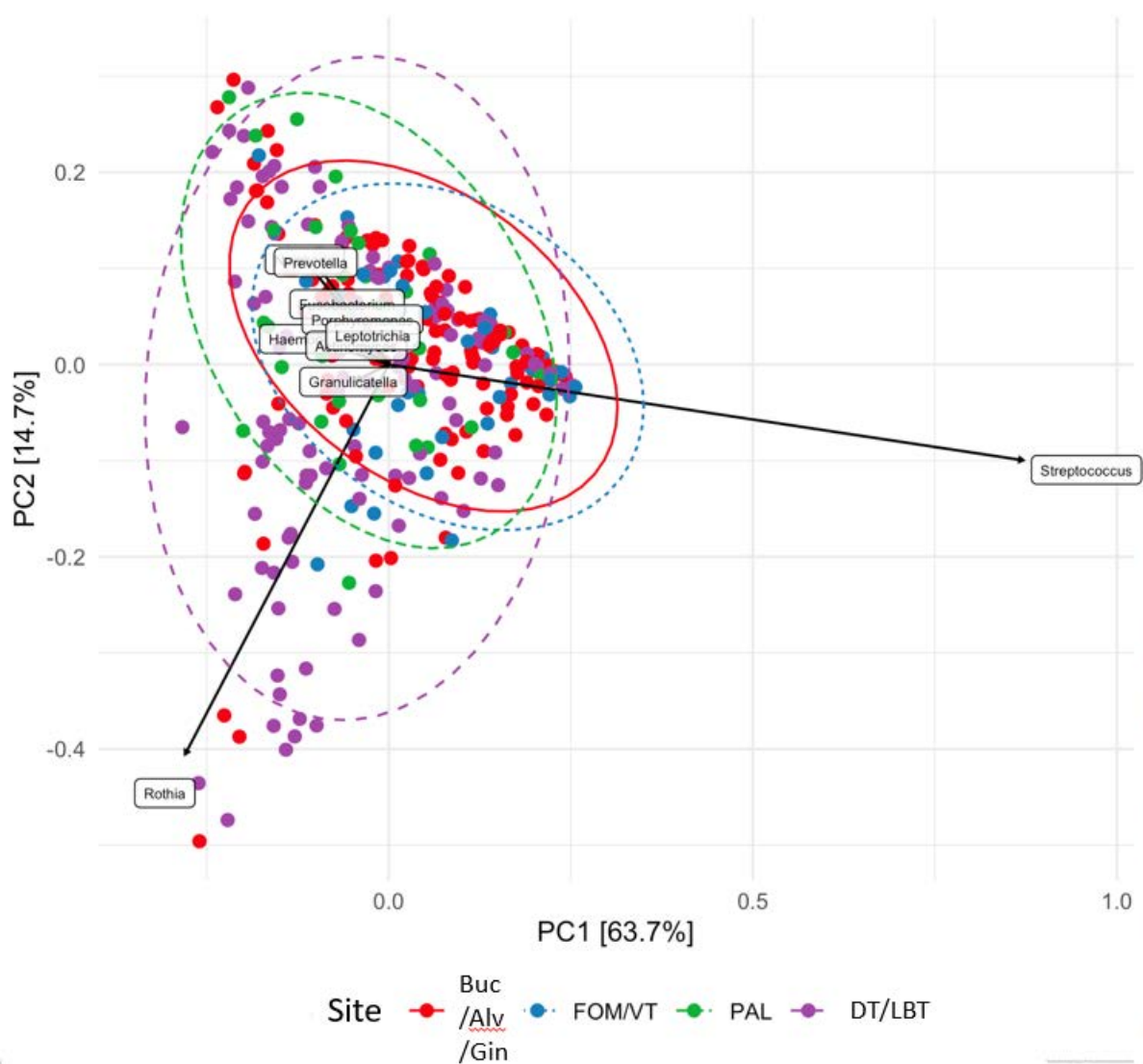


**Figure 28.** Axis 1 and 2 PCA plot by site showing separation of the microbial communities of different mucosal sites. The labelled arrows indicate the influence of specific genera in separating samples along the respective axis. DT/LBT (dorsum tongue/lateral border tongue) sites are clearly dominated by *Rothia* spp., FOM/VT (floor of mouth/ventral tongue) and Buc/Alv/Gin (buccal mucosa/alveolus/gingiva) sites appear to have similar colonisation patterns dominated by streptococci, while the separation of PAL (palatal) sites is influenced by the presence of *Prevotella* spp.

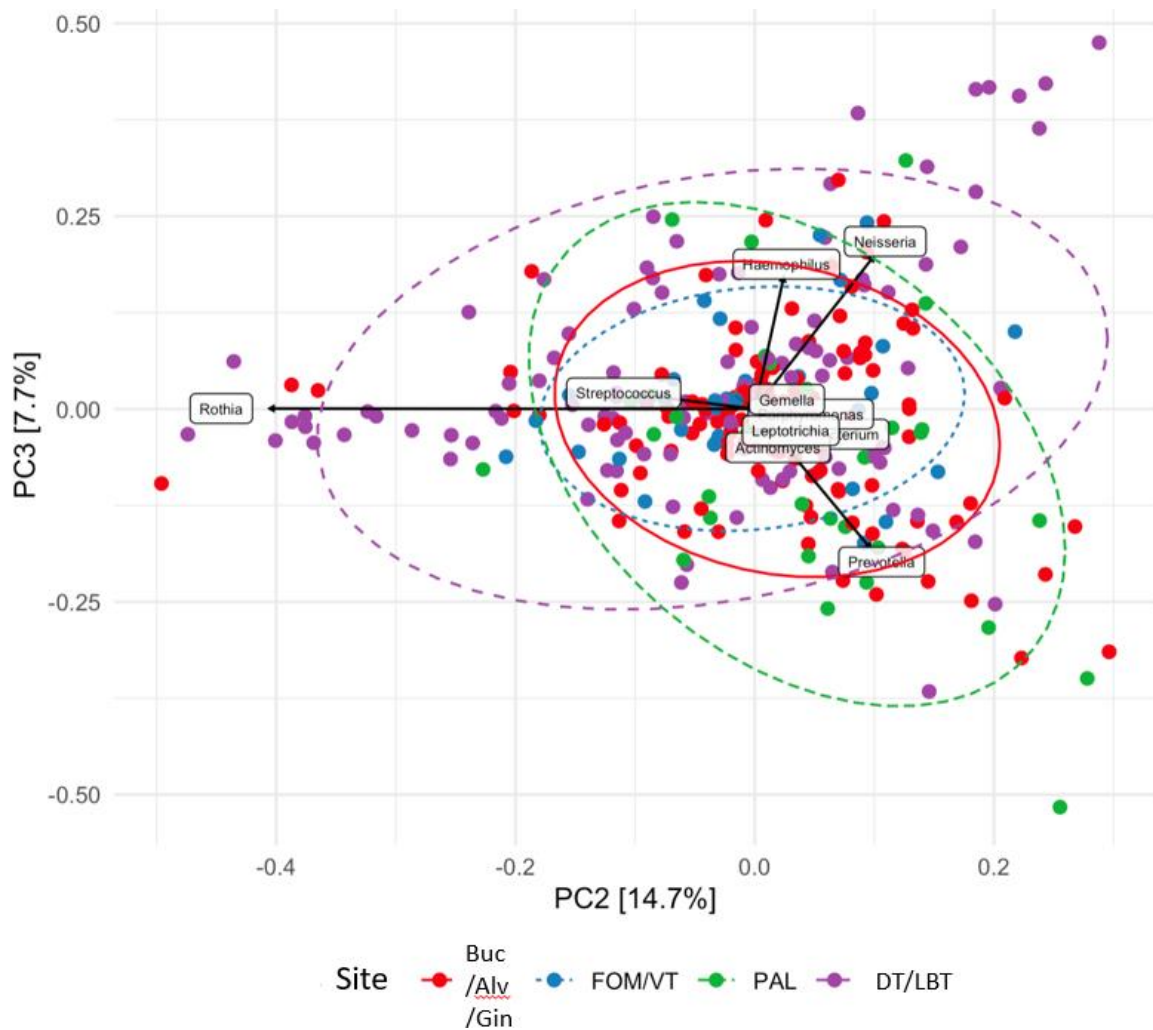


**Figure 29.** Axis 2 and 3 PCA plot by site showing separation of the microbial communities of different mucosal sites. The labelled arrows indicate the influence of specific genera in separating samples along the respective axis. DT/LBT (dorsum tongue/lateral border tongue) sites are clearly dominated by *Rothia* and *Neisseria* spp., FOM/VT (floor of mouth/ventral tongue) and Buc/Alv/Gin (buccal mucosa/alveolus/gingiva) sites have similar colonisation patterns, while the separation of PAL (palatal) sites is due in part to the presence of *Prevotella* spp.

**Table 36.** Pairwise PERMANOVA results for site comparisons (DT/LBT=Dorsum tongue/lateral border tongue, Buc/Alv/Gin=Buccal mucosa/Alveolus/Gingiva, PAL=Palate, FOM/VT=Floor of mouth/ventral tongue)

| Site comparisons      | p value | Adjusted p value |
|-----------------------|---------|------------------|
| DT/LBT vs PAL         | 0.004   | 0.004            |
| DT/LBT vs Buc/Alv/Gin | 0.0001  | 0.0002           |
| DT/LBT vs FOM/VT      | 0.0001  | 0.0002           |
| PAL vs Buc/Alv/Gin    | 0.0001  | 0.0002           |
| PAL vs FOM/VT         | 0.0001  | 0.0002           |
| Buc/Alv/Gin vs FOM/VT | 0.13    | 0.13             |

As before, MaAsLin2 was used to investigate which species were responsible for these changes. Using Buc/Alv/Gin as reference, 96 taxa showed significant changes ( $p_{adj} < 0.1$ ) in abundance: 44 on the DT/LBT, 48 on PAL, but only four on the FOM/VT (**Table 36**). The most significant changes at DT/LBT sites were increased abundances of *Rothia* spp. including *Rothia mucilaginosa* and another taxa designated *Rothia aerea/dentocariosa*, three *Veillonella* spp. and three *Actinomyces* spp. Abundances of eight *Streptococcus* spp. along with *Granulicatella adiacens*, *Gemella sanguinis*, one *Prevotella* sp., and *Haemophilus pittmaniae* were also significantly increased. Twenty-four taxa showed reduced abundance at DT/LBT sites including seven streptococci, *Gemella haemolysans*, *Corynebacterium matruchotii*, *Campylobacter gracilis*, *Rothia dentocariosa*, *Mogibacterium timidum*, *Parvimonas micra*, *Capnocytophaga granulosa*, *Bulleidia extructa*, two *Peptostreptococcus* sp., two *Prevotella* sp., two *Actinomyces* spp., and singles species of *Fusobacterium*, *Freitbacterium* and *Leptotrichia* genera.

Notable changes at PAL sites included increased abundance of six *Prevotella* spp., four *Actinomyces* spp., five *Streptococcus* spp., and two *Leptotrichia* sp. (HMT215 and HMT417). *Rothia mucilaginosa*, an *Alloprevotella* sp. (HMT308), *Oribacterium parvum*, *Mogibacterium diversum*, a *Peptostreptococcaceae* sp., a *Saccharibacteria* sp., a

*Selenomonas* sp., an *Alloprevotella* sp., *Atopobium parvulum*, *Lachnoanaerobaculum orale*, *Campylobacter concisus*, *Solobacterium moorei* and two *Stomatobaculum* spp. also showed increased abundance at PAL sites. Eighteen taxa, including a *Fusobacterium* sp., *Haemophilus haemolyticus*, *Leptotrichia hongkongensis*, *Rothia dentocariosa*, two *Tannerella* spp., two *Campylobacter* spp., a *Veillonella* sp., *Eikenella corrodens*, *Porphyromonas endodontalis*, one *Streptococcus* sp., one *Peptostreptococcus* sp., and *Parvimonas micra* showed reduced abundance at PAL sites. Four taxa showed significantly reduced abundance at FOM/VT sites; a streptococcal taxa designated *Streptococcus constellatus intermedius*, *Parvimonas micra*, *Gemella haemolysans*, and *Mycoplasma faucium*.

The different sites had several species with significant changes in common; two *Actinomyces* species, *Actinomyces lingnae* and *Actinomyces odontolyticus*, *Rothia mucilaginosa* and one streptococcal species showed increased abundance at both DT/LBT and PAL sites relative to Buc/Alv/Gin sites, while *Campylobacter gracilis*, *Mogibacterium timidum*, *Prevotella denticola.*, *Rothia dentocariosa*, one streptococcal and one peptostreptococcal species showed reduced abundance at both DT/LBT and PAL sites. *Gemella haemolysans* was the only species with a change common to DT/LBT and FOM/VT, showing reduced abundance at both sites. Finally, there was only one species with a change common to three sites, *Parvimonas micra* showing reduced abundance at DT/LBT, PAL and FOM/VT sites relative to Buc/Alv/Gin sites. The heatmap generated in MaAsLin2 (**Figure 29**) demonstrates more species with common changes as MaAsLin2 uses a default cut-off value of  $p_{adj}=0.25$ .

**Table 37.** Results of MaAsLin2 analysis showing taxa with significant changes in abundance associated with site. Only results with  $\text{padj} < 0.05$  shown due to volume of results. (Buc/Alv/Gin used as reference) (negative values denote reduced abundance). Full results available in **Appendix A**

| Taxa                                                                        | Site   | Coefficient | p value  | Adjusted p value |
|-----------------------------------------------------------------------------|--------|-------------|----------|------------------|
| <i>Rothia aerea dentocariosa</i>                                            | DT/LBT | 0.84        | 6.52E-12 | 3.66E-09         |
| <i>Veillonella rogosae dispar</i>                                           | DT/LBT | 0.85        | 2.28E-09 | 4.27E-07         |
| <i>Rothia mucilaginosa</i>                                                  | DT/LBT | 0.51        | 4.56E-09 | 6.40E-07         |
| <i>Granulicatella adiacens</i>                                              | DT/LBT | 0.39        | 1.71E-06 | 0.0002           |
| <i>Veillonella rogosae dispar parvula atypica</i>                           | DT/LBT | 0.38        | 3.42E-06 | 0.0003           |
| <i>Veillonella denticariosi</i>                                             | DT/LBT | 0.51        | 1.17E-05 | 0.0006           |
| <i>Streptococcus</i> sp. HMT061                                             | DT/LBT | 0.48        | 1.82E-05 | 0.0008           |
| <i>Actinomyces</i> sp. HMT180 odontolyticus                                 | DT/LBT | 0.51        | 2.89E-05 | 0.001            |
| <i>Streptococcus parasanguinis</i> clade 411 <i>parasanguinis</i> clade 721 | DT/LBT | 0.58        | 7.01E-05 | 0.002            |
| <i>Actinomyces odontolyticus</i>                                            | DT/LBT | 0.38        | 0.002    | 0.02             |
| <i>Streptococcus australis</i>                                              | DT/LBT | 0.38        | 0.002    | 0.02             |
| <i>Haemophilus pittmaniae</i>                                               | DT/LBT | 0.16        | 0.003    | 0.03             |
| <i>Streptococcus parasanguinis</i> clade 411                                | DT/LBT | 0.54        | 0.0002   | 0.004            |
| <i>Streptococcus infantis</i> clade 638                                     | DT/LBT | 0.48        | 0.0003   | 0.006            |
| <i>Actinomyces lingnae</i> Not Validly Published                            | DT/LBT | 0.38        | 0.001    | 0.01             |
| <i>Gemella sanguinis</i>                                                    | DT/LBT | 0.43        | 0.001    | 0.01             |
| <i>Streptococcus thermophilus salivarius</i>                                | DT/LBT | 0.21        | 0.001    | 0.01             |
| <i>Streptococcus</i> sp. HMT423 mitis                                       | DT/LBT | -0.51       | 8.09E-05 | 0.002            |
| <i>Gemella haemolysans</i>                                                  | DT/LBT | -0.63       | 0.0001   | 0.003            |
| <i>Actinomyces oris</i>                                                     | DT/LBT | -0.27       | 0.0001   | 0.003            |
| <i>Streptococcus mitis pneumoniae</i>                                       | DT/LBT | -0.71       | 0.0002   | 0.003            |
| <i>Streptococcus</i> sp. HMT064 mitis                                       | DT/LBT | -0.51       | 0.0002   | 0.003            |
| <i>Streptococcus sanguinis</i>                                              | DT/LBT | -0.37       | 0.001    | 0.02             |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 070               | DT/LBT | -0.34       | 0.003    | 0.04             |
| <i>Prevotella</i> sp. HMT300                                                | DT/LBT | -0.19       | 0.003    | 0.04             |
| <i>Parvimonas micra</i>                                                     | DT/LBT | -0.37       | 0.004    | 0.04             |
| <i>Mogibacterium timidum</i>                                                | DT/LBT | -0.25       | 0.004    | 0.04             |
| <i>Corynebacterium matruchotii</i>                                          | DT/LBT | -0.27       | 0.005    | 0.045            |
| <i>Streptococcus constellatus intermedius</i>                               | FOM/VT | -0.54       | 0.004    | 0.04             |
| <i>Prevotella</i> sp. HMT305                                                | PAL    | 0.56        | 3.74E-10 | 1.05E-07         |
| <i>Peptostreptococcaceae</i> XI G 1 <i>sulci</i>                            | PAL    | 0.41        | 2.08E-07 | 2.34E-05         |
| <i>Mogibacterium diversum</i>                                               | PAL    | 0.65        | 4.43E-06 | 0.0003           |
| <i>Saccharibacteria</i> TM7 G1 <i>bacterium</i> HMT352                      | PAL    | 0.55        | 6.49E-06 | 0.0004           |
| <i>Prevotella pallens</i>                                                   | PAL    | 0.72        | 9.99E-06 | 0.0006           |
| <i>Prevotella salivae</i>                                                   | PAL    | 0.96        | 1.38E-05 | 0.0006           |
| <i>Prevotella histicola</i>                                                 | PAL    | 1.05        | 2.00E-05 | 0.0008           |
| <i>Selenomonas</i> sp. HMT136                                               | PAL    | 0.40        | 2.84E-05 | 0.001            |
| <i>Atopobium parvulum</i>                                                   | PAL    | 0.82        | 5.94E-05 | 0.002            |
| <i>Alloprevotella</i> sp. HMT308                                            | PAL    | 0.66        | 5.78E-05 | 0.002            |
| <i>Solobacterium moorei</i>                                                 | PAL    | 0.62        | 5.48E-05 | 0.002            |
| <i>Streptococcus salivarius</i>                                             | PAL    | 0.42        | 6.38E-05 | 0.002            |

|                                                           |     |       |        |       |
|-----------------------------------------------------------|-----|-------|--------|-------|
| <i>Streptococcus vestibularis thermophilus salivarius</i> | PAL | 0.47  | 0.0001 | 0.003 |
| <i>Lachnoanaerobaculum orale</i>                          | PAL | 0.34  | 0.0002 | 0.005 |
| <i>Stomatobaculum</i> sp. HMT097                          | PAL | 0.42  | 0.001  | 0.009 |
| <i>Streptococcus parasanguinis</i> clade 721              | PAL | 0.68  | 0.001  | 0.01  |
| <i>Actinomyces</i> sp. HMT172                             | PAL | 0.60  | 0.001  | 0.01  |
| <i>Prevotella</i> sp. HMT313                              | PAL | 0.65  | 0.001  | 0.01  |
| <i>Actinomyces odontolyticus</i>                          | PAL | 0.62  | 0.001  | 0.01  |
| <i>Actinomyces graevenitzi</i>                            | PAL | 0.69  | 0.001  | 0.01  |
| <i>Actinomyces lingnae</i> Not Validly Published          | PAL | 0.49  | 0.002  | 0.03  |
| <i>Leptotrichia</i> sp. HMT215                            | PAL | 0.45  | 0.003  | 0.04  |
| <i>Streptococcus</i> sp. HMT057                           | PAL | 0.56  | 0.004  | 0.04  |
| <i>Campylobacter concisus</i>                             | PAL | 0.46  | 0.004  | 0.04  |
| <i>Prevotella melaninogenica</i>                          | PAL | 0.55  | 0.005  | 0.04  |
| <i>Rothia dentocariosa</i>                                | PAL | -0.62 | 0.001  | 0.009 |
| <i>Haemophilus haemolyticus</i>                           | PAL | -0.45 | 0.001  | 0.01  |
| <i>Leptotrichia hongkongensis</i>                         | PAL | -0.37 | 0.003  | 0.04  |
| <i>Parvimonas micra</i>                                   | PAL | -0.62 | 0.001  | 0.02  |
| <i>Fusobacterium nucleatum</i> subsp. <i>vincentii</i>    | PAL | -0.56 | 0.002  | 0.02  |



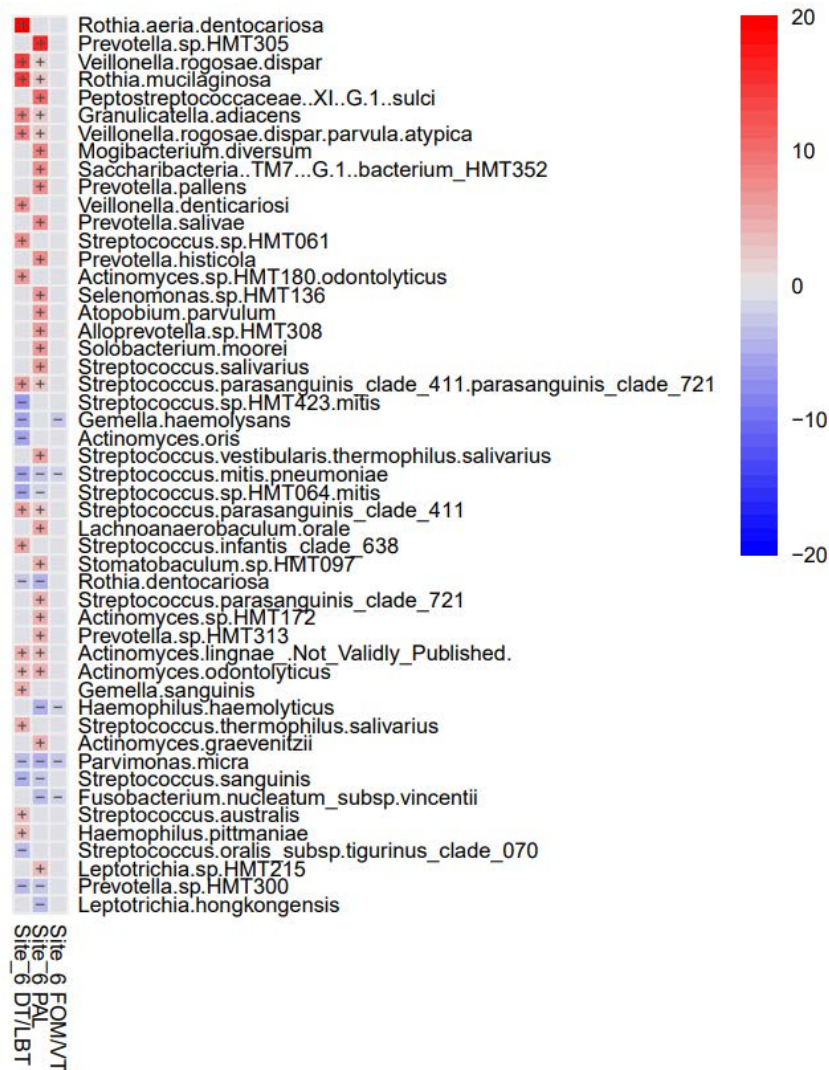


**Figure 30.** Heatmap showing taxa with changes in abundance associated with mucosal site on MaAsLin2 analysis (Effect size calculated as  $-\log(q\text{-value}) \times (\text{coefficient})$ ). Red=increased, blue=reduced relative to Buc/Alv/Gin)

## 6.4 Multivariate model analysis

### 6.4.1 Smoking, alcohol consumption and multiple measures of oral hygiene

A multivariate model including smoking, alcohol consumption, brushing, oral hygiene score, denture wear, mouthwash use, and missing teeth was created and analysed to assess the combined influence of these variables on mucosal microbial colonisation. The most significant changes in abundance in this model were associated with smoking, and agreed with the changes seen on the individual smoking analysis (**Table 38, Figure 31**). Current smoking was associated with significantly reduced abundance of eleven taxa including *Lautropia mirabilis*, a *Neisseria* species designated *Neisseria perflava flavescentis*, *Haemophilus parainfluenzae*, *Kingella oralis*, *Streptococcus intermedius*, a *Stomatobaculum* sp., *Abiotrophia defectiva*, *Cardiobacterium hominis*, *Selenomonas artemidis* and two unspecified streptococcal taxa, and increased abundance of three taxa, *Prevotella pleuritidis*, a *Clostridiales* sp. and *Peptostreptococcaceae XI G5 saphenum*. Former smoking was associated with reduced abundance of four species, *Streptococcus intermedius*, *Selenomonas artemidis*, *Streptococcus gordonii* and *Campylobacter gracilis*. Alcohol consumption, while not associated with any significant changes in abundance when analysed separately, was associated with significantly reduced abundance of *Filifactor alocis* and *Campylobacter showae* at one to 10 units/week in the multivariate model, but not at greater levels of consumption. Brushing once or less times per day was associated with significantly reduced abundance of *Oribacterium sinus* and significantly increased abundance of *Granulicatella elegans*, while fair oral hygiene was associated with significantly increased abundance of *Streptococcus cristatus* clade 578. Finally missing 5 to 15 teeth was associated with reduced abundance of *Haemophilus sputorum* and *Haemophilus haemolyticus* and increased abundance of two streptococcal species.

Mouthwash use and denture wear were not associated with any significant ( $p_{adj} < 0.1$ ) microbiome changes in this multivariate model.

**Table 38.** Significant results from multivariate MaAsLin2 analysis of smoking, alcohol (units/week), brushing (times per day), oral hygiene score, denture wear, mouthwash use and number of missing teeth (negative coefficient values denote reduced abundance)

| Taxa                                          | Variable           | Coefficient | p value  | Adjusted p value |
|-----------------------------------------------|--------------------|-------------|----------|------------------|
| <i>Prevotella pleuritidis</i>                 | Current smoker     | 0.63        | 4.56E-06 | 0.005            |
| <i>Peptostreptococcaceae XI G 5 saphenum</i>  | Current smoker     | 0.54        | 0.0001   | 0.02             |
| <i>Lautropia mirabilis</i>                    | Current smoker     | -0.69       | 6.02E-06 | 0.005            |
| <i>Neisseria perflava flavescens</i>          | Current smoker     | -1.08       | 8.49E-06 | 0.005            |
| <i>Haemophilus parainfluenzae</i>             | Current smoker     | -0.63       | 1.69E-05 | 0.007            |
| <i>Clostridiales F 1 G 1 bacterium HMT093</i> | Current smoker     | 0.31        | 1.62E-05 | 0.007            |
| <i>Kingella oralis</i>                        | Current smoker     | -0.72       | 3.14E-05 | 0.01             |
| <i>Streptococcus intermedius</i>              | Current smoker     | -0.46       | 7.27E-05 | 0.02             |
| <i>Stomatobaculum</i> sp. HMT097              | Current smoker     | -0.43       | 0.0001   | 0.03             |
| <i>Abiotrophia defectiva</i>                  | Current smoker     | -0.33       | 0.0003   | 0.05             |
| <i>Cardiobacterium hominis</i>                | Current smoker     | -0.32       | 0.0004   | 0.07             |
| <i>Streptococcus</i> sp. HMT066 sp. HMT074    | Current smoker     | -0.64       | 0.0006   | 0.07             |
| <i>Selenomonas artemidis</i>                  | Current smoker     | -0.40       | 0.0006   | 0.07             |
| <i>Streptococcus</i> sp. HMT066               | Current smoker     | -0.61       | 0.0007   | 0.08             |
| <i>Streptococcus intermedius</i>              | Former smoker      | -0.58       | 3.76E-06 | 0.005            |
| <i>Selenomonas artemidis</i>                  | Former smoker      | -0.44       | 0.0004   | 0.06             |
| <i>Streptococcus gordonii</i>                 | Former smoker      | -0.76       | 0.0006   | 0.07             |
| <i>Campylobacter gracilis</i>                 | Former smoker      | -0.52       | 0.001    | 0.09             |
| <i>Streptococcus cristatus</i> clade 578      | Fair OH            | 0.49        | 0.0009   | 0.08             |
| <i>Campylobacter showae</i>                   | 1-10 units/week    | -0.33       | 0.0007   | 0.08             |
| <i>Filifactor alocis</i>                      | 1-10 units/week    | -0.56       | 0.0008   | 0.08             |
| <i>Streptococcus constellatus intermedius</i> | 5-15 teeth missing | 0.57        | 0.0002   | 0.03             |
| <i>Haemophilus sputorum</i>                   | 5-15 teeth missing | -0.36       | 0.0003   | 0.05             |
| <i>Streptococcus parasanguinis</i> clade 721  | 5-15 teeth missing | 0.55        | 0.0005   | 0.07             |
| <i>Haemophilus haemolyticus</i>               | 5-15 teeth missing | -0.38       | 0.0006   | 0.07             |
| <i>Granulicatella elegans</i>                 | Brushing once/less | 0.56        | 6.89E-05 | 0.02             |
| <i>Oribacterium sinus</i>                     | Brushing once/less | -0.57       | 0.001    | 0.09             |



**Figure 31.** Heatmap of MaAsLin2 analysis of a multivariate model of smoking, alcohol, oral hygiene, tooth number, mouthwash use, denture wear and brushing showing taxa with significant changes in abundance associated with these variables. (Effect size calculated as  $-\log(q\text{-value}) \times (\text{coefficient})$ ). Red=increased, blue=decreased. Never smoker, no alcohol, OHI-S<0.5, missing < 5 teeth, no denture wear and brushing >1/day used as references) (Brushing\_2=brushing once or less/day)

## 6.4.2 Age and Gender

Differences in mucosal surface colonisation patterns associated with age and gender were investigated in a multivariate model with smoking, tooth number, brushing habit and oral hygiene, as tooth number and oral hygiene were significantly associated with age ( $p=1.11E-09$  and  $p=0.048$  respectively) and current smoking and less frequent brushing were associated with male gender ( $p=0.047$  and  $p=0.01$  respectively) in this cohort (data not shown). Age over 60 years was associated with reduced abundance of *Actinomyces* sp.

HMT175 and *Streptococcus oralis* subsp. *dentisani* clade 398, and increased abundance of *Veillonella rogosae:dispar*, while those in the 40–60-year age group showed increased abundance of *Bulleidia extructa*. Surprisingly male gender was associated with significantly ( $p_{adj} < 0.1$ ) increased abundance of *Selenomonas sputigena*, *Streptococcus parasanguinis* clade 721 and *Actinomyces* sp. HMT180. Increased abundance of the periodontal disease associated pathogens *Treponema socranskii* and *Peptostreptococcaceae* XI G5 *saphenum* were also found in males although the adjusted p values were not significant ( $p_{adj}$  0.11 and 0.12 respectively). Finally, male gender was associated with significantly reduced abundance of *Porphyromonas catoniae* (**Table 39**). Although there were significant associations between age and tooth number and oral hygiene, and between male gender and current smoking and brushing habit, the microbial changes associated with age and gender would appear to be independent of these variables.

**Table 39.** Taxa with significant changes in abundance associated with age and gender after adjusting for smoking habit, number of missing teeth and brushing habit (negative coefficient value denotes reduced abundance) (\*= $p_{adj}<0.1$ )

| Taxa                                                          | Variable    | Coefficient | p value | Adjusted p value |
|---------------------------------------------------------------|-------------|-------------|---------|------------------|
| <i>Bulleidia extructa</i>                                     | 40-60 years | 0.84        | 0.002   | 0.10             |
| <i>Veillonella rogosae dispar</i>                             | >60 years   | 2.92        | 0.0007  | 0.06*            |
| <i>Streptococcus oralis</i> subsp. <i>dentisani</i> clade 398 | >60 years   | -2.75       | 0.0002  | 0.03*            |
| <i>Actinomyces</i> sp. HMT175                                 | >60 years   | -2.32       | 0.0003  | 0.03*            |
| <i>Selenomonas sputigena</i>                                  | Male gender | 1.21        | 0.0002  | 0.03*            |
| <i>Streptococcus parasanguinis</i> clade 721                  | Male gender | 1.72        | 0.0002  | 0.03*            |
| <i>Actinomyces</i> sp. HMT180                                 | Male gender | 0.73        | 0.0004  | 0.04*            |
| <i>Treponema socranskii</i>                                   | Male gender | 0.75        | 0.002   | 0.11             |
| <i>Peptostreptococcaceae</i> XI G 5 <i>saphenum</i>           | Male gender | 1.09        | 0.003   | 0.12             |
| <i>Porphyromonas catoniae</i>                                 | Male gender | -0.51       | 0.001   | 0.09*            |

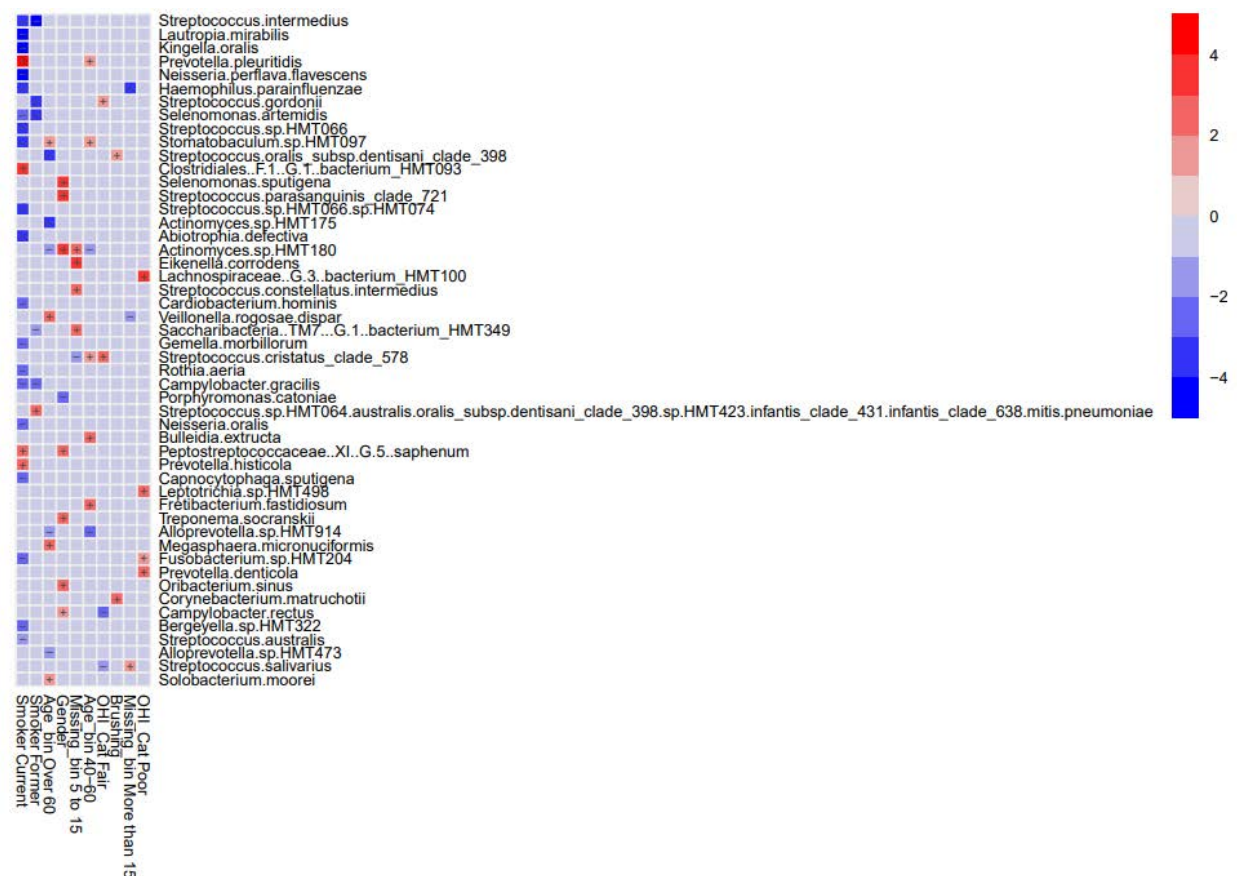


**Figure 32.** Heatmap of MaAsLin2 analysis of a multivariate model of age and gender adjusted for smoking, tooth number and brushing habit showing taxa with significant changes in abundance associated with these variables. (Effect size calculated as  $-\log(q\text{-value}) \times (\text{coefficient})$ ). (Red=increased, blue=decreased. Never smoker, female gender, age<40, brushing >1/day and missing <5 teeth used as references)



### 6.4.3 Smoking and site

As smoking and oral site had such significant influences on the oral microbiome, they were combined into a multivariate model to investigate if these changes were independent of each other. Seventy six of the 77 taxa that demonstrated significant ( $p_{adj}<0.1$ ) changes in abundance associated with current or former smoking (**Table 25**), demonstrated the same changes on the site and smoking model (**Table 40, Figure 33**) confirming that the association with smoking was largely independent of oral site. *Haemophilus sputorum* was associated with current smoking on the individual smoking analysis but the adjusted p value was not significant on the combined smoking and site model ( $p_{adj}=0.1$ ). Two additional species demonstrated changes associated with smoking on the site and smoking model; *Bulleidia extructa* demonstrated increased abundance, and *Neisseria bacilliformis* demonstrated reduced abundance, associated with current smoking on this model.

With regards to site, all 44 taxa showing significant changes associated with DT/LBT sites showed the same changes on the site and smoking model. An additional seven showed changes at DT/LBT sites after adjusting for smoking: two *Streptococcus* spp., a *Neisseria* sp., and *Haemophilus influenzae* were increased in abundance, while *Tannerella forsythia*, *Abitrophia defectiva* and an *Actinomyces* sp. were decreased in abundance. Twelve species showed common changes associated with both current smoking and DT/LBT location on this model: four *Streptococcus* spp., two *Actinomyces* spp. and two *Veillonella* spp. showed increased abundance while *Streptococcus sanguinis*, *Abitrophia defectiva* and two *Actinomyces* spp. showed decreased abundance, suggesting that smoking has an additional influence on DT/LBT colonisation.

PAL sites also showed the same significant changes on the site and smoking model as on the separate site analysis along with a further seven species showing significant changes in abundance independent of smoking: two *Actinomyces* sp., *Streptococcus parasanguinis* clade 411 and *Megasphaera micronuciformis* were increased in abundance while a *Fusobacterium nucleatum* subspecies, *Porphyromonas catoniae* and *Peptostreptococcaceae* XI G4 bacterium HMT369 showed reduced abundance at PAL sites after adjusting for smoking. Twelve species had changes common to both current smoking and PAL location on the site and smoking model suggesting that smoking had an additional influence on palatal colonisation patterns: two *Actinomyces* spp., *Prevotella histicola*, another *Prevotella* sp. (designated HMT313), *Stomatobaculum longum* and four *Streptococcus* spp. showed increased abundance, while *Abiotrophia defectiva*, *Leptotrichia hongkongensis* and *Porphyromonas catoniae* showed reduced abundance.

Finally, FOM/VT sites also showed more changes on the site and smoking model than when analysed alone; 11 species, two *Streptococcus* spp., *Bulleidia extructa*, a *Fusobacterium nucleatum* subsp., *Gemella haemolysans*, *Mycoplasma faucium*, *Parvimonas micra*, *Porphyromonas endodontalis*, *Treponema socranskii*, a *Peptostreptococcus* sp., and *Mogibacterium timidum* showed reduced abundance on the site and smoking model and a single *Actinomyces* sp., showed increased abundance, while only *Gemella haemolysans*, *Mycoplasma faucium*, *Parvimonas micra* and one of the streptococcal species showed reduced abundance on the site only model. Interestingly no species showed changes common to both current smoking and FOM/VT location.



**Table 40.** Results of MaAsLin2 analysis of a site and smoking model showing taxa with significant changes in abundance (only padj<0.05 values shown due to volume of results, full results available in **Appendix B**). (Negative coefficient values denote reduced abundance)

| Taxa                                                         | Variable       | Coefficient | p value  | Adjusted p value |
|--------------------------------------------------------------|----------------|-------------|----------|------------------|
| <i>Neisseria perflava flavescens</i>                         | Current smoker | -1.21       | 9.52E-11 | 2.97E-08         |
| <i>Lautropia mirabilis</i>                                   | Current smoker | -0.75       | 1.49E-11 | 6.95E-09         |
| <i>Haemophilus parainfluenzae</i>                            | Current smoker | -0.71       | 1.12E-08 | 1.31E-06         |
| <i>Streptococcus</i> sp. HMT066 sp HMT074                    | Current smoker | -0.70       | 5.05E-07 | 3.93E-05         |
| <i>Rothia aeria</i>                                          | Current smoker | -0.69       | 3.85E-09 | 6.00E-07         |
| <i>Streptococcus australis</i>                               | Current smoker | -0.66       | 5.64E-08 | 5.86E-06         |
| <i>Kingella oralis</i>                                       | Current smoker | -0.66       | 6.39E-07 | 4.60E-05         |
| <i>Streptococcus sanguinis</i>                               | Current smoker | -0.53       | 3.71E-06 | 0.0002           |
| <i>Neisseria macacae flava mucosa sicca</i>                  | Current smoker | -0.51       | 0.0003   | 0.004            |
| <i>Actinomyces</i> sp. HMT169                                | Current smoker | -0.49       | 9.30E-06 | 0.0003           |
| <i>Streptococcus</i> sp. HMT066                              | Current smoker | -0.49       | 0.0005   | 0.007            |
| <i>Fusobacterium periodonticum hwasookii</i>                 | Current smoker | -0.46       | 0.0002   | 0.003            |
| <i>Stomatobaculum</i> sp. HMT097                             | Current smoker | -0.41       | 7.63E-07 | 5.10E-05         |
| <i>Capnocytophaga sputigena</i>                              | Current smoker | -0.39       | 4.35E-05 | 0.001            |
| <i>Bergeyella</i> sp. HMT322                                 | Current smoker | -0.39       | 3.14E-06 | 0.0002           |
| <i>Abiotrophia defectiva</i>                                 | Current smoker | -0.32       | 3.82E-06 | 0.0002           |
| <i>Neisseria oralis</i>                                      | Current smoker | -0.31       | 2.23E-05 | 0.0006           |
| <i>Gemella morbillorum</i>                                   | Current smoker | -0.31       | 0.004    | 0.04             |
| <i>Streptococcus intermedius</i>                             | Current smoker | -0.31       | 0.0005   | 0.007            |
| <i>Campylobacter concisus</i>                                | Current smoker | -0.30       | 0.007    | 0.046            |
| <i>Cardiobacterium hominis</i>                               | Current smoker | -0.29       | 6.54E-06 | 0.0003           |
| <i>Selenomonas artemidis</i>                                 | Current smoker | -0.25       | 0.003    | 0.03             |
| <i>Neisseria elongata</i>                                    | Current smoker | -0.22       | 0.007    | 0.045            |
| <i>Aggregatibacter</i> sp HMT512                             | Current smoker | -0.21       | 0.002    | 0.02             |
| <i>Capnocytophaga gingivalis</i>                             | Current smoker | -0.21       | 0.005    | 0.04             |
| <i>Aggregatibacter</i> sp. HMT458                            | Current smoker | -0.20       | 0.006    | 0.04             |
| <i>Lachnoanaerobaculum umeaense</i>                          | Current smoker | -0.19       | 0.006    | 0.04             |
| <i>Porphyromonas catoniae</i>                                | Current smoker | -0.12       | 0.004    | 0.04             |
| <i>Treponema</i> sp. HMT237                                  | Current smoker | 0.20        | 0.005    | 0.04             |
| <i>Prevotella</i> sp. HMT526                                 | Current smoker | 0.20        | 0.003    | 0.03             |
| <i>Fusobacterium</i> sp. HMT203                              | Current smoker | 0.21        | 0.004    | 0.03             |
| <i>Veillonella rogosae dispar parvula atypica</i>            | Current smoker | 0.23        | 0.006    | 0.04             |
| <i>Streptococcus</i> sp. HMT064 sp. HMT423 mitis             | Current smoker | 0.23        | 0.006    | 0.04             |
| <i>Tannerella forsythia</i>                                  | Current smoker | 0.23        | 0.007    | 0.04             |
| <i>Peptostreptococcaceae</i> XI G 6 <i>nodatum</i>           | Current smoker | 0.23        | 0.004    | 0.03             |
| <i>Clostridiales</i> F1 G1 <i>bacterium</i> HMT093           | Current smoker | 0.23        | 1.70E-05 | 0.0005           |
| <i>Streptococcus thermophilus salivarius</i>                 | Current smoker | 0.24        | 0.0002   | 0.004            |
| <i>Bacteroidales</i> G2 <i>bacterium</i> HMT274              | Current smoker | 0.26        | 0.004    | 0.04             |
| <i>Peptostreptococcaceae</i> XI G4 <i>bacterium</i> HMT369   | Current smoker | 0.28        | 0.0006   | 0.008            |
| <i>Stomatobaculum longum</i>                                 | Current smoker | 0.32        | 0.0002   | 0.004            |
| <i>Streptococcus oralis</i> subsp <i>tigurinus</i> clade 070 | Current smoker | 0.33        | 0.005    | 0.04             |

|                                                                                                                                                                                  |                |       |          |          |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|-------|----------|----------|
| <i>Mogibacterium timidum</i>                                                                                                                                                     | Current smoker | 0.36  | 5.42E-05 | 0.001    |
| <i>Fusobacterium nucleatum</i> subsp. <i>vincentii</i>                                                                                                                           | Current smoker | 0.36  | 0.004    | 0.03     |
| <i>Veillonella denticariosi</i>                                                                                                                                                  | Current smoker | 0.36  | 0.002    | 0.02     |
| <i>Mycoplasma faucium</i>                                                                                                                                                        | Current smoker | 0.36  | 3.45E-05 | 0.0009   |
| <i>Filifactor alocis</i>                                                                                                                                                         | Current smoker | 0.36  | 0.004    | 0.03     |
| <i>Alloprevotella tannerae</i>                                                                                                                                                   | Current smoker | 0.37  | 0.008    | 0.049    |
| <i>Actinomyces lingnae</i> Not Validly Published                                                                                                                                 | Current smoker | 0.39  | 0.0004   | 0.006    |
| <i>Parvimonas micra</i>                                                                                                                                                          | Current smoker | 0.40  | 0.002    | 0.02     |
| <i>Prevotella</i> sp. HMT313                                                                                                                                                     | Current smoker | 0.42  | 0.001    | 0.01     |
| <i>Streptococcus</i> sp. HMT057                                                                                                                                                  | Current smoker | 0.43  | 0.001    | 0.01     |
| <i>Porphyromonas endodontalis</i>                                                                                                                                                | Current smoker | 0.45  | 0.0003   | 0.005    |
| <i>Actinomyces graevenitzii</i>                                                                                                                                                  | Current smoker | 0.50  | 0.0005   | 0.007    |
| <i>Peptostreptococcaceae</i> XI G5 <i>saphenum</i>                                                                                                                               | Current smoker | 0.50  | 1.99E-06 | 0.0001   |
| <i>Prevotella pleuritidis</i>                                                                                                                                                    | Current smoker | 0.52  | 9.62E-07 | 6.00E-05 |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 071                                                                                                                    | Current smoker | 0.53  | 3.49E-06 | 0.0002   |
| <i>Prevotella histicola</i>                                                                                                                                                      | Current smoker | 0.59  | 0.0004   | 0.007    |
| <i>Atopobium parvulum</i>                                                                                                                                                        | Current smoker | 0.59  | 1.51E-05 | 0.0005   |
| <i>Streptococcus parasanguinis</i> clade 411 <i>parasanguinis</i> clade 721                                                                                                      | Current smoker | 0.76  | 1.24E-07 | 1.16E-05 |
| <i>Streptococcus gordonii</i>                                                                                                                                                    | Former smoker  | -0.62 | 0.001    | 0.008    |
| <i>Actinomyces</i> sp. HMT169                                                                                                                                                    | Former smoker  | -0.50 | 7.46E-05 | 0.001    |
| <i>Streptococcus intermedius</i>                                                                                                                                                 | Former smoker  | -0.41 | 3.76E-05 | 0.001    |
| <i>Streptococcus sanguinis</i>                                                                                                                                                   | Former smoker  | -0.36 | 0.005    | 0.04     |
| <i>Actinomyces</i> sp. HMT175                                                                                                                                                    | Former smoker  | -0.33 | 0.004    | 0.04     |
| <i>Selenomonas artemidis</i>                                                                                                                                                     | Former smoker  | -0.32 | 0.001    | 0.01     |
| <i>Actinomyces</i> sp. HMT448                                                                                                                                                    | Former smoker  | -0.27 | 0.005    | 0.04     |
| <i>Cardiobacterium hominis</i>                                                                                                                                                   | Former smoker  | -0.23 | 0.001    | 0.02     |
| <i>Streptococcus</i> sp. HMT064 <i>australis oralis</i> subsp. <i>dentisani</i> clade 398 sp. HMT423 <i>infantis</i> clade 431 <i>infantis</i> clade 638 <i>mitis pneumoniae</i> | Former smoker  | 0.25  | 0.006    | 0.04     |
| <i>Streptococcus mitis pneumoniae</i>                                                                                                                                            | DT/LBT         | -0.70 | 0.0002   | 0.004    |
| <i>Gemella haemolysans</i>                                                                                                                                                       | DT/LBT         | -0.59 | 0.0002   | 0.004    |
| <i>Streptococcus</i> sp. HMT064 <i>mitis</i>                                                                                                                                     | DT/LBT         | -0.51 | 0.0002   | 0.003    |
| <i>Streptococcus</i> sp. HMT423 <i>mitis</i>                                                                                                                                     | DT/LBT         | -0.50 | 0.00011  | 0.002    |
| <i>Parvimonas micra</i>                                                                                                                                                          | DT/LBT         | -0.39 | 0.002    | 0.02     |
| <i>Streptococcus sanguinis</i>                                                                                                                                                   | DT/LBT         | -0.38 | 0.0007   | 0.009    |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 070                                                                                                                    | DT/LBT         | -0.36 | 0.002    | 0.02     |
| <i>Rothia dentocariosa</i>                                                                                                                                                       | DT/LBT         | -0.34 | 0.005    | 0.04     |
| <i>Actinomyces</i> sp. HMT169                                                                                                                                                    | DT/LBT         | -0.31 | 0.004    | 0.03     |
| <i>Peptostreptococcaceae</i> XI G 9 <i>brachy</i>                                                                                                                                | DT/LBT         | -0.31 | 0.005    | 0.04     |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 071                                                                                                                    | DT/LBT         | -0.30 | 0.007    | 0.04     |
| <i>Campylobacter gracilis</i>                                                                                                                                                    | DT/LBT         | -0.29 | 0.007    | 0.04     |
| <i>Actinomyces oris</i>                                                                                                                                                          | DT/LBT         | -0.28 | 7.44E-05 | 0.001    |
| <i>Corynebacterium matruchotii</i>                                                                                                                                               | DT/LBT         | -0.27 | 0.003    | 0.03     |
| <i>Mogibacterium timidum</i>                                                                                                                                                     | DT/LBT         | -0.26 | 0.003    | 0.03     |
| <i>Peptostreptococcaceae</i> XI G 4 <i>bacterium</i> HMT369                                                                                                                      | DT/LBT         | -0.22 | 0.005    | 0.04     |
| <i>Prevotella</i> sp. HMT300                                                                                                                                                     | DT/LBT         | -0.19 | 0.003    | 0.03     |
| <i>Haemophilus pittmaniae</i>                                                                                                                                                    | DT/LBT         | 0.17  | 0.002    | 0.02     |

|                                                                      |        |       |          |          |
|----------------------------------------------------------------------|--------|-------|----------|----------|
| <i>Prevotella</i> sp. HMT396                                         | DT/LBT | 0.19  | 0.006    | 0.04     |
| <i>Streptococcus thermophilus salivarius</i>                         | DT/LBT | 0.21  | 0.0008   | 0.01     |
| <i>Actinomyces odontolyticus</i>                                     | DT/LBT | 0.36  | 0.003    | 0.03     |
| <i>Actinomyces lingnae</i> Not Validly Published                     | DT/LBT | 0.37  | 0.0007   | 0.009    |
| <i>Veillonella rogosae dispar parvula atypica</i>                    | DT/LBT | 0.38  | 2.94E-06 | 0.0002   |
| <i>Granulicatella adiacens</i>                                       | DT/LBT | 0.39  | 2.13E-06 | 0.0001   |
| <i>Streptococcus australis</i>                                       | DT/LBT | 0.39  | 0.001    | 0.01     |
| <i>Streptococcus</i> sp. HMT066 sp HMT074                            | DT/LBT | 0.40  | 0.003    | 0.03     |
| <i>Gemella sanguinis</i>                                             | DT/LBT | 0.43  | 0.0009   | 0.01     |
| <i>Streptococcus</i> sp. HMT061                                      | DT/LBT | 0.48  | 1.60E-05 | 0.0005   |
| <i>Streptococcus infantis clade 638</i>                              | DT/LBT | 0.50  | 0.0002   | 0.003    |
| <i>Veillonella denticariosi</i>                                      | DT/LBT | 0.50  | 1.30E-05 | 0.0004   |
| <i>Rothia mucilaginoso</i>                                           | DT/LBT | 0.51  | 5.85E-09 | 7.81E-07 |
| <i>Actinomyces</i> sp. HMT180 odontolyticus                          | DT/LBT | 0.52  | 2.91E-05 | 0.0008   |
| <i>Streptococcus parasanguinis clade 411</i>                         | DT/LBT | 0.53  | 0.0003   | 0.004    |
| <i>Streptococcus parasanguinis clade 411 parasanguinis clade 721</i> | DT/LBT | 0.56  | 5.99E-05 | 0.001    |
| <i>Veillonella rogosae dispar</i>                                    | DT/LBT | 0.84  | 3.51E-09 | 6.00E-07 |
| <i>Rothia aerea dentocariosa</i>                                     | DT/LBT | 0.85  | 5.41E-12 | 5.06E-09 |
| <i>Streptococcus constellatus intermedius</i>                        | FOM/VT | -0.56 | 0.003    | 0.03     |
| <i>Parvimonas micra</i>                                              | FOM/VT | -0.49 | 0.005    | 0.04     |
| <i>Mycoplasma faucium</i>                                            | FOM/VT | -0.34 | 0.004    | 0.03     |
| <i>Mogibacterium timidum</i>                                         | FOM/VT | -0.32 | 0.007    | 0.04     |
| <i>Parvimonas micra</i>                                              | PAL    | -0.62 | 0.0010   | 0.01     |
| <i>Rothia dentocariosa</i>                                           | PAL    | -0.61 | 0.0006   | 0.008    |
| <i>Fusobacterium nucleatum subsp. vincentii</i>                      | PAL    | -0.57 | 0.001    | 0.02     |
| <i>Porphyromonas endodontalis</i>                                    | PAL    | -0.50 | 0.006    | 0.04     |
| <i>Prevotella denticola</i>                                          | PAL    | -0.48 | 0.006    | 0.04     |
| <i>Haemophilus haemolyticus</i>                                      | PAL    | -0.46 | 0.0008   | 0.009    |
| <i>Leptotrichia hongkongensis</i>                                    | PAL    | -0.37 | 0.004    | 0.03     |
| <i>Abiotrophia defectiva</i>                                         | PAL    | -0.27 | 0.006    | 0.04     |
| <i>Prevotella</i> sp. HMT300                                         | PAL    | -0.26 | 0.007    | 0.04     |
| <i>Lachnoanaerobaculum orale</i>                                     | PAL    | 0.34  | 0.0003   | 0.004    |
| <i>Selenomonas</i> sp. HMT136                                        | PAL    | 0.40  | 2.76E-05 | 0.0007   |
| <i>Peptostreptococcaceae XI G1 sulci</i>                             | PAL    | 0.41  | 2.09E-07 | 1.78E-05 |
| <i>Streptococcus salivarius</i>                                      | PAL    | 0.42  | 7.02E-05 | 0.001    |
| <i>Stomatobaculum</i> sp. HMT097                                     | PAL    | 0.43  | 0.0003   | 0.005    |
| <i>Leptotrichia</i> sp. HMT215                                       | PAL    | 0.45  | 0.003    | 0.03     |
| <i>Oribacterium sinus</i>                                            | PAL    | 0.46  | 0.007    | 0.04     |
| <i>Campylobacter concisus</i>                                        | PAL    | 0.46  | 0.004    | 0.03     |
| <i>Streptococcus vestibularis thermophilus salivarius</i>            | PAL    | 0.47  | 0.0001   | 0.003    |
| <i>Actinomyces lingnae</i> Not Validly Published                     | PAL    | 0.49  | 0.002    | 0.02     |
| <i>Saccharibacteria TM7 G1 bacterium HMT352</i>                      | PAL    | 0.55  | 6.95E-06 | 0.0003   |
| <i>Prevotella melaninogenica</i>                                     | PAL    | 0.55  | 0.004    | 0.04     |
| <i>Streptococcus parasanguinis clade 411 parasanguinis clade 721</i> | PAL    | 0.55  | 0.007    | 0.04     |
| <i>Streptococcus</i> sp. HMT057                                      | PAL    | 0.56  | 0.004    | 0.03     |

|                                              |     |      |          |          |
|----------------------------------------------|-----|------|----------|----------|
| <i>Prevotella</i> sp. HMT305                 | PAL | 0.56 | 4.27E-10 | 9.97E-08 |
| <i>Actinomyces</i> sp. HMT172                | PAL | 0.60 | 0.0006   | 0.008    |
| <i>Leptotrichia</i> sp. HMT417               | PAL | 0.61 | 0.007    | 0.04     |
| <i>Solobacterium moorei</i>                  | PAL | 0.62 | 5.78E-05 | 0.001    |
| <i>Actinomyces odontolyticus</i>             | PAL | 0.63 | 0.0005   | 0.007    |
| <i>Prevotella</i> sp. HMT313                 | PAL | 0.65 | 0.0006   | 0.008    |
| <i>Mogibacterium diversum</i>                | PAL | 0.65 | 4.69E-06 | 0.0002   |
| <i>Alloprevotella</i> sp. HMT308             | PAL | 0.66 | 6.25E-05 | 0.001    |
| <i>Streptococcus parasanguinis</i> clade 721 | PAL | 0.67 | 0.0006   | 0.008    |
| <i>Actinomyces graevenitzi</i>               | PAL | 0.68 | 0.0010   | 0.01     |
| <i>Prevotella pallens</i>                    | PAL | 0.72 | 1.08E-05 | 0.0004   |
| <i>Atopobium parvulum</i>                    | PAL | 0.81 | 3.93E-05 | 0.0009   |
| <i>Prevotella salivae</i>                    | PAL | 0.96 | 1.27E-05 | 0.0004   |
| <i>Prevotella histicola</i>                  | PAL | 1.05 | 1.66E-05 | 0.0005   |

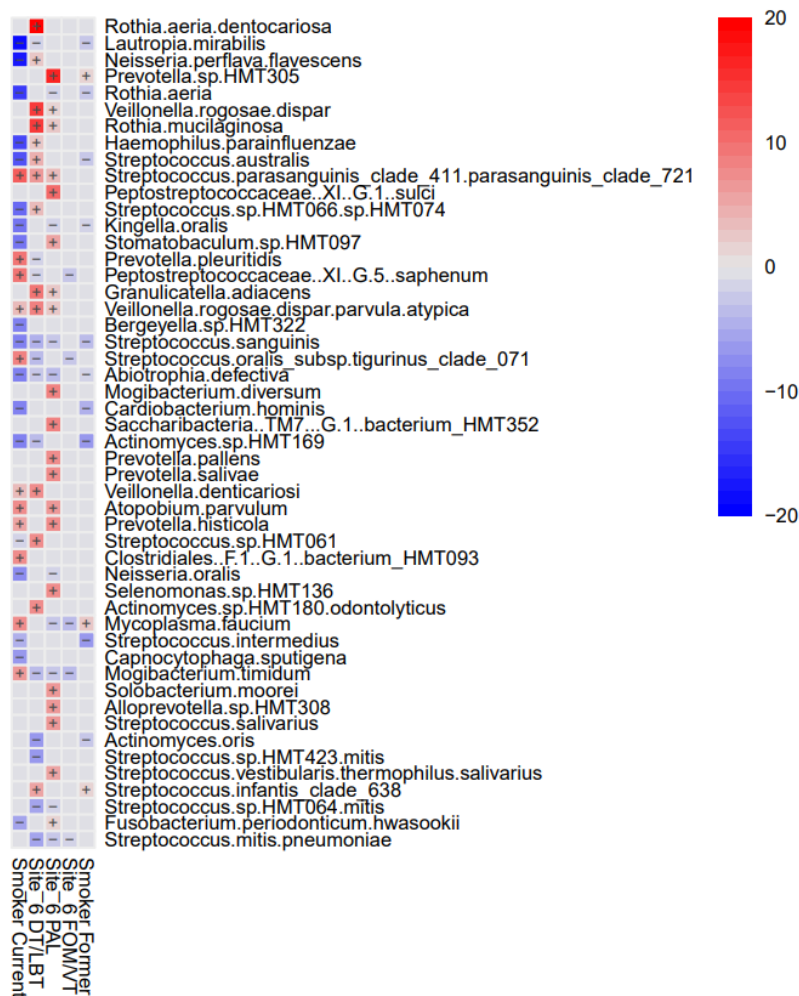


**Figure 33.** MaAsLin2 generated heatmap showing taxa with changes in abundance associated with smoking and oral site. (Effect size calculated as  $-\log(q\text{-value}) \times (\text{coefficient})$ ). Buc/Alv/Gin and never smokers used as references. Red=increased, blue=reduced)

## 6.5 Discussion:

### 6.5.1 Influence of plaque, oral hygiene measures and tooth number on the mucosal microbiome

Periodontal disease is an inflammatory disease of the supporting tissues of the teeth caused by specific, or groups of specific, micro-organisms<sup>234</sup> and preceded by gingival inflammation due to plaque accumulation from poor oral hygiene. Early plaque colonisers are predominantly Gram-positive, facultatively aerobic, non-motile cocci and rods of

*Streptococcus* and *Actinomyces* genera respectively. As plaque develops, and the bacterial population increases, there is a shift to a more anaerobic population dominated by Gram-negative rods and filamentous organisms, with the final phase of plaque development involving proliferation of motile, tissue-invasive anaerobic spirochetes<sup>235</sup>. A healthy periodontium is therefore associated with predominance of *Streptococcus* and *Actinomyces* spp., with only small proportions of Gram-negative periodontal disease-associated species. Chronic periodontal disease, on the other hand, is associated with increased abundances of *Campylobacter* spp., *Eikenella corrodens*, *Fusobacterium* spp., *Parvimonas micra*, *Prevotella* spp., *Leptotrichia* spp., *Selenomonas* spp., and *Treponema* spp., while the three red-complex species, *Porphyromonas gingivalis*, *Tannerella forsythia* and *Treponema denticola*, are associated with severe disease<sup>235</sup>.

Although I did not measure periodontal disease *per se*, I measured plaque levels using the Simplified Greene and Vermillion Index (OHI-S)<sup>205</sup>, along with brushing habits, mouthwash use, tooth number and denture wear, and found that these factors have an influence on the microbial colonisation of mucosal surfaces. Poor oral hygiene was associated with increased abundance ( $p < 0.05$ , but  $\text{padj} > 0.1$ ) of several periodontal pathogens on clinically normal oral mucosa, including several *Prevotella* spp. and a *Tannerella* sp., along with plaque-associated species of genera *Leptotrichia*, *Saccharibacteria* and *Lachnospiraceae*, and reduced abundance of the early plaque-coloniser *Streptococcus sanguinis*. Brushing habits also had some impact on the oral microbiome, with brushing once or less per day associated with reduced abundance of the commensal *Rothia dentocariosa*, along with *Kingella oralis*, *Oribacterium sinus*, and an unspecified *Streptococcus* sp., and increased abundance of *Granulicatella elegans*. As it can take up to four days without oral hygiene measures for dental plaque to begin to significantly change its structure<sup>235</sup>, and only 5.1% (13 of 256) of

the cohort brushed less than once per day, it is not unexpected that brushing habits did not have a profound influence on the microbiome. Mouthwash use was associated with increased abundance of the commensal mucosal colonisers *Granulicatella adiacens* and *Rothia mucilaginosa* (although  $\text{padj} > 0.1$ ), and significantly reduced abundance of *Streptococcus mitis pneumoniae*.

I also found that missing more than 15 teeth was significantly associated with reduced abundance of several plaque-associated species including two *Haemophilus* spp., a *Fusobacterium nucleatum* subsp., *Streptococcus sanguinis*, two other unspecified *Streptococcus* spp., *Rothia aeria*, *Rothia dentocariosa*, two *Actinomyces* spp., an *Alloprevotella* sp. and *Lautropia mirabilis*. Missing five to 15 teeth was associated with increased abundance of three streptococcal species, a *Saccharibacteria* sp., *Eikenella corrodens*, a *Fretibacterium* sp., *Parvimonas micra* and a *Treponema* species, and reduced abundance of *Streptococcus australis*, *Rothia aeria*, two *Haemophilus* spp. and an *Alloprevotella* sp. The detection of more plaque-associated bacteria in the latter group is likely related to the availability of more tooth surfaces for early plaque colonisers. Finally, due to the absence of teeth, complete denture wear was associated with significantly reduced abundance of *Rothia dentocariosa*, two *Streptococcus* spp., *Parvimonas micra*, a subspecies of *Fusobacterium nucleatum* and an *Alloprevotella* species, and increased abundance of another streptococcal species. Wearing a partial denture was associated with reduced abundance of *Streptococcus sanguinis*, *Lautropia mirabilis*, *Haemophilus parainfluenzae*, *Corynebacterium durum*, *Rothia aeria*, *Filifactor alocis* and an *Actinomyces* species, and increased abundance of *Streptococcus infantis clade 431 mitis*, although none of the adjusted p values were significant.

Socranksy and Huffajee<sup>236</sup> were one of the few to compare the mucosal microbial colonisation patterns of dentate and edentulous (complete denture-wearing) subjects, sampling supragingival plaque, saliva and various mucosal sites (dorsal tongue, lateral tongue, ventral tongue, floor of mouth, buccal mucosa, hard palate, vestibule, attached gingiva and lip vestibule) in 28 periodontally healthy, 8 periodontitis affected, and 18 fully edentulous denture-wearing subjects. Checkerboard DNA–DNA hybridization was used to assess the presence of 41 bacteria in each group and comparisons then made of the relative abundances between the groups. The microbial profiles of the dentate healthy and periodontitis subjects were found to be more similar to each other than to the edentulous subjects. In agreement with my findings, edentulous subjects were found to have significantly reduced abundance of a *Fusobacterium nucleatum* subspecies (all sites except hard palate, lip vestibule and ventral tongue) compared with dentate subjects. The edentulous group were also found to have significantly reduced abundances of *Eikenella corrodens* (all sites except dorsum tongue) and *Veillonella parvula* (all sites except ventral tongue and floor of mouth), and increased abundances of *Fusobacterium periodonticum* (all sites), *Streptococcus mitis* (all sites except lateral tongue and dorsum tongue, where it was significantly reduced) and *Streptococcus gordonii* (all sites except dorsum and lateral tongue), findings I did not replicate. I also found complete denture-wearers to have significantly reduced abundances of *Streptococcus sanguinis*, a finding not seen in Socranksy and Huffajee's study, and of *Streptococcus cristatus* clade 578, *Parvimonas micra*, *Alloprevotella tanneriae* and *Rothia dentocariosa*, species not tested for by Socranksy and Huffajee's study. Despite the differences between my findings and Socranksy and Huffajee's, it is clear that the soft tissue microbiota differs between dentate and edentulous subjects, likely due to the different hard tissue surfaces present, although the different oral hygiene measures of dentate and edentulous patients may also be a factor.



### 6.5.2 Influence of age and gender on the mucosal microbiome

I also identified age and gender associated differences in a multivariate analysis with smoking, tooth-loss and oral hygiene measures. Loss of *Actinomyces* sp. HMT175 and *Streptococcus oralis* subsp. *dentisani* clade 398, along with an increase in *Veillonella rogosae:dispar* was deemed highly significant in those over the age of 60. As these changes were independent of tooth loss, other age-related factors such immunosenescence could be responsible. Interestingly, men from this cohort (as detailed in **section 6.2**) tended to harbour increased levels of the periodontal pathogens *Treponema socranskii* and *Peptostreptococcaceae XI G5 saphenum* with multivariate analysis indicating that this was independent of smoking and tooth loss. The incidence of periodontal disease is known to be higher in males than females which could be related to different abundances of periodontal pathogens. Liu *et al.*<sup>237</sup> also recently identified differences in the oral microbiome between females and males based on 3504 saliva and 3694 dorsum tongue samples in a Chinese cohort. After adjusting for age, body mass index, mouthwash use and respiratory disease, women were found to harbour more species of genera *Streptococcus*, *Prevotella* and *Granulicatella*, while men harboured more potential periodontal pathogens including those of genera *Campylobacter*, *Porphyromonas*, *Fusobacterium* and *Capnocytophaga*. In contrast to my results, species of genera *Treponema* and *Peptostreptococcus* were not identified. They linked some of these differences to hormonal and metabolic differences between females and males. Female-specific factors including age of menarche, menstrual cycle, pregnancy, lactation and menopause were found to be responsible for 0.1-0.7% of the variance in salivary microbial composition, and 0.1-1.2% of the dorsum tongue microbial differences, between females and males. Further, serum steroid metabolites, aldosterone, progesterone, cortisol and testosterone, known to differ between females and males, were responsible for 0.1-0.4% of the salivary variance and 0.1-

0.4% of the tongue variance. However, smoking status, also recognised to differ between males and females, and tooth number were not clearly recorded or adjusted for.

### 6.5.3 Influence of mucosal site on the mucosal microbiome

There is general agreement in the literature that the oral microbiome displays site-specificity, and that certain sites can be clustered according to their colonisation patterns, with most agreeing that the dorsum and lateral tongue have similar microbiota while buccal mucosa, palate, floor of mouth and gingival sites also tend to cluster.<sup>236,238–242</sup>

Prior to the advent of NGS, in a study of 225 healthy individuals, Mager *et al.*<sup>238</sup> used checkerboard DNA-DNA hybridisation to investigate the presence of 40 common bacteria in saliva and supra- and sub-gingival plaque and on eight mucosal sites: dorsum tongue, ventral tongue, lateral tongue, maxillary anterior gingiva, floor of mouth, buccal mucosa and anterior vestibule/upper and lower labial mucosa. However, the palate was not sampled, smoking habits were not recorded and both periodontally healthy and disease subjects were included but differences not measured. The profiles of the 40 bacteria measured were found to form three distinct clusters; plaque samples forming the first cluster, saliva, lateral and dorsal tongue samples the second, while the remaining soft tissue sites (ventral tongue, floor of mouth, gingiva, buccal mucosa and labial mucosa) constituted the third cluster. Mager's saliva/tongue cluster was marked by significantly higher proportions of *Veillonella parvula*, *Prevotella melaninogenica*, *Eikenella corrodens*, *Neisseria mucosa*, *Actinomyces odontolyticus*, *Fusobacterium periodonticum*, *Fusobacterium nucleatum* subsp. *vincentii* and *Porphyromonas gingivalis*, while *Streptococcus mitis*, *Streptococcus oralis* and *Selenomonas noxia* were significantly more abundant in the third cluster (ventral tongue,

floor of mouth, gingiva, buccal mucosa and labial mucosa). The dominant colonisers of the dorsum and lateral tongue were found to be anaerobic or facultatively anaerobic Gram-negative species such as *Prevotella melaninogenica*, *Veillonella parvula* and *Capnocytophaga gingivalis*, rather than streptococci<sup>238</sup>.

The study of Aas *et al.*<sup>240</sup> was also pre-NGS, using 16S rRNA cloning, and was much smaller, sampling only five individuals. Samples were taken of supragingival and subgingival plaque and from seven intra-oral sites: dorsum tongue, lateral borders of tongue, buccal mucosa, hard palate, soft palate, labial gingiva, and tonsils. Subjects were periodontally and dentally healthy but again smoking was not recorded. The findings were broadly similar to those of Mager *et al.*, although they did not attempt to cluster samples together and the dominant colonisers of the tongue differed, likely in part due to the different methodology used for bacterial identification. 141 taxa from six different phyla were found, although approximately 60% of the sample was represented by as yet uncultivated phylotypes. The microbiome of the buccal mucosa was dominated by *Streptococcus mitis* and *Gemella haemolysans* and that of the labial gingiva by *Streptococcus mitis*, *Granulicatella* spp. and *Gemella* spp. The dominant species of the lateral tongue were *Streptococcus mitis*, *Streptococcus mitis* biovar 2, *Streptococcus australis*, two unnamed *Streptococcus* clones, the facultative anaerobes *Granulicatella adiacens* and *Gemella haemolysans* and anaerobic *Veillonella* spp., while the dorsum of the tongue was populated predominantly by several streptococcal species including *Streptococcus mitis*, *Streptococcus australis*, *Streptococcus parasanguinis* and *Streptococcus salivarius*, along with *Granulicatella adiacens* and *Veillonella* spp. *Streptococcus mitis* again was the predominant species on the hard palate, along with *Streptococcus* biovar 2, *Streptococcus infantis*, another *Streptococcus* clone, *Granulicatella elegans*, *Gemella haemolysans* and

*Neisseria subflava*. The colonisation pattern of the soft palate was similar with *Streptococcus mitis*, another not-yet-cultivable *Streptococcus* species, *Granulicatella adiacens* and *Gemella haemolysans* predominant.

More recently, Segata *et al.*<sup>241</sup> sampled 200 subjects at seven oral sites: buccal mucosa, keratinised attached gingiva, hard palate, saliva, tongue and two tooth surfaces. The specific tongue surface sampled was not stated. Tonsils, throat and stool were also sampled but smoking habits and periodontal health were not recorded. Four distinct community types were found within the ten sites sampled. Buccal mucosa, gingiva and hard palate samples clustered together (Group 1), while saliva and tongue samples grouped with throat and tonsils (Group 2). The buccal mucosa, gingiva and hard palate cluster consisted mostly of Firmicutes, followed by Proteobacteria, Bacteroides, Actinobacteria and Fusobacteria. The second group (saliva, tongue, tonsils and throat) however had reduced relative abundance of Firmicutes and increased abundance of Bacteroides, Fusobacteria, Actinobacteria and TM7 (now known as Saccharibacteria). At the genera level, Group 1 was dominated by *Streptococcus*, while Group 2 had a more even distribution of *Streptococcus*, *Veillonella*, *Prevotella*, *Neisseria*, *Fusobacterium*, *Actinomyces* and *Leptotrichia*.

Finally, McLean *et al.*<sup>243</sup> described site-tropism for streptococci, showing that *Streptococcus mitis* was most abundant on buccal mucosa and keratinised gingiva, *Streptococcus infantis* most abundant on the tongue and *Streptococcus oralis* most abundant in dental plaque.

My investigation showed dorsum and lateral tongue sites were dominated by *Rothia* spp., *Veillonella* spp., *Actinomyces* spp., several streptococci, and *Granulicatella adiacens*, thus showing similar colonisation patterns to previous studies. Although the PCA plot

demonstrated *Neisseria* at tongue sites, this change was most likely related to tongue sites in non-smokers as MaAsLin2 analysis did not identify any significant changes in *Neisseria* abundance at DT/LBT sites. However, interestingly, the increased abundance of *Rothia* spp. that I noted on the tongue was not seen in other studies<sup>238,240,241</sup>. Furthermore, I did not find an increased abundance of *Prevotella* or *Fusobacterium* spp. on the tongue, as noted by Segata *et al.*<sup>241</sup>. The other notable difference between my findings and those of others, was in abundances of *Gemella haemolysans*; I found its abundance decreased on the tongue while other studies reported it to be increased<sup>238,240,241</sup>.

My buccal mucosa, alveolus/gingiva (Buc/Alv/Gin) and floor of mouth/ventral tongue findings also agreed with previous studies<sup>236,238,240–242</sup>, having very similar microbiomes dominated by streptococcal species, of which only one, a taxa designated *Streptococcus constellatus intermedius*, differed between the sites.

However, my findings and those of previous studies differed significantly with regards to palatal colonisation. Some authors<sup>240,241</sup> have reported that buccal mucosa and palatal sites have similar microbial patterns and can be clustered together but I found the colonisation patterns of these sites to be significantly different. Palatal sites were dominated by *Prevotella* spp., *Actinomyces* spp. and *Streptococcus* spp., contrasting with other studies which demonstrated streptococcal dominance<sup>238,240,241</sup> without any significant increase in *Prevotella* or *Actinomyces* spp. Significantly however, as mentioned, none of the previously mentioned site studies recorded subjects' smoking habits which may explain some of the differences noted. My analysis demonstrated that smoking appeared to have more of an effect on palatal colonisation than on other sites, with ten species showing changes common to both current smoking and palatal site, compared with six on the tongue and none on the

floor of mouth/ventral tongue. Some differences are also likely due to the different sample processing techniques used and may in some way explain the absence of *Rothia* spp. from previous studies.

#### 6.5.4 Influence of smoking on the mucosal microbiome

Smoking is arguably the most significant influencer of the oral microbiome. Numerous studies have documented the changes smoking imparts on oral microbial colonisation. There is a lack of consistency as to the exact changes amongst studies, likely in part related to differing sampling methods (saliva collection, mucosal swab, supra- and sub-gingival plaque sampling), and inclusion criteria, particularly with regards to periodontal disease, employed. However, all agree that smoking has a profound effect on the oral microbiome (**Table 41**). It should also be noted that most smoking-related microbiome studies have significantly more non-smokers than smokers in their sample populations, which potentially influences reported results. Smoking is known to have a profound inhibitory effect on *Neisseria* and is also associated with reduced abundances of *Staphylococcus aureus*, *Capnocytophaga* spp., *Gemella* spp. and certain streptococci such as *Streptococcus pneumoniae*<sup>57–59,61,63,71</sup>. Most studies also agree that smokers have increased abundances of *Veillonella*,<sup>59,65,70</sup> while several have demonstrated increased abundance of periodontal pathogens such as *Tannerella*, *Parvimonas* and *Prevotella*<sup>60,69,70</sup> in smokers. However, reported abundances of *Fusobacteria* in smokers vary, with some showing increased abundance associated with smoking<sup>68–71</sup> and others showing reduced abundance<sup>59,61</sup>. Despite the disagreements in the literature, it is clear that smoking potentiates a switch to a facultatively anaerobic or frankly anaerobic Gram-negative heavy microbiome.

**Table 41.** Significant differences between the microbiomes of smokers and non-smokers from previous studies

| Author                                    | Sampling method     | Smokers/non-smokers                                                                                          | Periodontal health                                                                       | Increased in smokers v non-smokers                                                                                                                                                                                                                                                                                                                                                            | Decreased in smokers v non-smokers                                                                                                                                                                                           |
|-------------------------------------------|---------------------|--------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Charlson <i>et al.</i> 2010 <sup>59</sup> | Oropharyngeal swabs | 29 smokers<br>33 non-smokers                                                                                 | Not documented<br>No antibiotics for 3 months prior to sampling                          | <i>Megasphaera</i><br><i>Veillonella</i>                                                                                                                                                                                                                                                                                                                                                      | <i>Neisseria</i><br><i>Fusobacterium</i><br><i>Capnocytophaga</i>                                                                                                                                                            |
| Morris <i>et al.</i> 2013 <sup>63</sup>   | Saliva              | 19 smokers<br>45 non-smokers                                                                                 | Not measured                                                                             | -                                                                                                                                                                                                                                                                                                                                                                                             | <i>Neisseria</i><br><i>Gemella</i><br><i>Porphyromonas</i>                                                                                                                                                                   |
| Thomas <i>et al.</i> 2014 <sup>60</sup>   | Mucosal swabs       | Mucosal swabs<br>13 smokers<br>9 never smokers                                                               | Not measured                                                                             | <i>Fusobacteria</i><br><i>Actinobacteria</i><br><i>Capnocytophaga</i><br><i>Prevotella</i>                                                                                                                                                                                                                                                                                                    | Firmicutes<br>Cyanobacteria<br>TM7<br>BD1-5<br><i>Neisseria</i><br><i>Peptostreptococcus</i><br><i>Gemella</i><br><i>Granulicatella</i><br><i>Staphylococcus</i>                                                             |
| Belstrom <i>et al.</i> 2014 <sup>64</sup> | Saliva              | 21 smokers<br>123 former smokers<br>148 never smokers<br>Note no difference between former and never smokers | Low levels of caries and periodontal disease – significant disease an exclusion criteria | <i>Streptococcus sobrinus</i><br><i>Eubacterium brachy</i>                                                                                                                                                                                                                                                                                                                                    | -                                                                                                                                                                                                                            |
| Moon <i>et al.</i> 2015 <sup>70</sup>     | Subgingival plaque  | 134 smokers<br>134 non-smokers                                                                               | Moderate periodontitis                                                                   | <i>Fusobacteria</i><br><i>Fusobacterium</i><br><i>Freitbacterium</i><br><i>Streptococcus</i><br><i>Veillonella</i><br><i>Corynebacterium</i><br>TM7<br><i>Filifactor</i><br><i>Parvimonas</i><br><i>Tannerella</i><br><i>Rothia</i><br><i>Lautropia</i><br><i>Eubacterium</i><br><i>Gemella</i><br><i>Granulicatella</i><br><i>Mogibacterium</i><br><i>Bacteroides</i><br><i>Leptotrichia</i> | Bacteroidetes<br><i>Neisseria</i><br><i>Campylobacter</i><br><i>Haemophilus</i><br><i>Prevotella</i><br><i>Aggregatibacter</i><br><i>Selenomonas</i><br><i>Veillonella</i><br><i>Porphyromonas</i><br><i>Cardiobacterium</i> |

|                                           |                                            |                                                                    |                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                               |
|-------------------------------------------|--------------------------------------------|--------------------------------------------------------------------|--------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Mason <i>et al.</i> 2015 <sup>71</sup>    | Subgingival plaque                         | 100 smokers<br>100 never smokers                                   | Periodontally healthy                      | <i>Fusobacterium nucleatum</i><br><i>Fusobacterium naviforme</i><br><i>Filifactor alocis</i><br><i>Dialister microaerophilus</i><br><i>Megasphaera sueciensis</i><br><i>Megasphaera geminatus</i><br><i>Megasphaera elsdenii</i><br><i>Megasphaera micronuciformis</i><br><i>Acinetobacter johnsonii</i><br><i>Acinetobacter guillouiae</i><br><i>Acinetobacter schindleri</i><br><i>Acinetobacter baumannii</i><br><i>Acinetobacter haemolyticus</i><br><i>Pseudomonas pseudoalcaligenes</i><br><i>Pseudoramibacter alactolyticus</i> | <i>Streptococcus sanguinis</i><br><i>Streptococcus parasanguinis</i><br><i>Streptococcus oralis</i><br><i>Gemella elegans</i><br><i>Gemella adiacens</i><br><i>Acinetobacter viscosus</i><br><i>Acinetobacter israelii</i><br><i>Acinetobacter dentalis</i><br><i>Neisseria subflava</i><br><i>Haemophilus parainfluenzae</i> |
| Wu <i>et al.</i> 2016 <sup>65</sup>       | Saliva                                     | 112 smokers<br>571 former smokers<br>521 never smokers             | Not measured, no oral examinations         | Actinobacteria<br>Firmicutes<br><i>Atopobium</i><br><i>Streptococcus</i><br><i>Bifidobacterium</i><br><i>Lactobacillus</i><br><i>Veillonella</i>                                                                                                                                                                                                                                                                                                                                                                                       | Proteobacteria<br><i>Neisseria</i><br><i>Haemophilus</i><br><i>Aggregatibacter</i><br><i>Capnocytophaga</i><br><i>Corynebacterium</i><br><i>Porphyromonas</i><br><i>Prevotella</i><br><i>Peptostreptococcus</i><br><i>Leptotrichia</i><br><i>Abiotrophia</i><br><i>Selenomonas</i>                                            |
| Amer <i>et al.</i> 2017 <sup>61</sup>     | Mucosal swabs (OLK study, subset analysis) | 19 current smokers<br>10 former smokers<br>7 never smokers         | Not measured                               | -                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | <i>Fusobacterium nucleatum</i><br><i>Neisseria</i><br><i>Leptotrichia</i>                                                                                                                                                                                                                                                     |
| Hsiao <i>et al.</i> 2018 <sup>68</sup>    | Saliva                                     | 104 male smokers<br>47 male never smokers                          | Not measured                               | <i>Fusobacterium nucleatum</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | -                                                                                                                                                                                                                                                                                                                             |
| Wirth <i>et al.</i> 2020 <sup>69</sup>    | Saliva                                     | 11 smokers<br>11 non-smokers (8 never, 3 former - > 1yr cessation) | Moderate and severe periodontitis excluded | <i>Prevotella</i><br><i>Megasphaera</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | <i>Neisseria</i><br><i>Capnocytophaga</i><br><i>Porphyromonas</i><br><i>Oribacterium</i>                                                                                                                                                                                                                                      |
| Gopinath <i>et al.</i> 2021 <sup>62</sup> | Buccal mucosa swabs                        | 17 male smokers<br>13 male non-smokers                             | Not measured, no oral examinations         | <i>Fusobacterium</i> spp. incl<br><i>Fusobacterium nucleatum</i><br><i>Saccheribacterium</i> spp.<br><i>Shuttleworthia</i> spp.                                                                                                                                                                                                                                                                                                                                                                                                        | -                                                                                                                                                                                                                                                                                                                             |

My findings in general agreed with the above, current smoking being associated with reduced abundance of several species of *Neisseria* and *Streptococcus* genera and *Lautropia*



*mirabilis*, *Gemella morbillorum*, *Rothia aeria*, *Selenomonas artemidis* and *Abiotrophia defectiva*. Several Gram-negative and plaque and periodontal disease-associated species from genera *Aggregatibacter*, *Campylobacter*, *Kingella*, *Bergeyella*, *Fusobacterium*, *Capnocytophaga* also showed significant decreases in abundance in association with smoking (**Table 25, Figure 16**). The species showing increased abundance associated with smoking were, as expected, predominantly plaque and periodontal disease-associated pathogens including species of *Prevotella*, *Actinomyces*, *Porphyromonas*, *Fusobacterium*, *Tannerella*, *Parvimonas*, *Filifactor*, *Clostridiales* and *Peptostreptococcus* genera. Several *Streptococcus* spp. were also significantly increased. I did not identify any red complex periodontal disease organisms (*Porphyromonas gingivalis*, *Tannerella forsythia*, *Treponema denticola*)<sup>236</sup> associated with smoking, which is not unexpected given that we sampled mucosa and not plaque. Furthermore, although there is agreement that smoking is a significant risk factor for periodontal disease, it has been shown that the risk is not simply due to smoking's effect on the subgingival microbiota which can be variable<sup>236</sup>.

## 6.6 Conclusion

Oral site and smoking were found to be the most significant influences on the microbiome of normal oral mucosal surfaces, the influence of site being such that mucosal microbial colonisation should be considered site specific. Tongue sites were dominated by *Rothia* and *Veillonella* spp. along with *Granulicatella adiacens*, *Gemella sanguinis* and *Haemophilus pittmaniae*. Palatal colonisation was dominated by *Prevotella* spp., *Actinomyces* spp. and *Streptococcus* spp. FOM/VT sites differed little from Buc/Alv/Gin sites, both sites being dominated by streptococci.

Smoking also had a profound influence on the microbiome. Smoking was associated with a shift to a more periodontal pathogen-heavy microbiome dominated by *Prevotella*, *Actinomyces*, *Porphyromonas*, *Fusobacterium*, *Tannerella*, *Parvimonas*, *Filifactor*, *Clostridiales* and *Peptostreptococcus* species, and diminished in healthy mucosa-associated species such as those of *Neisseria*, *Streptococcus* and *Rothia* genera, mirroring many of the mucosal microbial changes associated with OSCC and raising the question of the role of these changes in the initiation of OSCC.

Brushing habits influenced the microbiome with brushing once or less per day associated with reduced abundance of several early plaque-associated species. Poor oral hygiene was associated with an increase in periodontal pathogens, although the adjusted p values were not significant. Mouthwash use had minimal influence on the microbiome, being associated with reduced abundance of only one streptococcal species. Tooth number also had an effect on the microbiome. Missing 5 to 15 teeth was associated with an increase in some streptococcal and plaque-associated species, while missing more than 15 teeth was mostly associated with loss of species, including several oral commensals. Complete denture wear was associated with reduced abundance of several plaque- and periodontal-disease associated species, in keeping with the absence of teeth. Finally, alcohol consumption did not have a significant influence on the microbiome of healthy mucosal surfaces.

In conclusion, I have demonstrated that smoking and oral site are the most significant influencers of the oral microbiome, the influence of site being such that the different oral sites, palate, dorsum/lateral tongue, and buccal mucosa/alveolus/gingiva, should ideally be investigated separately. Interestingly floor of mouth/ventral tongue sites showed very similar microbial colonisation patterns to buccal mucosa/alveolus/gingiva and so could

potentially be combined for future analysis. However, as the floor of mouth/ventral tongue is considered a high-risk site for OSCC, I chose to analyse these sites separately.

Smoking has arguably the most profound influence on the microbiome, potentiating a switch to a Gram-negative anaerobic heavy microbiome, similar to that of OSCC, with associated loss of mucosal-health associated species. Many OSCC and OLK associated species are influenced by smoking with *Fusobacterium* spp. increased in smoking, OSCC and OLK; *Prevotella*, *Porphyromonas*, *Parvimonas*, *Filifactor* and *Peptostreptococcus* spp. increased in both smoking and OSCC; and *Veillonella* spp. increased in smoking and OLK (**Table 41**, **Figure 34**). Oral hygiene also has an effect on the microbiome, with poor oral hygiene associated with increased abundance of plaque and periodontal disease associated species, including *Prevotella* spp., which are also associated with smoking and OSCC.

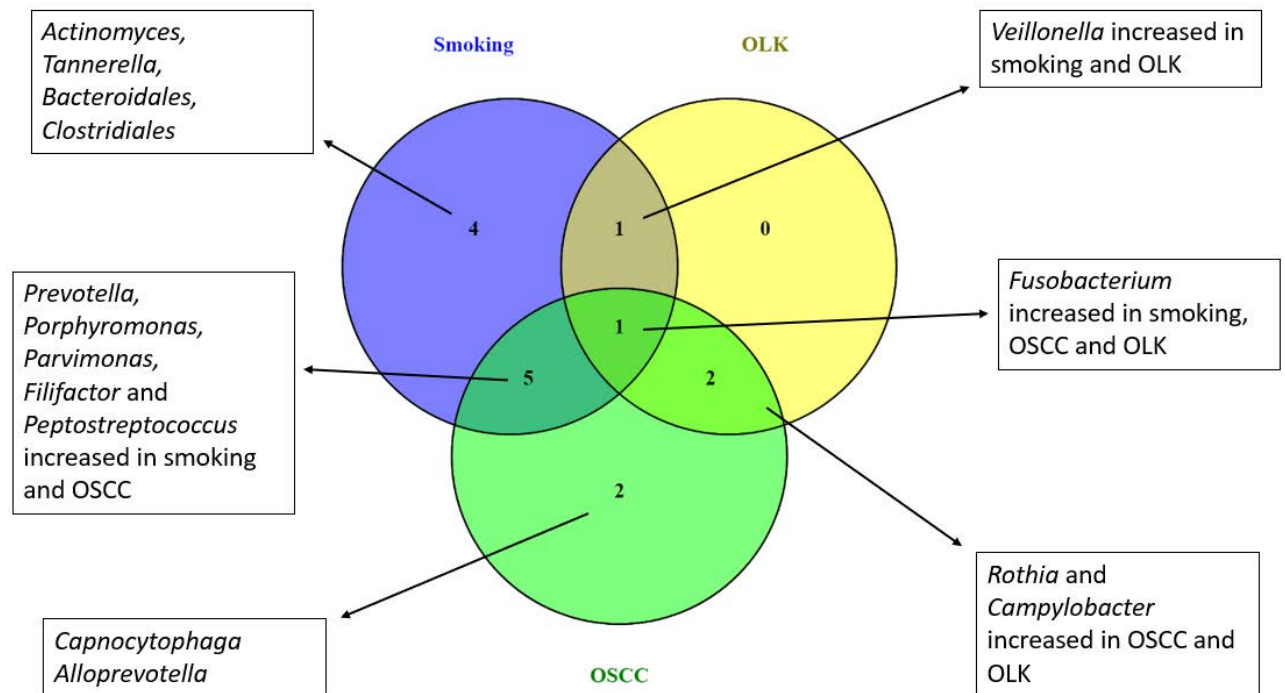


**Figure 34.** Venn diagram showing overlap of the genera associated with smoking, OSCC and OLK from previous studies.

Patient related factors which significantly influence the oral microbiome have been identified in this study, with some of these factors causing microbial shifts which mirror those of OSCC. These findings also emphasise the need for the use of contralateral normal sites, rather than control sites from healthy individuals, in OSCC and OLK microbiome studies to ensure that all possible confounding variables can be controlled for.

With regards to the aims of this chapter:

- Smoking, oral site, brushing habits, poor oral hygiene, tooth loss and denture wear were all found to influence the microbial colonisation of healthy mucosal surfaces to varying degrees.

## 7 The microbiome of OLK and the identification of bacterial taxa enriched on OLK

### 7.1 Introduction

The aim of this chapter was to investigate the primary hypothesis of the study, which was that the microbial colonisation of OLK differs to that of contralateral normal sites.

### 7.2 Methods

In order to determine if OLK was associated with specific microbiome changes, “Disease” (OLK) and “Contralateral normal” (CLN) sites from the patient cohort were compared. A total of 216 OLK from 177 patients, and 202 CLN sites from 180 patients (the same CLN cohort as Chapter 6) were available for analysis (**Figure 35**). The patient numbers in the OLK and CLN cohorts differed due to the DNA analysis; if there was insufficient DNA in an OLK (or CLN) sample, but sufficient DNA in the CLN (or OLK) sample from the same patient, the CLN (or OLK) sample was included in the microbial analysis. Although all patients were sampled at both disease and healthy control sites, in some patients, contralateral sites were also affected by OLK and so an alternative similarly keratinised or non-keratinised site was chosen. Due to the nature of sublingual keratoses and the fact they can extend across the midline, there were significantly more floor of mouth/ventral tongue (FOM/VT) disease than control sites. Lateral tongue and buccal mucosa sites were chosen as alternative keratinised and non-keratinised control sites when a matched contralateral normal site was not available and so there were significantly more lateral tongue and buccal mucosa control than disease sites ( $p=0.0005$ ) (**Table 43**). Data were analysed as described

in **section 4.4**. MaAsLin2 was used for multivariate analysis to investigate: (1) if the microbial colonisation of OLK differed to that of contralateral normal sites, and (2) if smoking, alcohol consumption, brushing habit, oral hygiene, denture wear, mouthwash use, tooth number, mucosal site, age, gender or clinical appearance influenced the colonisation patterns of OLK. Dorsum tongue (DT) and lateral border tongue (LBT) samples and alveolus/gingiva (Alv/Gin) and buccal mucosa (BM) samples were combined for analysis as DT/LBT and Buc/Alv/Gin respectively as previously described (**section 6.2**). DESeq2 was used to analyse the species level count data of OLK versus CLN sites as a second measure to ensure consistency of results. An initial comparison of the OLK and CLN sites with the 120 healthy control (HC) samples detailed in **Chapter 6** was also carried out.

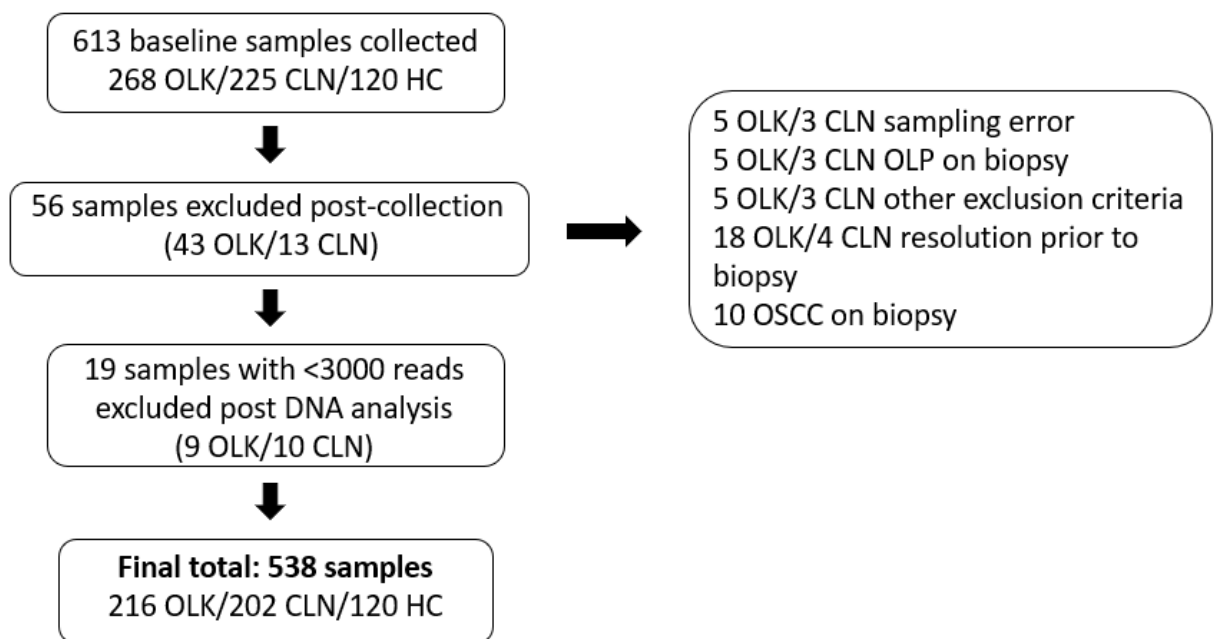


**Figure 35.** Flowchart of sample cohort

## 7.3 Results

The demographic data of the sample cohort and healthy controls are shown in **Table 42**.

**Table 42.** Demographic data of sample cohort and healthy controls

| Demographic               | OLK patients | OLK sites | Healthy controls |
|---------------------------|--------------|-----------|------------------|
| Total                     | 177          | 216       | 61               |
|                           |              |           |                  |
| <b>Gender</b>             |              |           |                  |
| Males                     | 92           | 111       | 21               |
| Females                   | 85           | 105       | 40               |
|                           |              |           |                  |
| <b>Age range</b>          |              |           |                  |
| <40                       | 18           | 20        | 17               |
| 40-60                     | 83           | 102       | 30               |
| >60                       | 76           | 94        | 13               |
| Not recorded              | 0            | 0         | 1                |
|                           |              |           |                  |
| <b>Smoking status</b>     |              |           |                  |
| Current smoker            | 92           | 109       | 15               |
| Former smoker             | 51           | 66        | 15               |
| Never smoker              | 34           | 41        | 31               |
|                           |              |           |                  |
| <b>Alcohol units/week</b> |              |           |                  |
| 0                         | 37           | 42        | 13               |
| 1 to 10                   | 93           | 111       | 41               |
| 11 to 20                  | 22           | 28        | 2                |
| More than 20              | 25           | 35        | 5                |
|                           |              |           |                  |
| <b>Missing teeth</b>      |              |           |                  |
| <5                        | 25           | 30        | 26               |
| 5 to 15                   | 75           | 101       | 26               |
| More than 15              | 34           | 42        | 7                |
| Not recorded              | 43           | 43        | 2                |
|                           |              |           |                  |
| <b>OHI Score</b>          |              |           |                  |
| Good (OHI-S < 0.5)        | 40           | 49        | 37               |
| Fair (OHI-S 0.5-1)        | 60           | 72        | 15               |
| Poor (OHI-S > 1)          | 37           | 45        | 7                |
| Not recorded              | 40           | 50        | 2                |
|                           |              |           |                  |
| <b>Brushing times/day</b> |              |           |                  |
| <1                        | 18           | 22        | 0                |
| 1                         | 28           | 37        | 5                |
| >1                        | 120          | 151       | 56               |
| Not recorded              | 11           | 6         | 0                |
|                           |              |           |                  |
| <b>Denture</b>            |              |           |                  |
| Complete                  | 7            | 8         | 1                |
| Partial                   | 32           | 38        | 7                |
| None                      | 138          | 170       | 53               |
|                           |              |           |                  |
| <b>Mouthwash use</b>      |              |           |                  |
| Yes                       | 83           | 96        | 26               |
| No                        | 94           | 120       | 35               |

**Table 43** shows the sites sampled and clinical appearance of the OLKs along with the CLN and HC sites.

**Table 43.** OLK, CLN and HC sites and clinical appearance of OLK of final sample cohort

| Sample data                        | % OLK (n=216)    | % CLN (n=202) | %HC (n=120) |
|------------------------------------|------------------|---------------|-------------|
| <b>Sites samples (n=428)</b>       |                  |               |             |
| Floor of mouth/ventral tongue      | 28.2 (61)        | 12.9 (26)     | 14.2 (17)   |
| Buccal mucosa/Alveolus/Gingiva     | 37 (80)          | 41.6 (84)     | 35.9 (43)   |
| Dorsum/lateral tongue              | 22.7 (49)        | 36.1 (73)     | 35.9 (43)   |
| Palate                             | 12 (26)          | 9.4 (19)      | 14.2 (17)   |
|                                    |                  |               |             |
| <b>Clinical appearance (n=216)</b> |                  |               |             |
| Homogenous                         | 83.8 (181)       | NA            | NA          |
| Non-homogenous                     | 16.2 (35)        | NA            | NA          |
| - of which ulcerated               | 3.2 of total (7) |               |             |

### 7.3.1 Comparison of the microbiome of OLK with CLN and HC sites

Significant differences between the alpha diversities of healthy controls and OLK were noted on the ACE and Shannon indices, reflecting differences in the microbial richness and evenness between these groups. The alpha diversities of HC and CLN sites also differed significantly on the Simpson index. No significant differences were noted between the alpha diversities of CLN and OLK which is not surprising as these samples came from the same patient cohort. (**Figure 36, Table 44**). The differences between the alpha diversities of HC and both CLN and OLK were also not unexpected considering that the healthy control cohort and patient cohort (OLK and CLN) differed significantly (**Table 22**) in all variables measured. However, it serves to emphasise that CLN samples are preferable to healthy control samples in studies aiming to establish if the microbiome of OLK differs from clinically normal mucosa. All further OLK analysis was therefore carried out based on comparison to CLN, and not healthy control, samples.



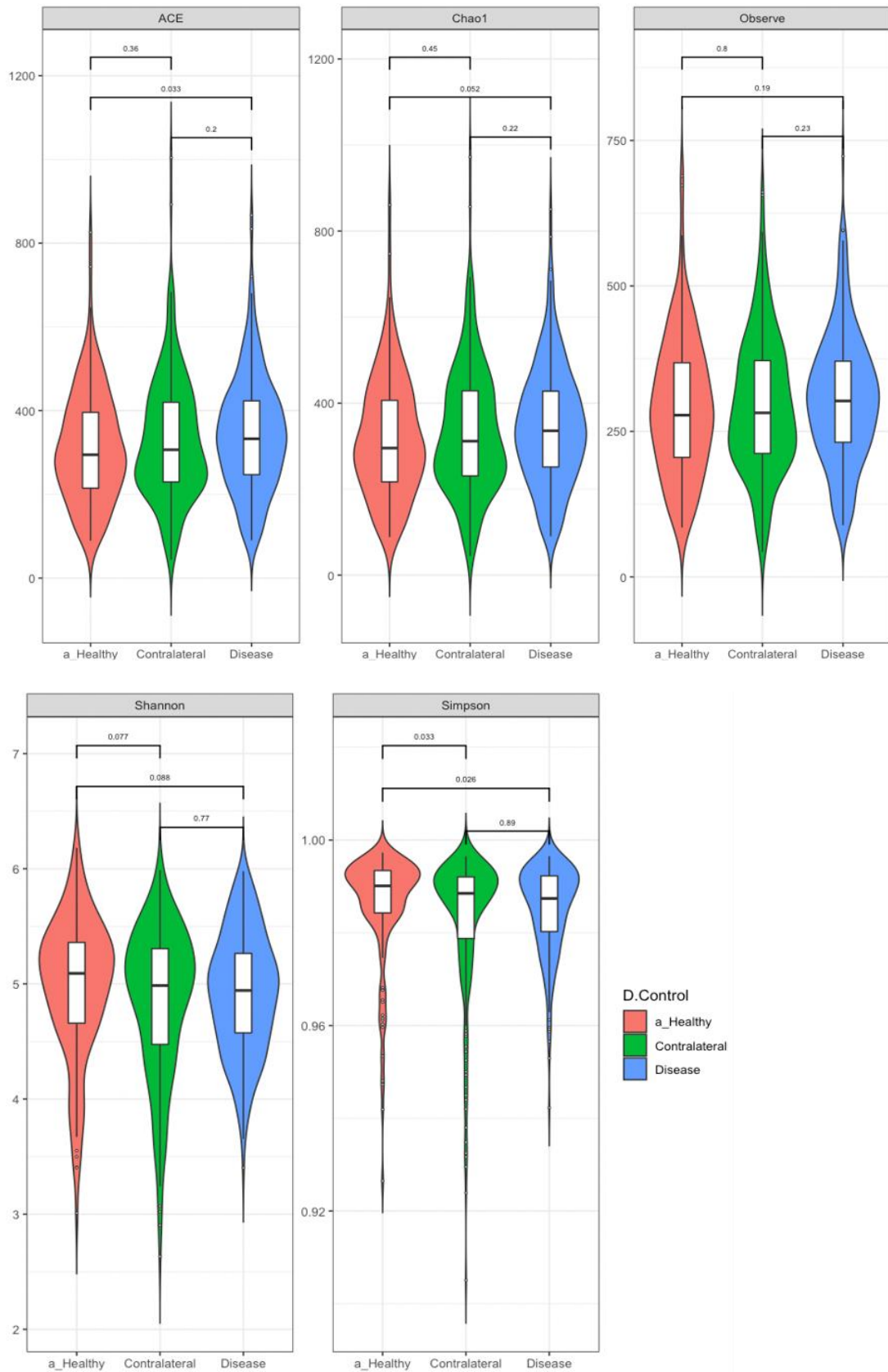


**Figure 36.** Alpha diversity plots of disease, contralateral normal and healthy controls using ACE, Chao1, Observe, Shannon and Simpson indices. Y axis shows index values, p values shown above square brackets and also in Table 44 below.

**Table 44.** Table showing p values from comparisons of alpha diversities of HC (Healthy controls), CLN (contralateral normal) and OLK sites on different indices (i.e., p values from **Figure 30**). Significant results in bold.

| Alpha diversity index | ACE          | Chao1 | Observe | Shannon | Simpson      |
|-----------------------|--------------|-------|---------|---------|--------------|
| Comparison            |              |       |         |         |              |
| HC v CLN              | 0.36         | 0.45  | 0.8     | 0.077   | <b>0.033</b> |
| HC v OLK              | <b>0.033</b> | 0.052 | 0.19    | 0.088   | <b>0.026</b> |
| CLN v OLK             | 0.2          | 0.22  | 0.23    | 0.77    | 0.23         |

Matrices of the beta-diversity values were visualised using non-metric multidimensional scaling (NMDS) (not shown). Pairwise PERMANOVA testing on these values confirmed significant separation of the communities of OLK and CLN ( $p_{adj}=0.01$ ). PCA plots of the relative abundances also demonstrated separation of disease (OLK) and CLN sites, with disease sites being dominated by species of genera *Rothia*, *Prevotella*, *Fusobacterium*, *Leptotrichia*, *Porphyromonas* and *Gemella* while contralateral normal sites were dominated by streptococci. (**Figure 37**).

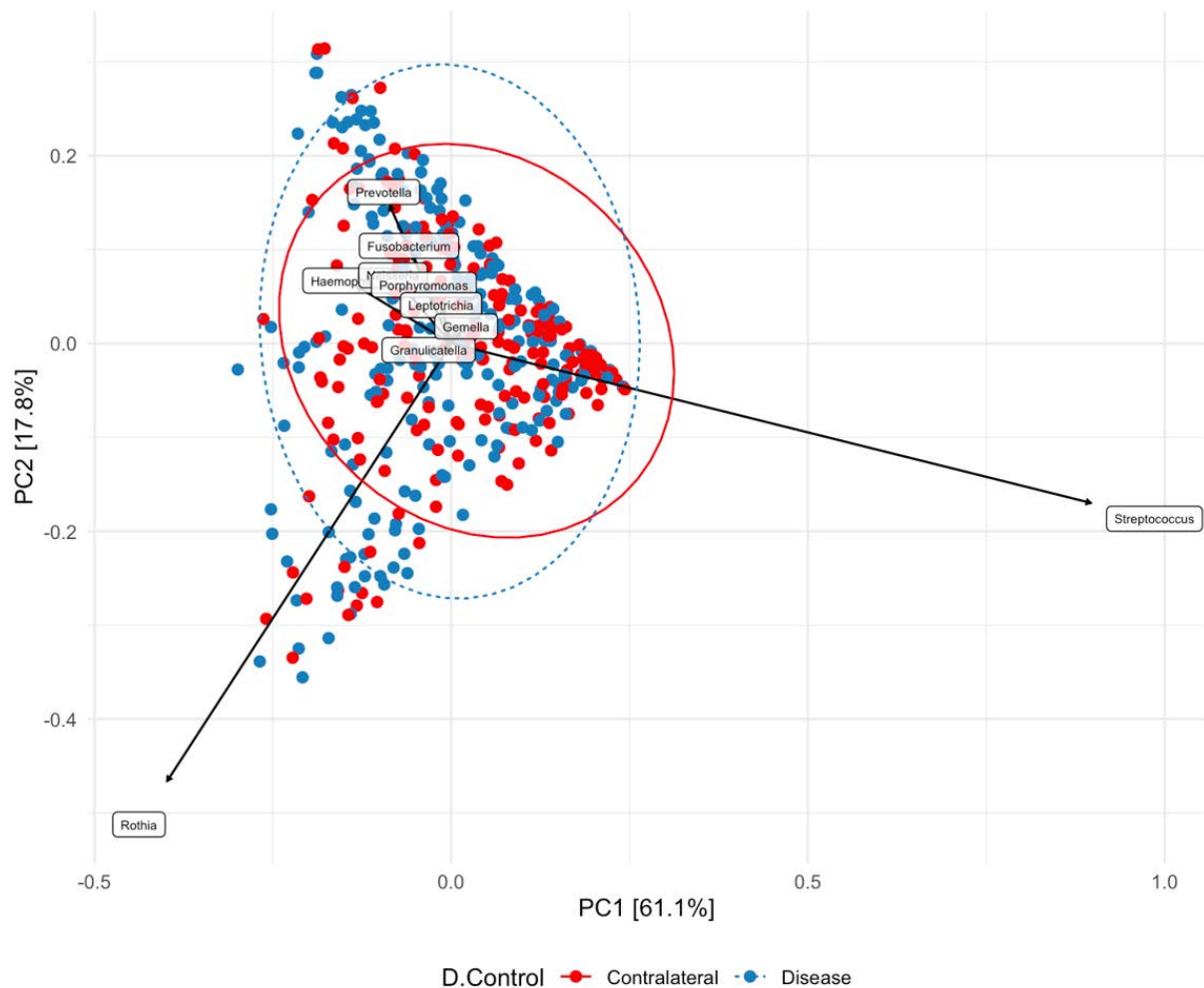


**Figure 37.** Axis 1 and 2 PCA plot of disease (OLK) and contralateral normal (CLN) sites showing separation of the microbial communities of OLK and CLN sites and the genera responsible for these population shifts. Contralateral sites show increased abundances of streptococci while disease sites are characterised by increased abundances of *Rothia*, *Prevotella*, *Fusobacterium*, *Porphyromonas*, *Leptotrichia* and *Gemella*.

Bar plots displaying the most abundant phyla and genera associated with CLN and OLK sites clearly demonstrate these differences in colonisation patterns (**Figure 38**, **Figure 39**).

Firmicutes was the most abundant phylum, followed by Actinobacteria, Proteobacteria, Bacteroidetes and Fusobacteria, in both CLN and OLK. However, there were noticeable differences between the cohorts with OLKs characterised by reduced abundance of Firmicutes and increased abundance of Actinobacteria, Proteobacteria, Bacteroidetes and Fusobacteria compared with CLN sites (**Figure 37**).

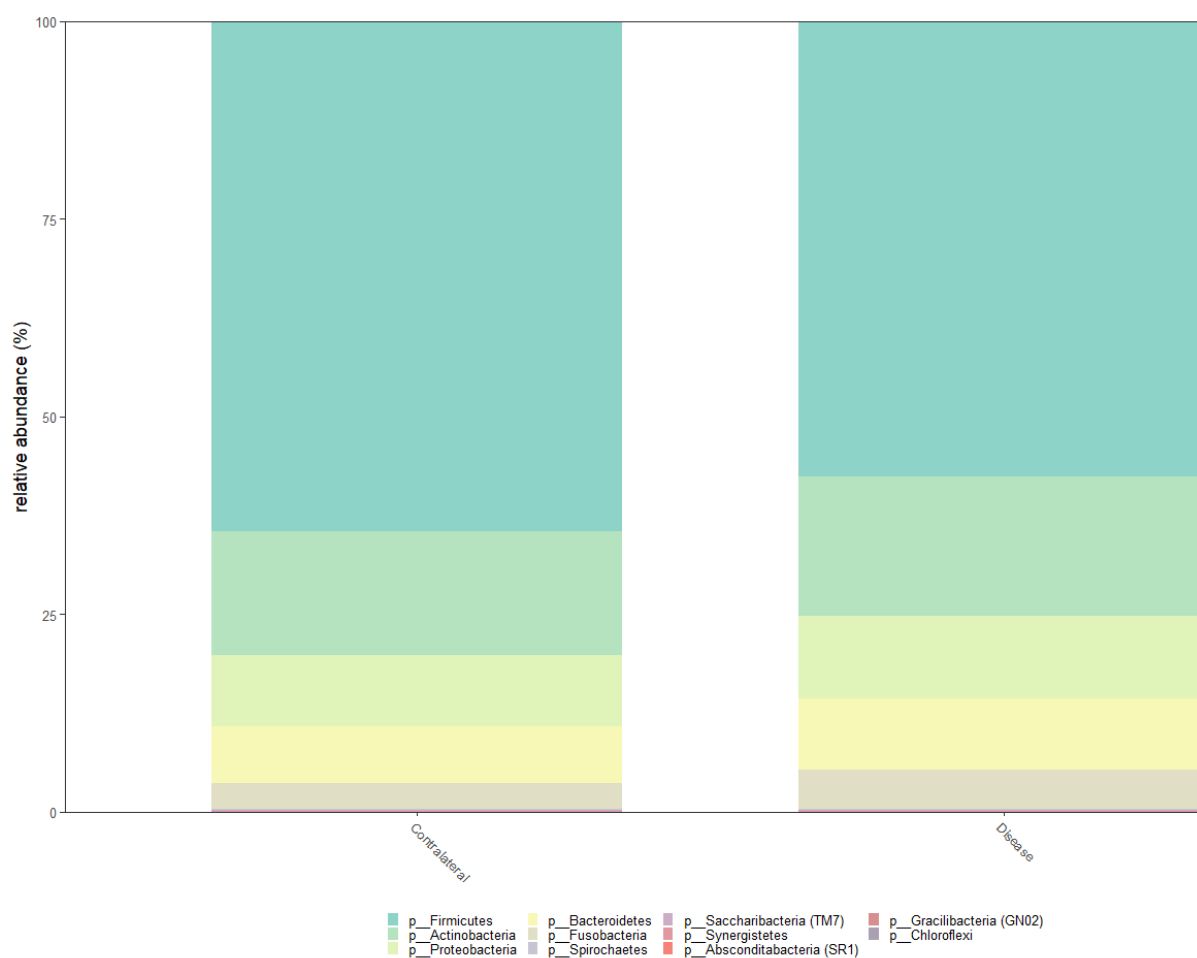


**Figure 38.** Barplot showing the most abundant phyla associated with CLN (Contralateral) and OLK (Disease)

*Streptococcus* was the dominant genus at both sites; however, as at phyla level, the abundance patterns differed between the sites, with OLK sites showing reduced abundance of streptococci compared with CLN sites. Abundances of *Rothia*, *Haemophilus*, *Prevotella*, *Fusobacterium* and *Porphyromonas* also differed between the sites; all showing increased abundance at OLK compared with CLN sites (**Figure 39**).

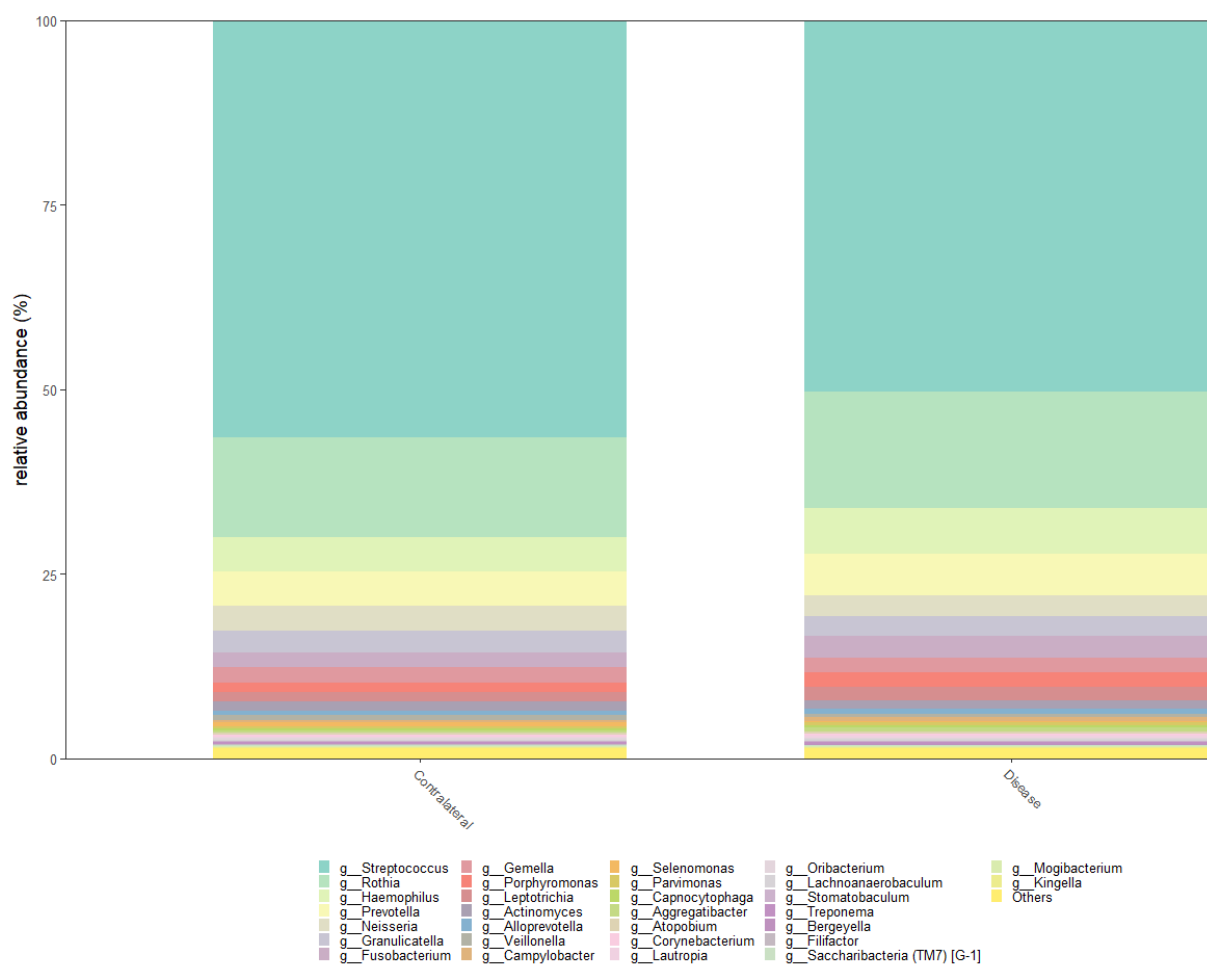


**Figure 39.** Bar plot showing the most abundant genera associated with CLN and OLK sites

MaAsLin2 was then used for multivariate analysis to compare the microbial colonisation patterns of OLK with CLN sites at both genera and species levels. Three genera, *Bergeyella*, *Leptotrichia* and *Fusobacterium*, showed increased abundance, and one genus, *Streptococcus* showed reduced abundance, associated with OLK, although the adjusted p values for *Fusobacterium* and *Streptococcus* did not reach significance (**Table 45**).

**Table 45.** Genera showing changes in abundance associated with OLK on MaAslin2 analysis (negative coefficient value denotes reduced abundance) \*=padj<0.1

| Genera               | Variable | Coefficient | p value | Adjusted p value |
|----------------------|----------|-------------|---------|------------------|
| <i>Bergeyella</i>    | OLK      | 0.96        | 0.0005  | 0.03*            |
| <i>Leptotrichia</i>  | OLK      | 0.99        | 0.001   | 0.04*            |
| <i>Fusobacterium</i> | OLK      | 0.68        | 0.02    | 0.24             |
| <i>Streptococcus</i> | OLK      | -0.18       | 0.01    | 0.24             |

At species level, five taxa, *Bergeyella* HMT322, *Leptotrichia* HMT215, *Streptococcus australis infantis* clade 431, *Gemella morbillorum*, and *Leptotrichia* HMT392, showed significantly increased abundance associated with OLK (all  $p \leq 0.002$ ,  $p_{adj} < 0.1$ ). Three other taxa, *Porphyromonas* sp. HMT284, *Fusobacterium periodonticum nucleatum* subsp. *polymorphum hwasookii* and *Lachnoanaerobaculum umeaense*, were also associated with OLK, although the adjusted p values were not significant ( $p < 0.05$ ,  $p_{adj} > 0.1$ ) (**Table 46**).

**Table 46.** Taxa with significant changes in abundance associated with OLK on MaAsLin2 analysis.

| Taxa                                                                                               | Variable | Coefficient | p value | Adjusted p value |
|----------------------------------------------------------------------------------------------------|----------|-------------|---------|------------------|
| <i>Bergeyella</i> sp. HMT322                                                                       | OLK      | 0.89        | 0.0001  | 0.02             |
| <i>Leptotrichia</i> sp. HMT215                                                                     | OLK      | 0.95        | 0.0007  | 0.06             |
| <i>Streptococcus australis infantis</i> clade 431                                                  | OLK      | 1.08        | 0.002   | 0.09             |
| <i>Gemella morbillorum</i>                                                                         | OLK      | 0.88        | 0.002   | 0.09             |
| <i>Leptotrichia</i> sp. HMT392                                                                     | OLK      | 0.48        | 0.002   | 0.09             |
| <i>Porphyromonas</i> sp. HMT284                                                                    | OLK      | 0.56        | 0.004   | 0.13             |
| <i>Fusobacterium periodonticum nucleatum</i> subsp. <i>polymorphum</i> sp. HMT204 <i>hwasookii</i> | OLK      | 1.06        | 0.007   | 0.18             |
| <i>Lachnoanaerobaculum umeaense</i>                                                                | OLK      | 0.56        | 0.008   | 0.18             |

A second, confirmatory analysis was carried out using DESeq2. This analysis of OLK sites also identified *Bergeyella* HMT322, *Streptococcus australis infantis* clade 431, *Haemophilus parainfluenzae* (associated with OLK after adjusting for site (see **section 7.2.9**)) and *Lachnoanaerobaculum umeaense* as associated with OLK, although the adjusted p values for *Haemophilus parainfluenzae* and *Lachnoanaerobaculum umeaense* were not significant (**Table 47**). *Streptococcus infantis* clade 431 and a *Fusobacterium* taxon denoted *periodonticum nucleatum* subsp. *polymorphum* also showed significantly increased abundance associated with OLK on DESeq2 analysis. Three species, *Streptococcus* sp. HMT423 *mitis*, another streptococcal species and *Gemella haemolysans*, showed significantly reduced abundance associated with OLK on the DESeq2 analysis.

**Table 47.** Taxa with significant changes in abundance associated with OLK on DESeq2 analysis (negative log2 values denote reduced abundance)

| Taxa                                                                   | Variable | log2 Fold Change | p value  | Adjusted p value |
|------------------------------------------------------------------------|----------|------------------|----------|------------------|
| <i>Streptococcus australis infantis</i> clade 431                      | OLK      | 1.98             | 6.52E-06 | 0.0006           |
| <i>Streptococcus infantis</i> clade 431                                | OLK      | 1.76             | 0.0002   | 0.005            |
| <i>Fusobacterium periodonticum nucleatum</i> subsp. <i>polymorphum</i> | OLK      | 1.18             | 0.003    | 0.04             |
| <i>Bergeyella</i> sp. HMT322                                           | OLK      | 1.07             | 0.003    | 0.04             |
| <i>Lachnoanaerobaculum umeaense</i>                                    | OLK      | 1.50             | 0.02     | 0.18             |
| <i>Haemophilus parainfluenzae</i>                                      | OLK      | 0.56             | 0.04     | 0.23             |
| <i>Streptococcus oralis infantis</i>                                   | OLK      | -0.57            | 0.002    | 0.04             |
| <i>Gemella haemolysans</i>                                             | OLK      | -1.00            | 0.002    | 0.04             |
| <i>Streptococcus</i> sp. HMT423 <i>mitis</i>                           | OLK      | -1.16            | 5.63E-05 | 0.003            |

### 7.3.2 Microbiome associated with OLK adjusted for smoking

MaAsLin2 was then used to investigate the effect of smoking and OLK on microbial colonisation and whether smoking habit influenced OLK colonisation. The five taxa identified as associated with OLK, *Bergeyella* HMT322, *Leptotrichia* HMT215, *Streptococcus australis infantis* clade 431, *Gemella morbillorum*, and *Leptotrichia* HMT392, remained significantly associated with OLK after adjusting for smoking (all padj <0.1). None of these taxa showed increased abundance associated with either current or former smoking, while the *Bergeyella* species was significantly reduced in current smokers (p=0.01, padj=0.09), emphasising that these associations with OLK were not smoking-associated changes. Furthermore, the three other OLK-associated taxa whose adjusted p values failed to reach significance (**Table 46**), *Porphyromonas* sp. HMT284, *Fusobacterium periodonticum nucleatum* subsp. *polymorphum* sp. HMT204 *hwasookii* and *Lachnoanaerobaculum umeaense*, were found to be significantly associated with OLK after adjusting for smoking (padj=0.03, 0.05 and 0.05 respectively), along with an additional *Fusobacterium* taxa, *Fusobacterium nucleatum* subsp. *polymorphum hwasookii*

(padj=0.09). As discussed in Chapter 6 (**section 6.3.1**), smoking has a profound influence on the oral microbiome and these same smoking-related changes, namely depletion of *Neisseria* spp., *Rothia* spp. and some streptococci, were also present when the presence of an OLK was taken into account (**Table 48**). Significantly, however, the microbial changes seen on OLKs were independent of smoking habit (**Figure 40**).

**Table 48.** Taxa with significant changes in abundance associated with OLK adjusted for smoking on MaAsLin2 analysis (negative coefficient values denote reduced abundance)

| Taxa                                                                                               | Variable       | Coefficient | p value  | Adjusted p value |
|----------------------------------------------------------------------------------------------------|----------------|-------------|----------|------------------|
| <i>Bergeyella</i> sp. HMT322                                                                       | OLK            | 0.87        | 0.0001   | 0.003            |
| <i>Leptotrichia</i> sp. HMT215                                                                     | OLK            | 0.93        | 0.0008   | 0.01             |
| <i>Gemella morbillorum</i>                                                                         | OLK            | 0.89        | 0.002    | 0.02             |
| <i>Leptotrichia</i> sp. HMT392                                                                     | OLK            | 0.49        | 0.002    | 0.02             |
| <i>Streptococcus australis infantis</i> clade 431                                                  | OLK            | 1.06        | 0.003    | 0.03             |
| <i>Porphyromonas</i> sp. HMT284                                                                    | OLK            | 0.58        | 0.004    | 0.03             |
| <i>Fusobacterium periodonticum nucleatum</i> subsp. <i>polymorphum</i> sp. HMT204 <i>hwasookii</i> | OLK            | 1.08        | 0.006    | 0.05             |
| <i>Lachnoanaerobaculum umeaense</i>                                                                | OLK            | 0.56        | 0.006    | 0.05             |
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum hwasookii</i>                                 | OLK            | 0.52        | 0.01     | 0.09             |
| <i>Neisseria perflava flavescens</i>                                                               | Current smoker | -6.40       | 1.61E-28 | 8.95E-26         |
| <i>Streptococcus australis</i>                                                                     | Current smoker | -3.52       | 7.03E-21 | 1.95E-18         |
| <i>Lautropia mirabilis</i>                                                                         | Current smoker | -2.17       | 5.49E-12 | 7.62E-10         |
| <i>Rothia aeria</i>                                                                                | Current smoker | -2.15       | 4.46E-12 | 7.62E-10         |
| <i>Prevotella histicola</i>                                                                        | Current smoker | 3.42        | 1.05E-10 | 1.17E-08         |
| <i>Atopobium parvulum</i>                                                                          | Current smoker | 2.64        | 3.43E-09 | 3.17E-07         |
| <i>Stomatobaculum</i> sp. HMT097                                                                   | Current smoker | -1.52       | 8.86E-08 | 6.14E-06         |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 071                                      | Current smoker | 2.20        | 1.01E-07 | 6.24E-06         |
| <i>Haemophilus parainfluenzae</i>                                                                  | Current smoker | -2.52       | 2.43E-07 | 1.35E-05         |
| <i>Streptococcus</i> sp. HMT066 sp. HMT074                                                         | Current smoker | -2.40       | 8.44E-07 | 3.91E-05         |
| <i>Porphyromonas endodontalis</i>                                                                  | Current smoker | 2.04        | 8.25E-07 | 3.91E-05         |
| <i>Bacteroidales</i> G 2 <i>bacterium</i> HMT274                                                   | Current smoker | 1.38        | 1.10E-06 | 4.69E-05         |
| <i>Fusobacterium periodonticum hwasookii</i>                                                       | Current smoker | -2.03       | 1.80E-06 | 7.13E-05         |
| <i>Streptococcus parasanguinis</i> clade 411 <i>parasanguinis</i> clade 721                        | Current smoker | 2.24        | 2.19E-06 | 8.10E-05         |
| <i>Stomatobaculum longum</i>                                                                       | Current smoker | 1.49        | 3.41E-06 | 0.0001           |
| <i>Tannerella forsythia</i>                                                                        | Current smoker | 1.42        | 4.36E-06 | 0.0001           |
| <i>Lachnoanaerobaculum umeaense</i>                                                                | Current smoker | -1.26       | 4.63E-06 | 0.0001           |
| <i>Prevotella pleuritidis</i>                                                                      | Current smoker | 1.51        | 4.89E-06 | 0.0001           |
| <i>Alloprevotella tannerae</i>                                                                     | Current smoker | 2.03        | 6.39E-06 | 0.0002           |
| <i>Peptostreptococcaceae</i> XI G9 <i>brachy</i>                                                   | Current smoker | 2.11        | 9.37E-06 | 0.0002           |
| <i>Streptococcus mutans</i>                                                                        | Current smoker | 1.49        | 1.35E-05 | 0.0003           |
| <i>Prevotella</i> sp. HMT313                                                                       | Current smoker | 1.81        | 3.33E-05 | 0.0008           |



|                                                                         |                |       |          |          |
|-------------------------------------------------------------------------|----------------|-------|----------|----------|
| <i>Streptococcus</i> sp. HMT423 <i>mitis</i>                            | Current smoker | 2.02  | 3.52E-05 | 0.0008   |
| <i>Streptococcus peroris</i>                                            | Current smoker | 0.95  | 0.0001   | 0.002    |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 070           | Current smoker | 1.56  | 0.0002   | 0.003    |
| <i>Peptostreptococcaceae</i> XI G5 <i>saphenum</i>                      | Current smoker | 1.26  | 0.0002   | 0.003    |
| <i>Filifactor alocis</i>                                                | Current smoker | 1.45  | 0.0002   | 0.003    |
| <i>Fusobacterium</i> sp. HMT203                                         | Current smoker | 1.28  | 0.0002   | 0.003    |
| <i>Actinomyces graevenitzi</i>                                          | Current smoker | 1.77  | 0.0002   | 0.003    |
| <i>Mycoplasma faucium</i>                                               | Current smoker | 1.09  | 0.0002   | 0.004    |
| <i>Veillonella</i> sp. HMT780                                           | Current smoker | 1.28  | 0.0003   | 0.004    |
| <i>Tannerella</i> sp. HMT286                                            | Current smoker | -1.04 | 0.0003   | 0.004    |
| <i>Parvimonas micra</i>                                                 | Current smoker | 1.47  | 0.0004   | 0.006    |
| <i>Porphyromonas pasteri</i>                                            | Current smoker | -1.57 | 0.0005   | 0.006    |
| <i>Clostridiales</i> F1 G1 <i>bacterium</i> HMT093                      | Current smoker | 0.60  | 0.0005   | 0.006    |
| <i>Abiotrophia defectiva</i>                                            | Current smoker | -0.76 | 0.0008   | 0.010    |
| <i>Actinomyces lingnae</i>                                              | Current smoker | 1.45  | 0.0008   | 0.01     |
| <i>Streptococcus thermophilus salivarius</i>                            | Current smoker | 1.22  | 0.0008   | 0.01     |
| <i>Capnocytophaga sputigena</i>                                         | Current smoker | -0.89 | 0.0009   | 0.01     |
| <i>Veillonella rogosae dispar parvula atypica</i>                       | Current smoker | 1.09  | 0.00     | 0.02     |
| <i>Peptostreptococcaceae</i> XI G4 <i>bacterium</i> HMT369              | Current smoker | 0.81  | 0.00     | 0.02     |
| <i>Fusobacterium nucleatum</i> subsp. <i>vincentii</i> <i>naviforme</i> | Current smoker | 1.62  | 0.00     | 0.02     |
| <i>Gemella sanguinis</i>                                                | Current smoker | -1.27 | 0.00     | 0.03     |
| <i>Streptococcus</i> sp. HMT066                                         | Current smoker | -1.30 | 0.00     | 0.03     |
| <i>Streptococcus vestibularis salivarius</i>                            | Current smoker | 1.22  | 0.00     | 0.03     |
| <i>Alloprevotella</i> sp. HMT473                                        | Current smoker | 0.93  | 0.00     | 0.03     |
| <i>Streptococcus parasanguinis</i> clade 411                            | Current smoker | 1.40  | 0.00     | 0.04     |
| <i>Streptococcus</i> sp. HMT061                                         | Current smoker | -1.06 | 0.00     | 0.04     |
| <i>Leptotrichia hofstadii</i>                                           | Current smoker | -0.90 | 0.01     | 0.05     |
| <i>Parvimonas</i> sp. HMT393                                            | Current smoker | -0.81 | 0.01     | 0.05     |
| <i>Fusobacterium nucleatum</i> subsp. <i>vincentii</i>                  | Current smoker | 1.09  | 0.01     | 0.06     |
| <i>Prevotella salivae</i>                                               | Current smoker | 1.32  | 0.01     | 0.06     |
| <i>Actinomyces odontolyticus</i>                                        | Current smoker | 0.88  | 0.01     | 0.08     |
| <i>Saccharibacteria</i> TM7 G1 <i>bacterium</i> HMT352                  | Current smoker | 0.69  | 0.01     | 0.08     |
| <i>Bergeyella</i> sp. HMT322                                            | Current smoker | -0.75 | 0.01     | 0.09     |
| <i>Streptococcus sanguinis</i>                                          | Current smoker | -1.23 | 0.01     | 0.09     |
| <i>Streptococcus</i> sp. HMT057                                         | Current smoker | 1.09  | 0.01     | 0.09     |
| <i>Prevotella denticola</i>                                             | Current smoker | 0.90  | 0.01     | 0.09     |
| <i>Veillonella denticariosi</i>                                         | Current smoker | 1.02  | 0.01     | 0.09     |
| <i>Streptococcus australis</i>                                          | Former smoker  | -2.35 | 4.49E-09 | 3.56E-07 |
| <i>Prevotella histicola</i>                                             | Former smoker  | 2.41  | 2.90E-05 | 0.001    |
| <i>Tannerella</i> sp. HMT286                                            | Former smoker  | -1.26 | 7.40E-05 | 0.002    |
| <i>Neisseria perflava flavescens</i>                                    | Former smoker  | -2.21 | 0.0002   | 0.003    |
| <i>Rothia aeria</i>                                                     | Former smoker  | -1.22 | 0.0003   | 0.004    |
| <i>Lautropia mirabilis</i>                                              | Former smoker  | -1.09 | 0.00     | 0.02     |
| <i>Fusobacterium</i> sp. HMT204                                         | Former smoker  | -0.99 | 0.00     | 0.02     |
| <i>Streptococcus</i> sp. HMT074                                         | Former smoker  | -1.62 | 0.00     | 0.03     |
| <i>Porphyromonas pasteri</i>                                            | Former smoker  | -1.48 | 0.00     | 0.03     |

|                                                   |               |       |      |      |
|---------------------------------------------------|---------------|-------|------|------|
| <i>Prevotella</i> sp. HMT305                      | Former smoker | 0.57  | 0.01 | 0.05 |
| <i>Stomatobaculum</i> sp. HMT097                  | Former smoker | -0.84 | 0.01 | 0.05 |
| <i>Streptococcus</i> sp. HMT423 <i>mitis</i>      | Former smoker | 1.44  | 0.01 | 0.06 |
| <i>Lachnospiraceae</i> G3 <i>bacterium</i> HMT100 | Former smoker | -0.75 | 0.01 | 0.07 |
| <i>Stomatobaculum longum</i>                      | Former smoker | 0.90  | 0.01 | 0.08 |
| <i>Atopobium parvulum</i>                         | Former smoker | 1.24  | 0.01 | 0.08 |
| <i>Alloprevotella</i> sp. HMT473                  | Former smoker | 0.90  | 0.01 | 0.08 |
| <i>Streptococcus sanguinis</i>                    | Former smoker | -1.37 | 0.01 | 0.09 |

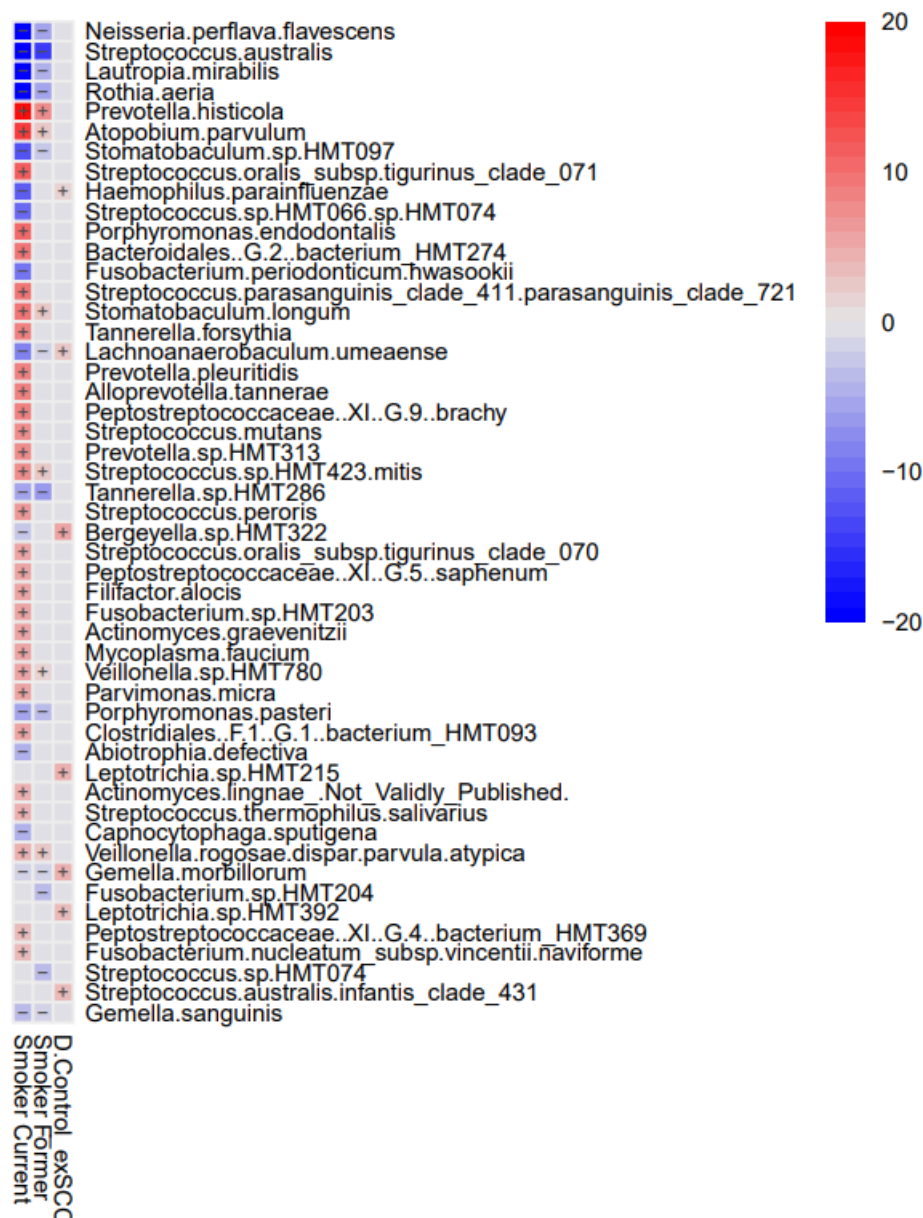


**Figure 40.** Heatmap showing taxa with significant changes in abundance associated with OLK and smoking with effect size calculated as  $-\log(q\text{-value}) * (\text{coefficient})$ . (D.Control\_exSCC=OLK excluding OSCC samples [n=10]. Red=increased, blue=decreased)

### 7.3.3 Microbiome associated with OLK adjusted for alcohol

The five OLK-associated taxa, *Bergeyella* HMT322, *Leptotrichia* HMT215, *Streptococcus australis infantis* clade 431, *Gemella morbillorum*, and *Leptotrichia* HMT392, remained associated with OLK after adjusting for alcohol consumption (all  $p \leq 0.007$ ), with the adjusted p values significant for *Bergeyella* HMT322 and *Leptotrichia* sp. HMT215. Three other previously identified OLK-associated taxa, *Porphyromonas* sp. HMT284, *Fusobacterium periodonticum nucleatum* subsp. *polymorphum hwasookii* and *Lachnoanaerobaculum umeaense*, were also associated with OLK after adjusting for alcohol, although the adjusted p values did not reach significance ( $p \leq 0.008$ ,  $p_{adj} > 0.1$ ). Consuming more than 20 units of alcohol per week was associated with significantly increased abundance of *Leptotrichia* sp. HMT221, *Prevotella salivae*, *Saccharibacteria* TM7 G1 bacterium HMT352, *Actinomyces odontolyticus* and *Megasphaera micronuciformis* (Table 49, Figure 41), changes not seen in the CLN and HC analysis (Chapter 6, section 6.3.2).

**Table 49.** Taxa significantly associated with OLK adjusted for alcohol consumption on MaAsLin2 analysis (negative values denote reduced abundance)

| Taxa                                                                                               | Variable       | Coefficient | p value  | Adjusted p value |
|----------------------------------------------------------------------------------------------------|----------------|-------------|----------|------------------|
| <i>Bergeyella</i> sp. HMT322                                                                       | OLK            | 0.89        | 0.0001   | 0.02             |
| <i>Leptotrichia</i> sp. HMT215                                                                     | OLK            | 0.95        | 0.0007   | 0.08             |
| <i>Gemella morbillorum</i>                                                                         | OLK            | 0.87        | 0.002    | 0.13             |
| <i>Leptotrichia</i> sp. HMT392                                                                     | OLK            | 0.48        | 0.002    | 0.13             |
| <i>Streptococcus australis infantis</i> clade 431                                                  | OLK            | 1.06        | 0.003    | 0.14             |
| <i>Porphyromonas</i> sp. HMT284                                                                    | OLK            | 0.56        | 0.005    | 0.16             |
| <i>Fusobacterium periodonticum nucleatum</i> subsp. <i>polymorphum</i> sp. HMT204 <i>hwasookii</i> | OLK            | 1.07        | 0.007    | 0.20             |
| <i>Lachnoanaerobaculum umeaense</i>                                                                | OLK            | 0.56        | 0.008    | 0.20             |
| <i>Leptotrichia</i> sp. HMT221                                                                     | >20 units/week | 2.15        | 6.21E-07 | 0.00             |
| <i>Prevotella salivae</i>                                                                          | >20 units/week | 2.47        | 5.99E-05 | 0.01             |
| <i>Saccharibacteria</i> TM7 G1 bacterium HMT352                                                    | >20 units/week | 1.35        | 4.60E-05 | 0.01             |
| <i>Actinomyces odontolyticus</i>                                                                   | >20 units/week | 1.50        | 0.0005   | 0.07             |
| <i>Megasphaera micronuciformis</i>                                                                 | >20 units/week | 1.20        | 0.0007   | 0.08             |

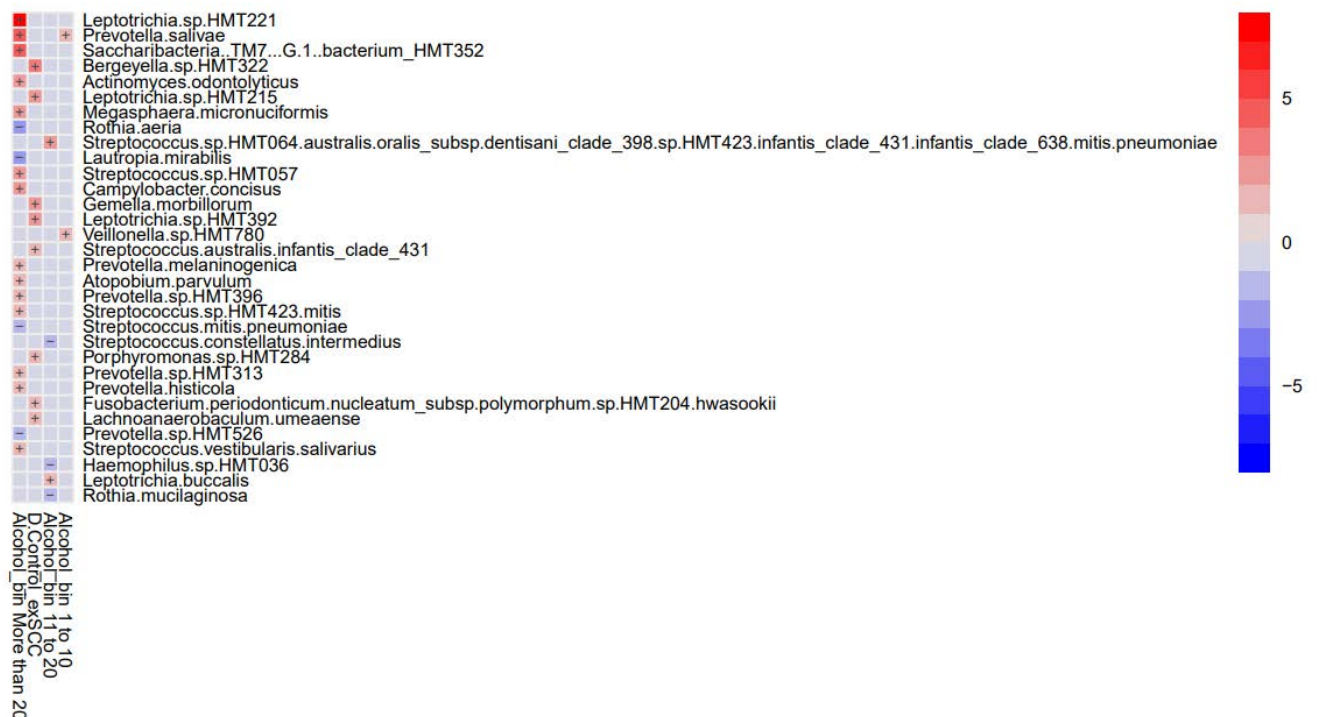


**Figure 41.** Heatmap showing taxa with significant changes in abundance associated with OLK and alcohol consumption. Effect size calculated as  $-\log(q\text{-value}) \times (\text{coefficient})$ . (D.Control\_exSCC=OLK excluding OSCC samples [n=10]. Alcohol\_bin 1 to 10 = 1-10 units/week, 11-20 = 11 to 20 units/week, more than 20 = >20 units/week. Red=increased, blue=decreased)

### 7.3.4 Microbiome associated with OLK adjusted for brushing habit

The association between brushing habit and OLK colonisation was also assessed using MaAsLin2. All previously identified taxa associated with OLK were also present on the OLK and brushing model (all  $p < 0.05$ ), with the adjusted p values significant for *Bergeyella* sp. HMT322, *Gemella morbillorum*, *Streptococcus australis infantis* clade 431 and *Leptotrichia* sp. HMT215. None of the OLK-associated taxa were associated with brushing habit (Table 50, Figure 42).

**Table 50.** Taxa significantly associated with OLK adjusted for brushing habit (per day) on MaAsLin2 analysis (negative coefficient values denote reduced abundance)

| Taxa                                                                                               | Variable  | Coefficient | p value | Adjusted p value |
|----------------------------------------------------------------------------------------------------|-----------|-------------|---------|------------------|
| <i>Bergeyella</i> sp. HMT322                                                                       | OLK       | 0.82        | 0.001   | 0.09             |
| <i>Leptotrichia</i> sp. HMT215                                                                     | OLK       | 0.89        | 0.002   | 0.09             |
| <i>Gemella morbillorum</i>                                                                         | OLK       | 0.91        | 0.002   | 0.09             |
| <i>Streptococcus australis infantis</i> clade 431                                                  | OLK       | 1.12        | 0.002   | 0.09             |
| <i>Fusobacterium periodonticum nucleatum</i> subsp. <i>polymorphum</i> sp. HMT204 <i>hwasookii</i> | OLK       | 1.17        | 0.004   | 0.11             |
| <i>Leptotrichia</i> sp. HMT392                                                                     | OLK       | 0.45        | 0.005   | 0.12             |
| <i>Porphyromonas</i> sp. HMT284                                                                    | OLK       | 0.59        | 0.005   | 0.12             |
| <i>Lachnoanaerobaculum umeaense</i>                                                                | OLK       | 0.52        | 0.01    | 0.25             |
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum</i> <i>hwasookii</i>                          | OLK       | 0.53        | 0.01    | 0.25             |
| <i>Gemella haemolysans</i>                                                                         | Once/less | 1.53        | 0.0009  | 0.09             |
| <i>Parvimonas</i> sp. HMT393                                                                       | Once/less | 0.79        | 0.002   | 0.09             |
| <i>Streptococcus mitis pneumoniae</i>                                                              | Once/less | 1.62        | 0.004   | 0.11             |
| <i>Abiotrophia defectiva</i>                                                                       | Once/less | 0.55        | 0.005   | 0.12             |
| <i>Haemophilus parahaemolyticus</i>                                                                | Once/less | 0.73        | 0.008   | 0.16             |
| <i>Granulicatella elegans</i>                                                                      | Once/less | 0.82        | 0.01    | 0.22             |
| <i>Streptococcus cristatus</i> clade 578                                                           | Once/less | -1.17       | 0.001   | 0.09             |
| <i>Oribacterium sinus</i>                                                                          | Once/less | -1.06       | 0.001   | 0.09             |
| <i>Streptococcus</i> sp. HMT064 <i>oralis</i> subsp. <i>dentisani</i> clade 398                    | Once/less | -0.78       | 0.001   | 0.09             |
| <i>Rothia dentocariosa</i>                                                                         | Once/less | -1.15       | 0.002   | 0.09             |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 071 <i>sanguinis</i>                     | Once/less | -0.89       | 0.004   | 0.11             |
| <i>Fusobacterium nucleatum</i> subsp. <i>vincentii naviforme</i>                                   | Once/less | -1.20       | 0.009   | 0.18             |
| <i>Corynebacterium matruchotii</i>                                                                 | Once/less | -0.76       | 0.01    | 0.25             |
| <i>Streptococcus sanguinis</i>                                                                     | Once/less | -1.03       | 0.02    | 0.25             |



**Figure 42.** Heatmap showing taxa with significant changes in abundance associated with OLK and brushing habit. Effect size calculated as  $-\log(q\text{-value}) * (\text{coefficient})$ . (D.Control\_exSCC=OLK excluding OSCC samples [n=10], Brushing\_2 = Brushing once or less than once per day. Red=increased, blue=decreased)

### 7.3.5 Microbiome associated with OLK adjusted for oral hygiene

Investigation of the influence of oral hygiene on OLK found that three of the OLK-associated taxa remained associated with OLK when oral hygiene category was considered. *Streptococcus australis infantis* clade 431, *Leptotrichia* sp. HMT215 and *Gemella morbillorum* showed increased abundance associated with OLK, although the adjusted p values were not significant (all  $p < 0.05$ ,  $\text{padj} > 0.1$ ). *Streptococcus australis infantis* clade 431 also showed increased abundance in association with fair oral hygiene ( $p = 0.0001$ ,  $\text{padj} = 0.07$ ), although interestingly not with poor oral hygiene. No other OLK-associated species was influenced by oral hygiene (Table 51, Figure 43).

**Table 51.** Taxa significantly associated with OLK adjusted for oral hygiene category on MaAsLin2 analysis (negative coefficient values denote reduced abundance, OH=oral hygiene)

| Taxa                                                          | Variable | Coefficient | p value | Adjusted p value |
|---------------------------------------------------------------|----------|-------------|---------|------------------|
| <i>Streptococcus australis infantis</i> clade 431             | OLK      | 1.30        | 0.001   | 0.12             |
| <i>Leptotrichia</i> sp. HMT215                                | OLK      | 0.93        | 0.004   | 0.16             |
| <i>Gemella morbillorum</i>                                    | OLK      | 0.88        | 0.009   | 0.24             |
| <i>Streptococcus australis infantis</i> clade 431             | Fair OH  | 1.87        | 0.0001  | 0.07             |
| <i>Streptococcus cristatus</i> clade 578                      | Fair OH  | 1.49        | 0.0006  | 0.11             |
| <i>Prevotella histicola</i>                                   | Fair OH  | 1.77        | 0.002   | 0.12             |
| <i>Actinomyces lingnae</i>                                    | Fair OH  | 1.43        | 0.002   | 0.12             |
| <i>Atopobium parvulum</i>                                     | Fair OH  | 1.45        | 0.002   | 0.14             |
| <i>Saccharibacteria</i> TM7 G1 bacterium HMT349               | Fair OH  | 0.76        | 0.004   | 0.16             |
| <i>Actinomyces graevenitzii</i>                               | Fair OH  | 1.37        | 0.007   | 0.22             |
| <i>Streptococcus thermophilus salivarius</i>                  | Fair OH  | 1.02        | 0.007   | 0.22             |
| <i>Selenomonas</i> sp. HMT136                                 | Fair OH  | -0.90       | 0.001   | 0.12             |
| <i>Leptotrichia hofstadii</i>                                 | Fair OH  | -0.97       | 0.008   | 0.23             |
| <i>Streptococcus oralis</i> subsp. <i>dentisani</i> clade 398 | Poor OH  | 1.55        | 0.0006  | 0.11             |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 071 | Poor OH  | 1.55        | 0.002   | 0.12             |
| <i>Granulicatella elegans</i>                                 | Poor OH  | 1.37        | 0.003   | 0.16             |
| <i>Streptococcus</i> sp. HMT423 <i>mitis</i>                  | Poor OH  | 1.57        | 0.003   | 0.16             |
| <i>Streptococcus mitis pneumoniae</i>                         | Poor OH  | 2.15        | 0.005   | 0.18             |
| <i>Streptococcus gordonii</i>                                 | Poor OH  | 1.78        | 0.006   | 0.21             |
| <i>Gemella haemolysans</i>                                    | Poor OH  | 1.64        | 0.008   | 0.23             |
| <i>Rothia aerea</i>                                           | Poor OH  | -1.10       | 0.005   | 0.19             |



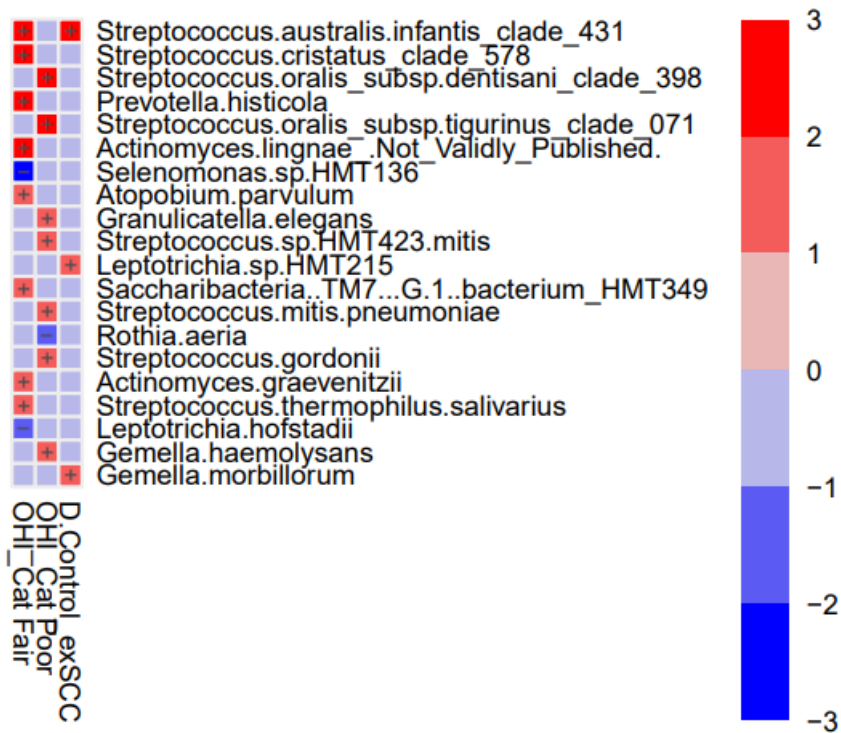


**Figure 43.** Heatmap showing taxa with significant changes in abundance associated with OLK and oral hygiene category. (Effect size calculated as  $-\log(q\text{-value}) \times (\text{coefficient})$ , Red=increased, blue=decreased. D.Control\_exSCC=OLK excluding OSCC samples [n=10], OHI\_Cat=oral hygiene category)

### 7.3.6 Microbiome associated with OLK adjusted for denture wear

Denture wear was not found to significantly affect the colonisation pattern of OLK. Six of the previously identified taxa associated with OLK were also significantly associated with OLK after adjusting for denture wear. *Bergeyella* HMT322, *Leptotrichia* HMT215, *Streptococcus australis infantis* clade 431, *Gemella morbillorum*, *Leptotrichia* HMT392 and *Porphyromonas* sp. HMT284 all showed significantly increased abundance associated with OLK in an OLK-Denture MaAsLin2 model (all  $p < 0.05$ ,  $\text{padj} < 0.1$ ) (**Table 52**). Only one OLK-associated species was influenced by denture wear, *Bergeyella* HMT322 showing increased abundance associated with complete denture wear, although the adjusted p value was not significant ( $p = 0.01$ ,  $\text{padj} = 0.13$  data not shown).



**Table 52.** Taxa significantly associated with OLK adjusted for denture wear on MaAsLin2 analysis (negative coefficient values denote reduced abundance)

| Taxa                                                             | Variable         | Coefficient | p value  | Adjusted p value |
|------------------------------------------------------------------|------------------|-------------|----------|------------------|
| <i>Bergeyella</i> sp. HMT322                                     | OLK              | 0.90        | 9.72E-05 | 0.009            |
| <i>Leptotrichia</i> sp. HMT215                                   | OLK              | 0.95        | 0.0007   | 0.03             |
| <i>Gemella morbillorum</i>                                       | OLK              | 0.86        | 0.002    | 0.06             |
| <i>Streptococcus australis infantis</i> clade 431                | OLK              | 1.08        | 0.002    | 0.07             |
| <i>Leptotrichia</i> sp. HMT392                                   | OLK              | 0.48        | 0.002    | 0.07             |
| <i>Porphyromonas</i> sp. HMT284                                  | OLK              | 0.56        | 0.004    | 0.09             |
| <i>Veillonella rogosae dispar parvula atypica</i>                | Complete denture | 1.91        | 0.005    | 0.096            |
| <i>Rothia dentocariosa</i>                                       | Complete denture | -4.56       | 2.55E-07 | 0.0001           |
| <i>Streptococcus cristatus</i> clade 578                         | Complete denture | -3.32       | 4.58E-05 | 0.007            |
| <i>Streptococcus sanguinis</i>                                   | Complete denture | -3.59       | 0.000    | 0.013            |
| <i>Parvimonas micra</i>                                          | Complete denture | -3.13       | 0.000    | 0.013            |
| <i>Porphyromonas endodontalis</i>                                | Complete denture | -3.11       | 0.000    | 0.013            |
| <i>Fusobacterium nucleatum</i> subsp. <i>vincentii naviforme</i> | Complete denture | -3.87       | 0.000    | 0.015            |
| <i>Alloprevotella tannerae</i>                                   | Complete denture | -2.99       | 0.001    | 0.039            |
| <i>Streptococcus constellatus intermedius</i>                    | Complete denture | -2.93       | 0.001    | 0.057            |
| <i>Corynebacterium matruchotii</i>                               | Complete denture | -2.12       | 0.003    | 0.070            |
| <i>Peptostreptococcaceae</i> XI G9 <i>brachy</i>                 | Complete denture | -2.71       | 0.005    | 0.096            |
| <i>Streptococcus sanguinis</i>                                   | Partial denture  | -2.45       | 3.66E-07 | 0.0001           |
| <i>Streptococcus cristatus</i> clade 578                         | Partial denture  | -1.62       | 6.71E-05 | 0.007            |
| <i>Gemella morbillorum</i>                                       | Partial denture  | -1.45       | 6.18E-05 | 0.007            |
| <i>Lautropia mirabilis</i>                                       | Partial denture  | -1.01       | 0.002    | 0.06             |
| <i>Prevotella</i> sp. HMT317                                     | Partial denture  | -0.74       | 0.002    | 0.06             |
| <i>Rothia aeria</i>                                              | Partial denture  | -0.97       | 0.002    | 0.06             |
| <i>Actinomyces</i> sp. HMT169                                    | Partial denture  | -0.91       | 0.003    | 0.07             |
| <i>Leptotrichia hofstadii</i>                                    | Partial denture  | -0.97       | 0.003    | 0.07             |
| <i>Parvimonas</i> sp. HMT110                                     | Partial denture  | -0.64       | 0.003    | 0.07             |
| <i>Corynebacterium durum</i>                                     | Partial denture  | -0.72       | 0.003    | 0.07             |
| <i>Streptococcus intermedius</i>                                 | Partial denture  | -0.73       | 0.003    | 0.07             |

### 7.3.7 Microbiome associated with OLK adjusted for mouthwash use

Mouthwash use did not influence the microbiome of OLK. All five of the initially identified OLK-associated taxa, *Bergeyella* HMT322, *Leptotrichia* HMT215, *Streptococcus australis infantis* clade 431, *Gemella morbillorum* and *Leptotrichia* HMT392, demonstrated significantly increased abundance associated with OLK after adjustment for mouthwash use (all  $p < 0.05$ ,  $p_{adj} < 0.1$ ) and no OLK-associated taxon showed a significant change in abundance associated with mouthwash use (**Table 53**).

**Table 53.** Taxa significantly associated with OLK adjusted for mouthwash use on MaAsLin2 analysis (negative coefficient values denote reduced abundance)

| Taxa                                              | Variable      | Coefficient | p value  | Adjusted p value |
|---------------------------------------------------|---------------|-------------|----------|------------------|
| <i>Bergeyella</i> sp. HMT322                      | OLK           | 0.90        | 0.0001   | 0.01             |
| <i>Leptotrichia</i> sp. HMT215                    | OLK           | 0.92        | 0.0009   | 0.06             |
| <i>Gemella morbillorum</i>                        | OLK           | 0.87        | 0.002    | 0.08             |
| <i>Streptococcus australis infantis</i> clade 431 | OLK           | 1.09        | 0.002    | 0.09             |
| <i>Leptotrichia</i> sp. HMT392                    | OLK           | 0.47        | 0.003    | 0.09             |
| <i>Granulicatella adiacens</i>                    | Mouthwash use | 0.87        | 2.01E-05 | 0.004            |
| <i>Rothia mucilaginosa</i>                        | Mouthwash use | 0.83        | 0.0009   | 0.06             |
| <i>Streptococcus mitis pneumoniae</i>             | Mouthwash use | -2.10       | 2.08E-05 | 0.004            |
| <i>Haemophilus</i> sp. HMT036                     | Mouthwash use | -0.67       | 0.0004   | 0.04             |
| <i>Veillonella</i> sp. HMT780                     | Mouthwash use | -0.84       | 0.002    | 0.08             |
| <i>Corynebacterium matruchotii</i>                | Mouthwash use | -0.86       | 0.002    | 0.08             |

### 7.3.8 Microbiome associated with OLK and missing teeth

Tooth number did not appear to significantly influence the microbial colonisation patterns of OLK. Five OLK-associated taxa remained significantly associated with OLK after adjustment for missing teeth, *Bergeyella* HMT322, *Leptotrichia* HMT215, *Streptococcus australis infantis* clade 431, *Leptotrichia* HMT392 and *Fusobacterium periodonticum nucleatum* subsp. *polymorphum* sp. HMT204 *hwasookii* demonstrating significantly increased abundance associated with OLK on this model ( $p < 0.05$ ,  $p_{adj} < 0.1$ ). Two OLK-

associated taxa showed reduced abundance associated with number of missing teeth; *Fusobacterium nucleatum* subsp. *polymorphum hwasookii* with missing 5 to 15 teeth ( $p=0.009$ ,  $padj=0.1$ ), and *Gemella morbillorum* with missing more than 15 teeth ( $p=0.001$ ,  $padj=0.1$ ), although the adjusted p values did not reach significance (**Table 54**).

**Table 54.** Taxa significantly associated with OLK after adjusting for missing teeth and OLK-associated taxa influenced by number of missing teeth on MaAsLin2 analysis (Negative coefficient values denote reduced abundance).

| Taxa                                                                                               | Variable              | Coefficient | p value  | Adjusted p value |
|----------------------------------------------------------------------------------------------------|-----------------------|-------------|----------|------------------|
| <i>Bergeyella</i> sp. HMT322                                                                       | OLK                   | 1.06        | 5.36E-05 | 0.004            |
| <i>Leptotrichia</i> sp. HMT215                                                                     | OLK                   | 1.09        | 0.001    | 0.02             |
| <i>Streptococcus australis infantis</i> clade 431                                                  | OLK                   | 1.34        | 0.001    | 0.03             |
| <i>Leptotrichia</i> sp. HMT392                                                                     | OLK                   | 0.53        | 0.002    | 0.05             |
| <i>Fusobacterium periodonticum nucleatum</i> subsp. <i>polymorphum</i> sp. HMT204 <i>hwasookii</i> | OLK                   | 1.27        | 0.005    | 0.07             |
| <i>Fusobacterium periodonticum nucleatum</i> subsp. <i>polymorphum hwasookii</i>                   | Missing 5 to 15 teeth | -0.78       | 0.009    | 0.1              |
| <i>Gemella morbillorum</i>                                                                         | Missing >15 teeth     | -1.12       | 0.01     | 0.12             |

### 7.3.9 Microbiome associated with OLK adjusted for mucosal site

Mucosal site was next assessed to determine if the changes in species abundance identified on OLK were independent of anatomical site, as oral site was found to have a significant influence on the microbiome of clinically normal mucosa (**section 6.3.8**). Ten taxa, including the initial five OLK-associated taxa (**Table 46**) and *Fusobacterium nucleatum* subsp. *polymorphum hwasookii*, *Porphyromonas* sp. HMT284, *Lachnoanaerobaculum umeaense*, *Haemophilus parainfluenzae* and *Prevotella* sp. HMT396, were significantly associated with OLK when OLK and site were analysed together in a MaAsLin2 model (**Table 55**) (all  $p<0.05$ ,  $padj<0.1$ ). *Haemophilus parainfluenzae* and *Prevotella* sp. HMT396 had not previously been found to be significantly associated with OLK using MaAsLin2,

although *Haemophilus parainfluenzae* was associated with OLK on DESeq2 analysis ( $p=0.04$ ,  $\text{padj}=0.23$ ) (**Table 47**). Unsurprisingly, given my previous findings of microbiome site specificity (**section 6.3.8**), several OLK-associated species also showed significant changes in abundance associated with mucosal site (**Table 55**). *Leptotrichia* sp. HMT215 and *Prevotella* sp. HMT396 were significantly increased at PAL sites and *Bergeyella* sp. HMT322 was significantly increased at FOM/VT sites, while *Gemella morbillorum* was significantly reduced at PAL and DT/LBT sites and *Leptotrichia* sp. HMT392 was significantly reduced at DT/LBT sites. Finally, FOM/VT sites showed significantly reduced abundance of the same *Prevotella* sp. that showed increased abundance at PAL sites.

**Table 55.** Taxa significantly associated with OLK after adjusting for mucosal site on MaAsLin2 analysis. Site associated (if any) also shown with p value and adjusted p value (negative coefficient values denote reduced abundance). (Buccal mucosa/alveolus/gingiva used as reference, DT/LBT=dorsum tongue/lateral border tongue, FOM/VT=floor of mouth/ventral tongue, PAL=palate)

| Taxa                                                               | Variables associated | Coefficient OLK/Site | p value | Adjusted p value |
|--------------------------------------------------------------------|----------------------|----------------------|---------|------------------|
| <i>Bergeyella</i> sp. HMT322                                       | OLK                  | 0.77                 | 0.001   | 0.02             |
|                                                                    | FOM/VT               | 0.93                 | 0.003   | 0.03             |
| <i>Streptococcus australis infantis</i> clade 431                  | OLK                  | 1.17                 | 0.001   | 0.02             |
| <i>Leptotrichia</i> sp. HMT215                                     | OLK                  | 0.9                  | 0.001   | 0.02             |
|                                                                    | PAL                  | 1.84                 | 0.0001  | 0.004            |
| <i>Gemella morbillorum</i>                                         | OLK                  | 0.9                  | 0.002   | 0.02             |
|                                                                    | PAL                  | -1.11                | 0.02    | 0.09             |
|                                                                    | DT/LBT               | -0.78                | 0.02    | 0.1              |
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum hwasookii</i> | OLK                  | 0.58                 | 0.006   | 0.05             |
| <i>Lachnoanaerobaculum umeaense</i>                                | OLK                  | 0.64                 | 0.003   | 0.03             |
| <i>Porphyromonas</i> sp. HMT284                                    | OLK                  | 0.59                 | 0.004   | 0.03             |
| <i>Haemophilus parainfluenzae</i>                                  | OLK                  | 1.05                 | 0.007   | 0.05             |
| <i>Prevotella</i> sp. HMT396                                       | OLK                  | 0.55                 | 0.01    | 0.06             |
|                                                                    | PAL                  | 1.04                 | 0.003   | 0.03             |
|                                                                    | FOM/VT               | -0.74                | 0.009   | 0.06             |
| <i>Leptotrichia</i> sp. HMT392                                     | OLK                  | 0.4                  | 0.01    | 0.07             |
|                                                                    | DT/LBT               | -0.5                 | 0.08    | 0.06             |

In view of the significance of site on mucosal microbial colonisation patterns, additional analysis using DESeq2, to compare OLK and CLN samples at each individual mucosal site, was carried out in order to determine if any site-specific microbiome changes were associated with OLK (**Table 56**). There were significant abundance changes common to both DESeq2 and MaAsLin2 (OLK and site) analyses, namely increased abundances of *Streptococcus australis infantis* clade 431 and *Prevotella* sp. HMT396 associated with OLKs, although DESeq2 identified these species on Buc/Alv/Gin sites, whereas MaAsLin2 identified *Streptococcus australis infantis* clade 431 as associated with all OLKs and *Prevotella* sp. HMT396 as associated with both OLK and PAL sites. DESeq2 found increased abundance of *Rothia mucilaginosa* associated with Buc/Alv/Gin and PAL OLKs, and increased abundance of *Rothia aeria dentocariosa* associated with Buc/Alv/Gin OLKs. MaAsLin2 also identified both *Rothia mucilaginosa* and *Rothia aeria dentocariosa* as associated with OLK, however the adjusted p values were not significant ( $p=0.09$ ,  $p_{adj}=0.25$ , and  $p=0.04$ ,  $p_{adj}=0.23$  respectively, results not shown), suggesting that these species could be specific to either Buc/Alv/Gin and PAL sites or to OLKs at these sites. Finally, the DESeq2 analysis also identified significantly reduced abundance of *Streptococcus mitis* associated with Buc/Alv/Gin OLKs.

**Table 56.** Results of DESeq2 analysis of OLK versus contralateral normal sites by mucosal site (negative log2 fold change denotes reduced abundance) (Buc/Alv/Gin=buccal mucosa/alveolus/gingiva, DT/LBT=dorsum tongue/lateral border tongue, FOM/VT=floor of mouth/ventral tongue, PAL=palate)

| Mucosal site | Taxa                                                                                                                                                                                    | Variable | log2 Fold Change | p value  | Adjusted p value |
|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|------------------|----------|------------------|
| Buc/Alv/Gin  | <i>Rothia mucilaginosa</i>                                                                                                                                                              | OLK      | 1.53             | 1.65E-07 | 2.72E-05         |
| Buc/Alv/Gin  | <i>Streptococcus australis infantis</i> clade 431                                                                                                                                       | OLK      | 2.25             | 0.0003   | 0.01             |
| Buc/Alv/Gin  | <i>Rothia aerea dentocariosa</i>                                                                                                                                                        | OLK      | 1.66             | 0.0003   | 0.01             |
| Buc/Alv/Gin  | <i>Prevotella</i> sp. HMT396                                                                                                                                                            | OLK      | 2.04             | 0.0005   | 0.01             |
| Buc/Alv/Gin  | <i>Leptotrichia</i> sp. HMT417                                                                                                                                                          | OLK      | 1.69             | 0.003    | 0.06             |
| Buc/Alv/Gin  | <i>Veillonella rogosae dispar</i>                                                                                                                                                       | OLK      | 1.11             | 0.005    | 0.09             |
| Buc/Alv/Gin  | <i>Streptococcus</i> sp. HMT423 <i>mitis</i>                                                                                                                                            | OLK      | -1.68            | 3.80E-05 | 0.003            |
| Buc/Alv/Gin  | <i>Streptococcus</i> sp. HMT064 <i>australis</i> <i>oralis</i> subsp. <i>dentisani</i> clade 398 sp. HMT423 <i>infantis</i> clade 431 <i>infantis</i> clade 638 <i>mitis pneumoniae</i> | OLK      | -0.98            | 0.0003   | 0.01             |
| Buc/Alv/Gin  | <i>Streptococcus</i> sp. HMT064 <i>mitis</i>                                                                                                                                            | OLK      | -1.02            | 0.002    | 0.05             |
| PAL          | <i>Rothia mucilaginosa</i>                                                                                                                                                              | OLK      | 2.48             | 0.0002   | 0.07             |
| FOM/VT       | <i>Fusobacterium periodonticum hwasookii</i>                                                                                                                                            | OLK      | 1.10             | 0.003    | 0.14             |
| FOM/VT       | <i>Streptococcus</i> sp. HMT064 <i>australis</i> <i>oralis</i> subsp. <i>dentisani</i> clade 398 sp. HMT423 <i>infantis</i> clade 431 <i>infantis</i> clade 638 <i>mitis pneumoniae</i> | OLK      | -1.59            | 1.62E-05 | 0.002            |
| DT/LBT       | <i>Streptococcus</i> sp. HMT066 sp. HMT074                                                                                                                                              | OLK      | -2.47            | 0.0006   | 0.09             |
| DT/LBT       | <i>Veillonella rogosae dispar</i>                                                                                                                                                       | OLK      | -1.61            | 0.0004   | 0.09             |
| DT/LBT       | <i>Streptococcus thermophilus salivarius</i>                                                                                                                                            | OLK      | -2.97            | 0.0005   | 0.09             |

### 7.3.10 Microbiome associated with OLK adjusted for age and gender

Age and gender were not found to influence the microbiome of OLK with the same five taxa, *Bergeyella* HMT322, *Leptotrichia* HMT215, *Streptococcus australis infantis* clade 431, *Gemella morbillorum*, and *Leptotrichia* HMT392, identified as associated with OLK (Table 46) continuing to show increased abundance associated with OLK after adjustment for age and gender (all  $p < 0.02$ ,  $p_{adj} < 0.1$ ).

### 7.3.11 Microbiome associated with OLK adjusted for clinical appearance

When the clinical appearance of the OLK, homogenous or non-homogenous (NHOLK), was considered, *Bergeyella* sp. HMT322 and *Capnocytophaga gingivalis* were significantly associated with NHOLK ( $p=0.0001$ ,  $padj=0.03$  and  $p=0.006$ ,  $padj=0.05$  respectively).

### 7.3.12 Microbiome associated with OLK adjusted for multiple measures of oral hygiene

A MaAsLin2 model incorporating OLK and multiple measures of oral hygiene and health (brushing habit, oral hygiene score, denture wear, mouthwash use and number of missing teeth) demonstrated that three of the previously identified OLK-associated taxa, *Bergeyella* sp. HMT322, *Leptotrichia* sp. HMT215 and *Streptococcus australis infantis* clade 431, remained associated with OLK, although only the adjusted  $p$  value for *Streptococcus australis infantis* clade 431 was significant ( $padj=0.098$ ). *Streptococcus australis infantis* clade 431 also remained significantly associated with fair oral hygiene on this model (OHI-S of 0.5-1,  $padj=0.03$ ) but not with poor oral hygiene ( $padj>0.1$ ), as previously found on the individual oral hygiene analysis. No other OLK-associated species was associated with any oral hygiene measure (**Figure 44**).

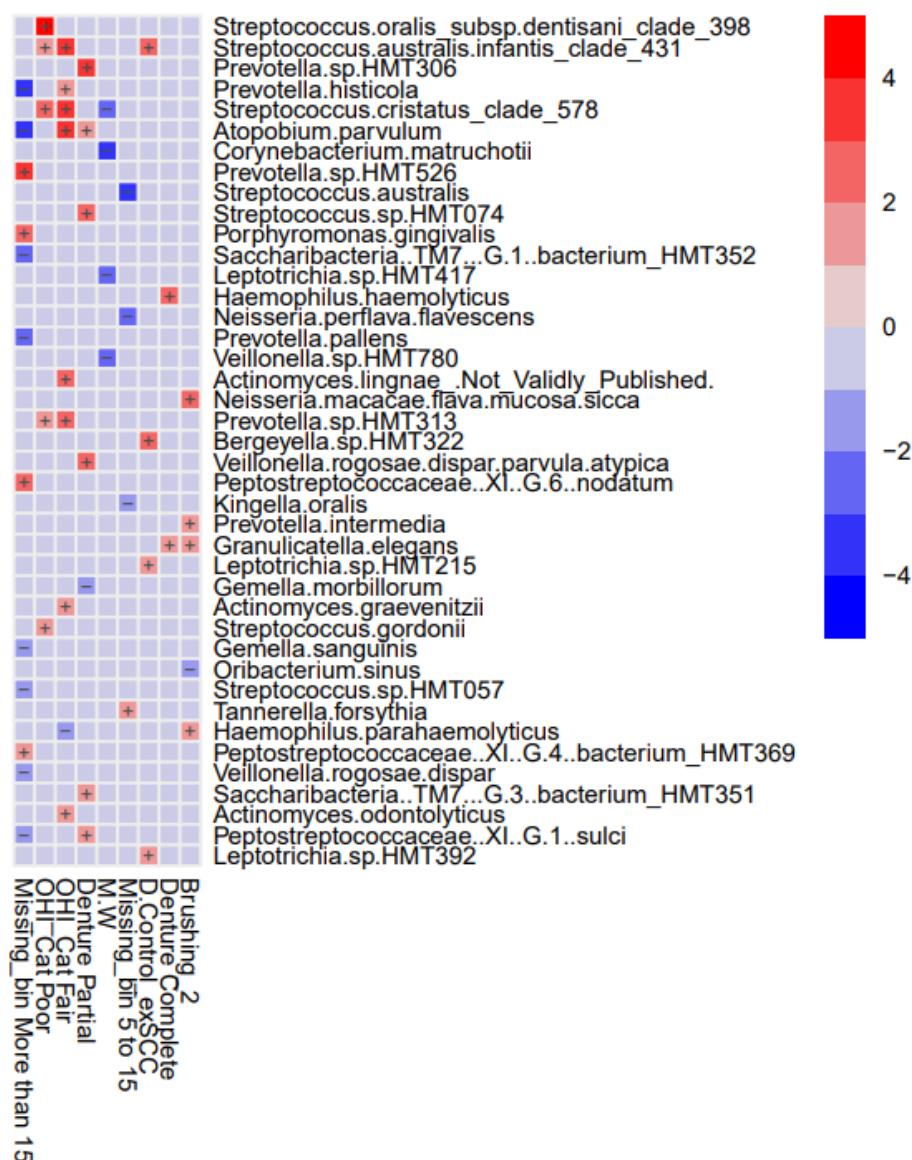


**Figure 44.** Heatmap showing the top 50 taxa with significant changes in abundance associated with OLK and brushing habit, oral hygiene category, denture wear, mouthwash use and number of missing teeth. Effect size calculated as  $-\log(q\text{-value}) \times (\text{coefficient})$ . (D.Control\_exSCC=OLK excluding OSCC samples [n=10], M.W=mouthwash use, Missing\_bin=number of missing teeth, Brushing\_2=brushing once or less/day, OHI\_Cat=oral hygiene category. Red=increased, blue=reduced)

### 7.3.13 Microbiome associated with OLK adjusted for site and smoking

Finally, as mucosal site and smoking were shown to be the most significant influencers of the oral microbiome, a MaAsLin2 model comprising OLK, mucosal site and smoking was constructed. All ten previously identified OLK-associated taxa remained significantly



associated with OLK after adjusting for oral site and smoking habit (**Table 57**), confirming that the microbial changes seen associated with OLK are independent of mucosal site and smoking.

**Table 57.** Results of MaAsLin2 analysis showing taxa significantly associated with OLK after adjusting for mucosal site and smoking. (Only OLK results shown)

| Taxa                                                                      | Variable | Coefficient | p value | Adjusted p value |
|---------------------------------------------------------------------------|----------|-------------|---------|------------------|
| <i>Bergeyella</i> sp. HMT322                                              | OLK      | 0.77        | 0.001   | 0.02             |
| <i>Streptococcus australis infantis</i> clade 431                         | OLK      | 1.17        | 0.001   | 0.02             |
| <i>Leptotrichia</i> sp. HMT215                                            | OLK      | 0.90        | 0.001   | 0.02             |
| <i>Gemella morbillorum</i>                                                | OLK      | 0.90        | 0.002   | 0.02             |
| <i>Lachnoanaerobaculum umeaense</i>                                       | OLK      | 0.64        | 0.003   | 0.03             |
| <i>Porphyromonas</i> sp. HMT284                                           | OLK      | 0.59        | 0.004   | 0.03             |
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum</i> <i>hwasookii</i> | OLK      | 0.58        | 0.006   | 0.05             |
| <i>Haemophilus parainfluenzae</i>                                         | OLK      | 1.05        | 0.007   | 0.05             |
| <i>Prevotella</i> sp. HMT396                                              | OLK      | 0.55        | 0.010   | 0.06             |
| <i>Leptotrichia</i> sp. HMT392                                            | OLK      | 0.40        | 0.01    | 0.07             |

## 7.4 Discussion

As discussed in Chapter 2 (**section 2.6**), there are few studies to date on the microbiome of OLK/OPMD<sup>61,79,81,93,94,96,98–100,103,152–154</sup> (**Table 4**). Further, many of these studies are limited by disparate study populations, the use of non-standard terminology to describe the included white patches and a lack of histological detail such that conditions not considered to be OPMDs, such a frictional keratosis, could have been included. As mentioned, only those conditions listed as OPMDs by Warnakulasuriya *et al.*<sup>116</sup> should have been included in these studies. However, there is agreement that firstly the microbiome of OLK/OPMD differs from that of healthy mucosa, and secondly that the difference is typically marked by increased abundance of anaerobic or facultatively anaerobic Gram-negative organisms of

phyla Fusobacteria and Bacteroidetes and reduced abundance of normal colonisers, such as *Streptococcus* spp.

The studies using swab sampling of OLK/OPMD<sup>61,79,81,152</sup> found increased abundances of Fusobacterium, Actinobacteria, Bacteroidetes, and Proteobacteria and reduced abundance of Firmicutes associated with OLK/OPMD. Genera level analysis was generally consistent amongst these studies in showing reduced abundance of *Streptococcus* associated with OLK/OPMD compared with controls. There are only four biopsy-based OLK/OPMD microbiome studies reported to date<sup>93,152–154</sup>; two by the same research group, one which included white patches (noted as hyperplasia and hyperkeratosis) that may not be true OPMDs and another which did not provide a clinical description of the ‘dysplastic lesions’ and only found differences in unmatched samples suggesting the influence of confounding factors. However, there was a trend of increased abundance of Gram-negative anaerobes or facultative anaerobes, including species of genera *Fusobacterium*, *Tannerella*, *Porphyromonas*, *Prevotella* and *Campylobacter*, and reduced abundance of *Streptococcus* spp. associated with OLK/OPMD. Interestingly, one study showed increased abundance of two *Streptococcus* spp. associated with OLK<sup>93</sup>. Finally, the salivary studies<sup>94,97–100,103</sup>, despite diverse, and at times unclear, inclusion criteria, in general also showed increased abundance of Gram-negative anaerobes and facultative anaerobes such as those of phyla Bacteroidetes and Fusobacterium, and reduced abundance of Firmicutes associated with OLK/OPMD.

Unfortunately, with the exception of Lee *et al.*<sup>99</sup> all of these studies are small, with less than 40 participants. The most common sampling method was saliva, with seven studies sampling OLK/OPMDs directly via either swab (n=4) or biopsy (n=3) (**Table 4**), of which only two<sup>61,79</sup> used CLN samples as controls. One used histologically normal adjacent

tissue<sup>152</sup>, another, with questionable inclusion criteria, used ipsilateral tissue<sup>153</sup> and three used healthy controls<sup>81,93,154</sup>. Also, other than Amer *et al.*<sup>61</sup> none investigated or controlled for other factors such as smoking, which are known to influence the oral microbiome. This study is therefore the first to comprehensively examine the microbiome of well-characterised OLK, while also investigating and controlling for various potential confounders including smoking, alcohol consumption, brushing habit, oral hygiene, denture wear, mouthwash use, tooth number and mucosal site.

As in the studies above, I also found OLK to be associated with significant changes in the oral microbiome. In agreement with the swab-based studies of Schmidt *et al.*<sup>79</sup> and Amer *et al.*<sup>61</sup>, I found increased abundance of *Fusobacterium* and *Leptotrichia* associated with OLK. Furthermore, agreeing with the findings of Schmidt *et al.*<sup>79</sup>, Mok *et al.*<sup>81</sup> and Radaic *et al.*<sup>152</sup>, I found reduced abundance of *Streptococcus* associated with OLK (**Table 45**). It is difficult to draw further comparisons however, as most of these swab studies did not carry out analysis to species level as I did, although Amer *et al.*<sup>61</sup> did report increased abundance of *Rothia mucilaginosa* associated with OLK, a finding I also identified using DESeq2 analysis (**Table 47**).

My results demonstrated that OLK is consistently associated with increased abundance of five taxa, *Bergeyella* sp. HMT322, *Leptotrichia* sp. HMT215, *Streptococcus australis infantis* clade 431, *Gemella morbillorum*, and *Leptotrichia* sp. HMT392, even when accounting for multiple variables. Interestingly *Bergeyella* was not identified in any previous OPMD study. It is difficult to compare my results with those of the biopsy and saliva studies because, as mentioned previously, different methodologies often lead to varied results. However, the trend of a switch to a Gram-negative anaerobic or facultatively anaerobic microbiome was partially observed; *Bergeyella* being Gram-negative and aerobic,

*Leptotrichia* Gram-negative and anaerobic and *Gemella* Gram-positive and facultatively anaerobic. I also found one streptococcal taxon, *Streptococcus australis infantis* clade 431, increased in abundance associated with OLK. Only one other study to date has identified an increase in abundance of streptococci, *Streptococcus parasanguinis* and *Streptococcus salivarius*, in OLK<sup>93</sup>. The only notable exception to this grouping's association with OLK was after adjusting for oral hygiene score, when none of the adjusted p values were significant. The other exceptions to the associations of these taxa with OLK were (1) after adjusting for alcohol consumption, when *Streptococcus australis infantis* clade 431, *Leptotrichia* sp. HMT392 and *Gemella morbillorum* were not significantly associated with OLK, (2) after adjusting for missing teeth when *Gemella morbillorum* was not associated, and (3) after adjusting for brushing habit when *Leptotrichia* sp. HMT392 was not associated.

Five other taxa, *Porphyromonas* sp. HMT284, *Fusobacterium nucleatum* subsp. *polymorphum hwasookii*, *Lachnoanaerobaculum umeaense*, *Haemophilus parainfluenzae* and *Prevotella* sp. HMT396 were associated with OLK after adjusting for mucosal site, demonstrating that these taxa are associated with OLK after site colonisation differences are considered. *Porphyromonas* sp. HMT284, *Fusobacterium nucleatum* subsp. *polymorphum hwasookii*, *Fusobacterium periodonticum nucleatum* subsp. *polymorphum* sp. HMT204 *hwasookii* and *Lachnoanaerobaculum umeaense* were also associated with OLK after adjusting for smoking. Furthermore, *Fusobacterium periodonticum nucleatum* subsp. *polymorphum* sp. HMT204 *hwasookii* and *Porphyromonas* sp. HMT284 were also associated with OLK after adjusting for missing teeth. Finally, *Porphyromonas* sp. HMT284 was also associated with OLK after adjusting for denture wear (**Table 57**).

**Table 57.** Summary table showing taxa significantly associated with OLK alone and associated with OLK after adjusting for various potentially influencing variables on MaAslin2 analysis (X denotes significant association where  $p_{adj} < 0.1$ ) (m/wash=mouthwash)

| Model adjusted for (OLK + variable)                                                                | OLK | Smoking | Alcohol use | Brushing habit | Oral hygiene | Denture wear | M/wash use | Missing teeth | Mucosal site |
|----------------------------------------------------------------------------------------------------|-----|---------|-------------|----------------|--------------|--------------|------------|---------------|--------------|
| <b>OLK associated taxa</b>                                                                         |     |         |             |                |              |              |            |               |              |
| <i>Bergeyella</i> sp. HMT322                                                                       | X   | X       | X           | X              |              | X            | X          | X             | X            |
| <i>Leptotrichia</i> sp. HMT215                                                                     | X   | X       | X           | X              |              | X            | X          | X             | X            |
| <i>Streptococcus australis infantis</i> clade 431                                                  | X   | X       |             | X              |              | X            | X          | X             | X            |
| <i>Gemella morbillorum</i>                                                                         | X   | X       |             | X              |              | X            | X          |               | X            |
| <i>Leptotrichia</i> sp. HMT392                                                                     | X   | X       |             |                |              | X            | X          | X             | X            |
| <i>Porphyromonas</i> sp. HMT284                                                                    |     | X       |             |                |              | X            |            | X             | X            |
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum</i> <i>hwasookii</i>                          |     | X       |             |                |              |              |            |               | X            |
| <i>Lachnoanaerobaculum umeaense</i>                                                                |     | X       |             |                |              |              |            |               | X            |
| <i>Fusobacterium periodonticum nucleatum</i> subsp. <i>polymorphum</i> sp. HMT204 <i>hwasookii</i> |     | X       |             |                |              |              |            | X             |              |
| <i>Haemophilus parainfluenzae</i>                                                                  |     |         |             |                |              |              |            |               | X            |
| <i>Prevotella</i> sp. HMT396                                                                       |     |         |             |                |              |              |            |               | X            |

As mentioned, both MaAsLin2 and DESeq2 analyses were carried out for OLK versus CLN sites and for specific OLK site analyses. The results were broadly consistent with both methodologies showing significantly increased abundance of *Bergeyella* sp. HMT322 and *Streptococcus australis infantis* clade 431 associated with OLK, DESeq2 also showing *Streptococcus australis infantis* clade 431 significantly associated with Buc/Alv/Gin OLKs. Finally, *Prevotella* sp. HMT396 was identified as associated with OLK after adjusting for site on MaAslin2 and associated with Buc/Alv/Gin OLKs on DESeq2. The other three taxa consistently associated with OLK on MaAslin2 analysis, *Leptotrichia* sp. HMT215, *Leptotrichia* sp. HMT392 and *Gemella morbillorum* (Table 46), were not identified on

DESeq2. *Haemophilus parainfluenzae* and *Lachnoanaerobaculum umeaense* were also identified as associated with OLK on both MaAslin2, after adjusting for mucosal site and smoking respectively, and DESeq2 analysis, although the adjusted p values on DESeq2 were not significant. It is not surprising that there were some differences in the MaAsLin2 and DESeq2 analyses as, although both packages detect changes in relative abundances, DESeq2 measures species counts associated with one specific variable (OLK, specific mucosal site etc.) and cannot adjust for multiple variables, whereas MaAsLin2 determines multivariable association in a combined model.

The most noteworthy differences between the MaAsLin2 and DESeq2 analyses were in the significance of changes in abundances of *Rothia* species. DESeq2 analysis found significantly increased abundance of *Rothia mucilaginosa* associated with Buc/Alv/Gin and PAL OLKs and significantly increased abundance of *Rothia aeria dentocariosa* associated with Buc/Alv/Gin OLKs. MaAsLin2 also identified both *Rothia mucilaginosa* and *Rothia aeria dentocariosa* as associated with OLK, however the adjusted p values were not significant ( $p=0.09$ ,  $\text{padj}=0.25$ ,  $p=0.04$  and  $\text{padj}=0.23$  respectively, results not shown). Previous MaAslin2 analysis of CLN and healthy control sites (**Chapter 6, section 6.3.8**) identified multiple *Rothia* species significantly associated with DT/LBT sites, *Rothia mucilaginosa* significantly associated with PAL sites ( $p=0.01$ ,  $\text{padj}=0.08$ ) and *Rothia dentocariosa* as associated with Buc/Alv/Gin sites, ( $p=0.0005$ ,  $\text{padj}=0.007$ , results not shown) suggesting that the changes seen on DESeq2 may in part be site, rather than OLK, related changes.

## 7.5 Conclusion

In agreement with previous studies (**Table 4**), I have demonstrated that the microbiome of OLK differs significantly from contralateral normal sites, OLK in general being associated with increased abundance of organisms that are anaerobic or facultatively anaerobic, (with the exception of *Bergeyella* HMT322 which is aerobic) and Gram-negative (with the exception of *Gemella morbillorum*, *Streptococcus australis infantis* clade 431 and *Lachnoanaerobaculum umeaense*). I have also identified five taxa, *Bergeyella* sp. HMT322, *Leptotrichia* sp. HMT215, *Streptococcus australis infantis* clade 431, *Gemella morbillorum*, and *Leptotrichia* sp. HMT392, consistently associated with OLK after adjusting for multiple variables. Adjusting for oral site identified several other OLK-associated taxa, including two *Fusobacterium* species, *Porphyromonas* sp. HMT284, *Lachnoanaerobaculum umeaense*, *Haemophilus parainfluenzae*, *Prevotella* sp. HMT396 on MaAsLin2 and *Rothia mucilaginosa* on DESeq2. Further, adjusting for smoking identified the same two *Fusobacterium* species, *Porphyromonas* sp. HMT284, and *Lachnoanaerobaculum umeaense* as associated with OLK.

In summary with regards to the aim of this chapter and primary hypothesis of the study:

- The microbial colonisation of OLK has been shown to be different to that of contralateral normal sites and these changes are independent of factors such as smoking and oral site which were previously shown to influence mucosal colonisation.

## 8 The microbiome of low-risk and high-risk dysplastic OLKs and OSCC

### 8.1 Introduction

Following on from Chapter 7, having established a core microbiome of OLK which is significantly different to that of contralateral normal mucosa, the aim of this chapter was to investigate if there were microbiome changes specific to different degrees of dysplasia of OLK which could be used to discriminate sites at higher or lower risk of malignant transformation.

### 8.2 Methods

As previously discussed, the diagnosis of OLK is confirmed by histological examination, with the degree of dysplasia, if present, graded according to the 3-tier WHO 2017 classification (mild, moderate, severe dysplasia)<sup>114</sup>. However, for the purposes of the microbiome analysis, the degree of dysplasia was re-classified in a binary manner (low-risk=no/mild dysplasia and high-risk=moderate/severe dysplasia)<sup>124,125,225</sup>, a binary approach being advocated for research studies<sup>122</sup>. Moderate/severe (high-risk) dysplasia has been shown to have a statistically greater risk of malignant transformation than mild dysplasia<sup>127</sup> and furthermore, my cohort included only eleven samples with no dysplasia and so it was preferable to combine these with the mild dysplasia samples as the low-risk group. Low-risk OLK (LROLK), high-risk OLK (HROLK), OSCC and contralateral normal (CLN) sites from the patient cohort were compared. A total of 91 LROLK, 125 HROLK and 10 OSCC sites from 187 patients were available for analysis (**Table 59**). The 10 OSCCs included seven pT1, (pT=pathological tumour stage, pT1=<2cm diameter and depth of invasion (DOI) <5mm), two pT2 (<2cm diameter and DOI 5-10 mm, or 2-4cm diameter and DOI



<10mm) and one pT4 tumour (moderately advanced local disease with invasion of adjacent local structures, in this case cortical bone) (**Figure 45**, **Figure 46**). All were lymph node negative (N0) on resection. Data analysis was as described in **section 4.4**. MaAsLin2 was used for multivariate analysis to investigate (i) if the microbial colonisation patterns of LROLK, HROLK and OSCC sites were different, and (ii) if smoking, alcohol consumption, brushing habit, oral hygiene, denture wear, mouthwash use, tooth number, mucosal site or clinical appearance influenced the colonisation patterns of LROLK, HROLK and OSCC. DESeq2 in R Studio was also used to analyse and compare the species level count data of LROLK, HROLK and OSCC sites as a second measure to ensure consistency of results.



**Figure 45.** pT1 OSCC left floor of mouth



**Figure 46.** pT2 OSCC left lateral tongue

**Table 60.** Final sample cohort sample sites and histological findings

| Sample data                                       | % OLK (n)  |
|---------------------------------------------------|------------|
| <b>Sites samples (n=226)</b>                      |            |
| Floor of mouth/ventral tongue                     | 28.3 (64)  |
| Buccal mucosa/Alveolus/Gingiva                    | 37.2 (84)  |
| Dorsum/lateral tongue                             | 23 (52)    |
| Palate                                            | 11.5 (26)  |
|                                                   |            |
| <b>Histological findings (n=226)</b>              |            |
| No dysplasia                                      | 4.9 (11)   |
| Mild dysplasia                                    | 35.4 (80)  |
| Moderate dysplasia                                | 33.6 (76)  |
| Severe dysplasia                                  | 21.7 (49)  |
| OSCC                                              | 4.4 (10)   |
|                                                   |            |
| <b>Binary categorisation of dysplasia (n=226)</b> |            |
| Low-risk OLK (No/mild dysplasia)                  | 40.3 (91)  |
| High-risk OLK (Moderate/Severe dysplasia)         | 55.3 (125) |
| OSCC                                              | 4.4 (10)   |

## 8.3 Results

### 8.3.1 Microbiome of low-risk and high-risk dysplasia and OSCC

The alpha diversities of CLN, LROLK, HROLK and OSCC were compared using ACE, Chao1, Observe, Simpson and Shannon indices. The only significant differences noted were between LROLK and HROLK on Shannon and Simpson indices, reflecting differences in relative abundances between these groups which may in part be site-related given that 84% of the HROLK were located on the tongue and floor of mouth. No other significant differences were noted, which is not surprising as these samples all came from the same patient population (**Figure 47, Table 61**).

**Table 61.** Table of alpha diversity measures from **Figure 41** (significant results in bold)

| Alpha diversity index | ACE  | Chao1 | Observe | Shannon      | Simpson     |
|-----------------------|------|-------|---------|--------------|-------------|
| Comparison            |      |       |         |              |             |
| CLN v LROLK           | 0.27 | 0.26  | 0.18    | 0.15         | 0.12        |
| CLN v HROLK           | 0.45 | 0.45  | 0.57    | 0.43         | 0.28        |
| CLN v OSCC            | 0.72 | 0.7   | 0.67    | 0.74         | 0.55        |
| LROLK v HROLK         | 0.67 | 0.59  | 0.41    | <b>0.039</b> | <b>0.02</b> |
| LROLK v OSCC          | 1    | 0.91  | 0.85    | 0.4          | 0.26        |
| HROLK v OSCC          | 0.96 | 0.91  | 0.94    | 0.87         | 0.95        |

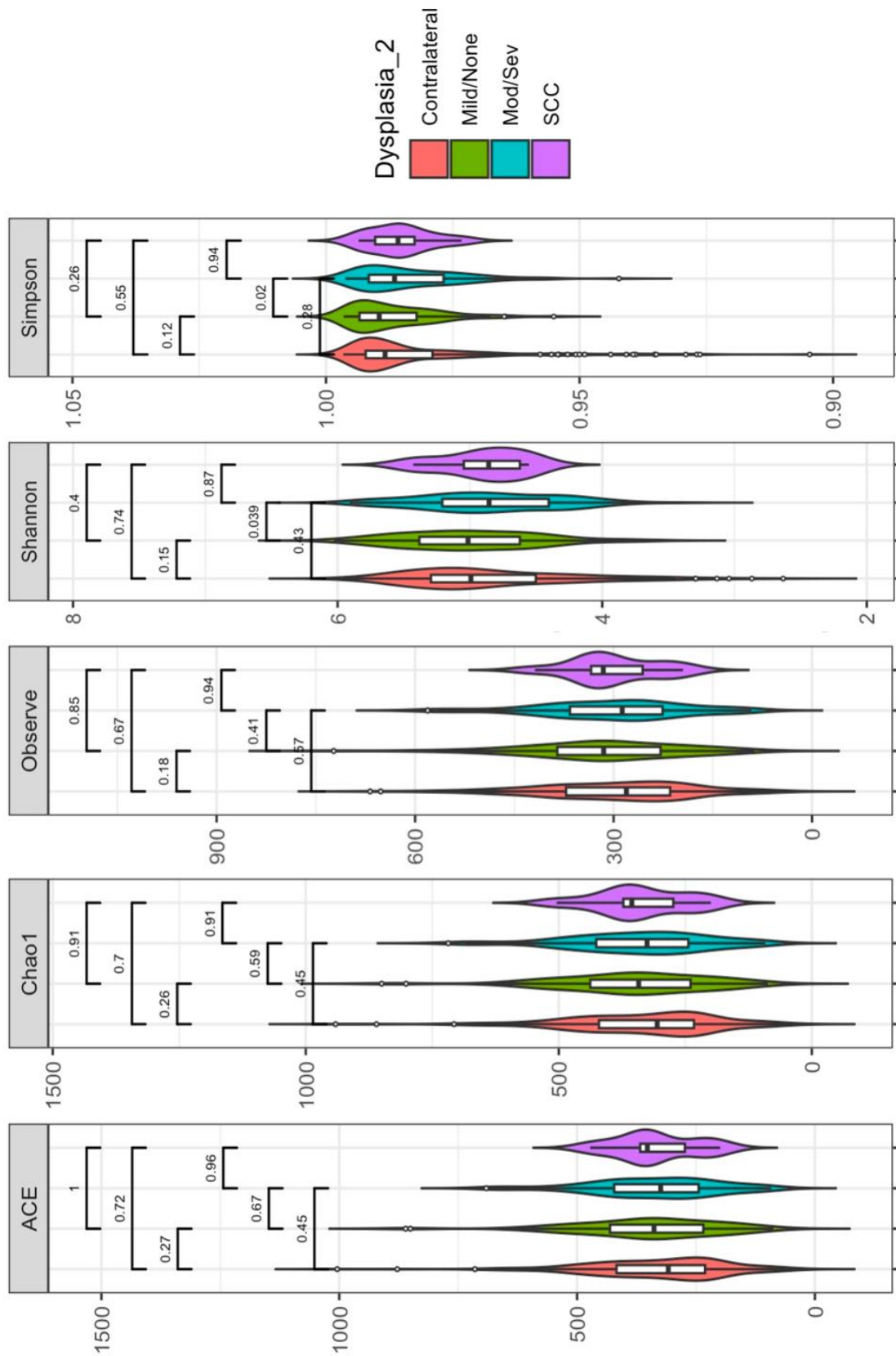


**Figure 47.** Alpha diversity plots of contralateral normal, none/mild dysplasia (LROLK), moderate/severe dysplasia (mod/sev [HROLK]) and OSCC (SCC) using ACE, Chao1, Observe, Shannon and Simpson indices. p values shown above square brackets

The microbiome composition of LROLK, HROLK and OSCC was compared by generating Bray-Curtis similarity values for each sample and assessing their relatedness using NMDS. (Figure 48, Figure 49). The plots demonstrated some shifts in community structure with PERMANOVA testing confirming a significant difference between the microbial communities of LROLK and HROLK ( $p_{adj}=0.004$ ). PERMANOVA testing comparing the communities of LROLK and HROLK with OSCC was not possible due to the small number of OSCC ( $n=10$ ) in the cohort.

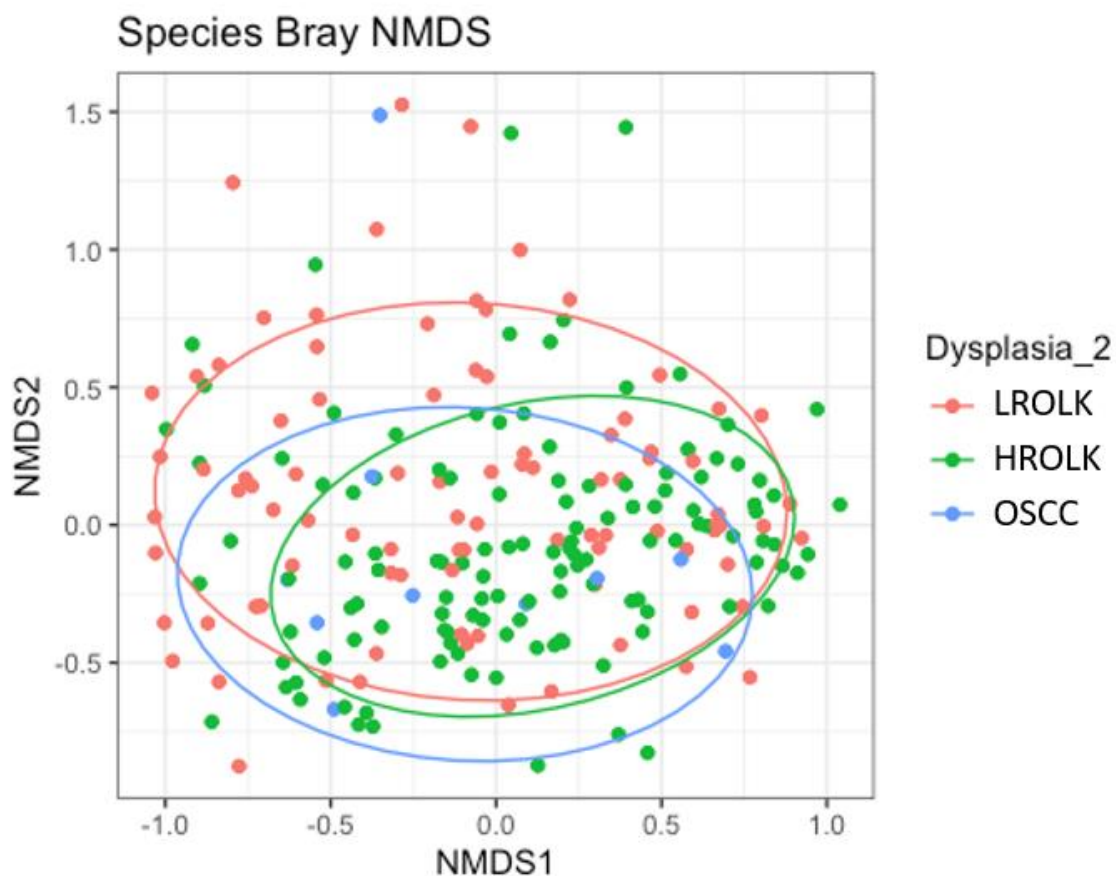


**Figure 48.** Axis 1 and 2 NMDS plot of LROLK, HROLK, and OSCC sites showing separation of the communities of LROLK, HROLK and OSCC.

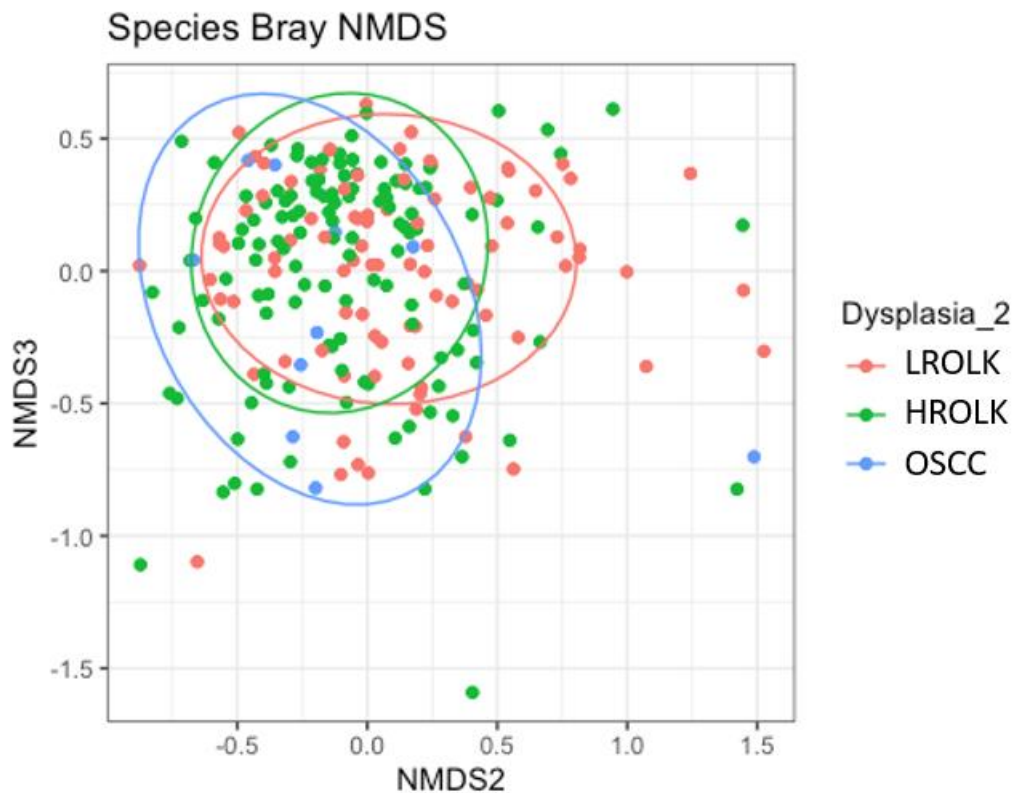


**Figure 49.** Axis 2 and 3 NMDS plot of LROLK, HROLK and OSCC sites showing separation of the communities of LROLK, HROLK and OSCC.

Qualitative analysis, using bar plots to visualise the most abundant phyla and genera associated with CLN, LROLK, HROLK and OSCC, clearly demonstrated these differences in colonisation patterns (**Figure 50, Figure 51**). At phyla level, all sites were dominated by Firmicutes, however there was a clear loss of Firmicutes associated with OSCC. Abundances of Actinobacteria did not appear to differ significantly between the groups, however there was a progressive increase in Proteobacteria from CLN to LROLK to HROLK to OSCC. Interestingly, similar abundances of Bacteroidetes and Fusobacteria were found in LROLK and OSCC, but not in HROLK where abundances of these phyla were more similar to CLN sites (**Figure 50**).



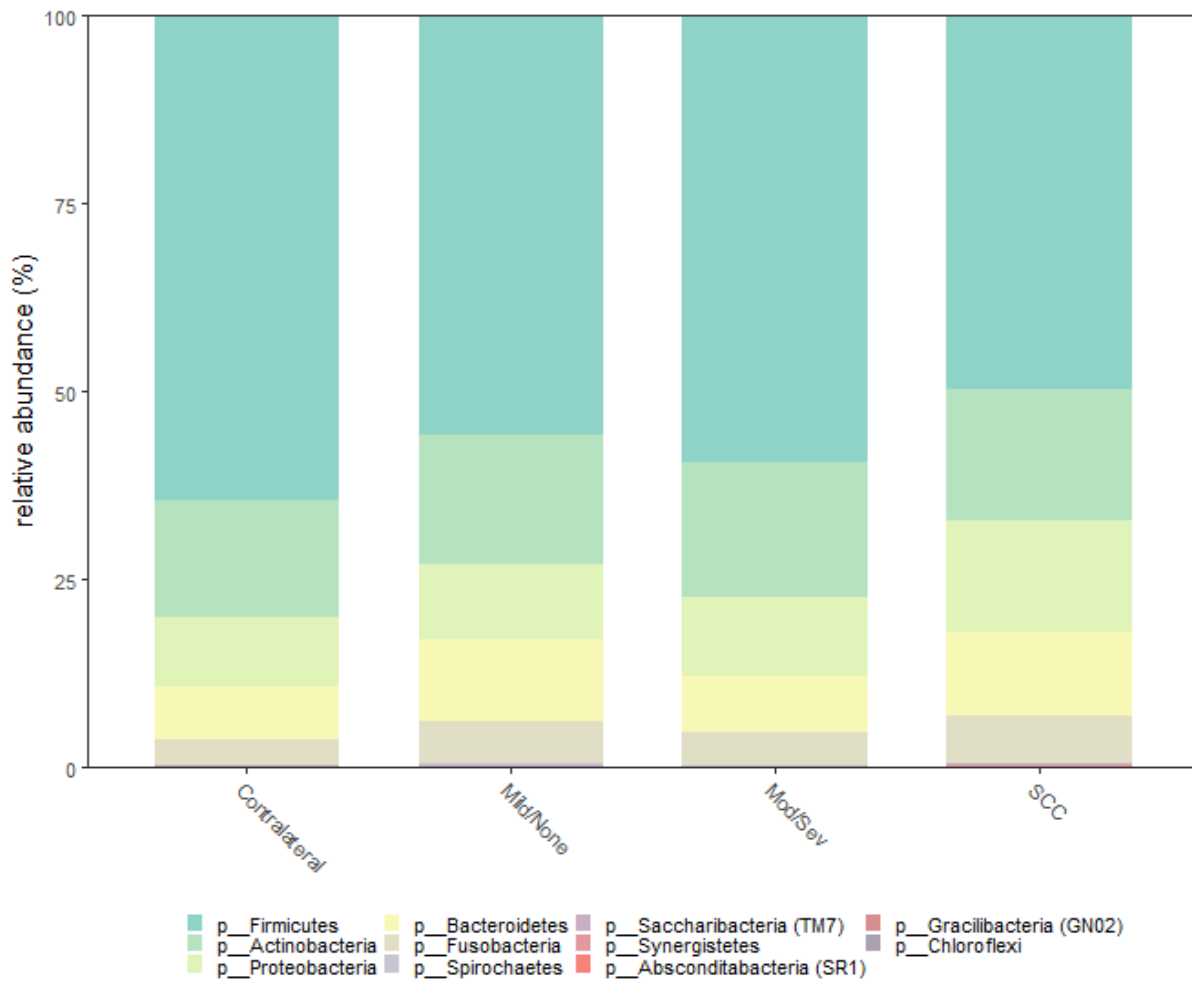


**Figure 50.** Barplot showing the most abundant phyla associated with CLN, LROLK (Mild/None), HROLK (Mod/Sev) and OSCC (SCC).

At the genera level, all sites were dominated by streptococci, with LROLK, HROLK and OSCC sites all showing reduced abundance of streptococci compared with CLN sites. *Rothia*, *Haemophilus* and *Fusobacterium* showed increased abundance at LROLK, HROLK and OSCC sites compared with CLN sites while *Neisseria*, and *Porphyromonas* were markedly increased in OSCC (**Figure 51**). Similar abundances of *Fusobacterium* and *Prevotella* were seen in LROLK and OSCC in keeping with the findings of the phyla level analysis.

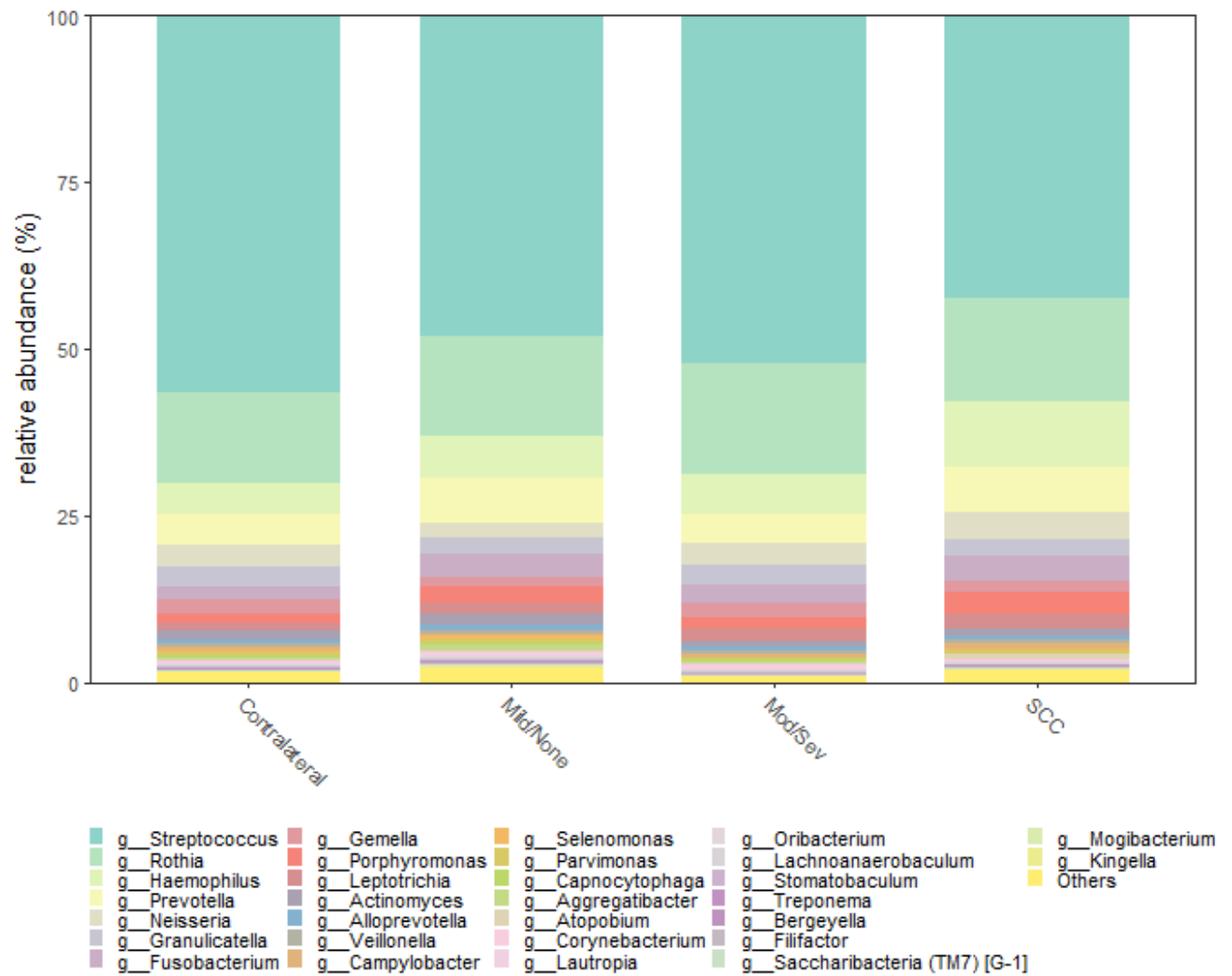


**Figure 51.** Barplot showing the most abundant genera associated with CLN, LROLK (Mild/None), HROLK (Mod/Sev) and OSCC (SCC).

MaAsLin2 was then used for multivariate analysis to compare the microbial colonisation patterns of LROLK, HROLK and OSCC with CLN sites at both genera and species levels. Nine genera showed significantly increased abundance associated with LROLK including the periodontal pathogens *Fusobacterium*, *Prevotella* and *Tannerella*, while streptococci were significantly reduced. HROLKs showed significantly increased abundance of *Bergeyella*, *Leptotrichia* and *Haemophilus* and reduced abundance of *Actinomyces*. OSCC was associated with increased abundance of *Haemophilus* and reduced abundance of *Streptococcus*, although the adjusted p values were not significant (Table 62).



**Table 62.** Genera with significant changes in abundance associated with LROLK, HROLK and OSCC compared with CLN sites on MaAsLin2 analysis (negative coefficient values denotes reduced abundance).

| Genera                             | Variable | Coefficient | p value | Adjusted p value |
|------------------------------------|----------|-------------|---------|------------------|
| <i>Fusobacterium</i>               | LROLK    | 1.23        | 0.0006  | 0.04             |
| <i>Mogibacterium</i>               | LROLK    | 1.45        | 0.0005  | 0.04             |
| <i>Prevotella</i>                  | LROLK    | 1.05        | 0.004   | 0.08             |
| <i>Campylobacter</i>               | LROLK    | 1.03        | 0.004   | 0.08             |
| <i>Tannerella</i>                  | LROLK    | 1.06        | 0.003   | 0.08             |
| <i>Bacteroidales G2</i>            | LROLK    | 0.86        | 0.002   | 0.08             |
| <i>Lachnospiraceae G3</i>          | LROLK    | 0.77        | 0.002   | 0.08             |
| <i>Aggregatibacter</i>             | LROLK    | 1.49        | 0.005   | 0.08             |
| <i>Peptostreptococcus</i>          | LROLK    | 0.98        | 0.006   | 0.08             |
| <i>Peptostreptococcaceae XI G1</i> | LROLK    | 0.61        | 0.005   | 0.08             |
| <i>Streptococcus</i>               | LROLK    | -0.27       | 0.006   | 0.08             |
| <i>Bergeyella</i>                  | HROLK    | 1.18        | 0.000   | 0.04             |
| <i>Leptotrichia</i>                | HROLK    | 1.01        | 0.004   | 0.08             |
| <i>Haemophilus</i>                 | HROLK    | 1.10        | 0.006   | 0.08             |
| <i>Actinomyces</i>                 | HROLK    | -1.02       | 0.003   | 0.08             |
| <i>Haemophilus</i>                 | OSCC     | 2.34        | 0.04    | 0.21             |
| <i>Streptococcus</i>               | OSCC     | -0.57       | 0.02    | 0.15             |

At the species level, LROLK was associated with significantly increased abundance of *Gemella morbillorum* and *Mogibacterium diversum*, while HROLK was associated with significantly increased abundance of *Bergeyella* sp. HMT322, *Streptococcus australis infantis* clade 431 and *Leptotrichia* sp. HMT215. Four of these species, *Gemella morbillorum*, *Bergeyella* sp. HMT322, *Streptococcus australis infantis* clade 431 and *Leptotrichia* sp. HMT215 were previously found to be associated with OLK (**Chapter 7, section 7.3.1**). Two species, *Fusobacterium nucleatum* subsp. *polymorphum hwasookii* (also previously found to be OLK-associated) and *Parvimonas* sp. HMT215, were associated with OSCC, although the adjusted p values were not significant (**Table 63**).

**Table 63.** Taxa with significant changes in abundance associated with LROLK, HROLK and OSCC compared with CLN sites on MaAsLin2 analysis.

| Taxa                                                               | Variable | Coefficient | p value  | Adjusted p value |
|--------------------------------------------------------------------|----------|-------------|----------|------------------|
| <i>Gemella morbillorum</i>                                         | LROLK    | 0.57        | 1.55E-07 | 8.63E-05         |
| <i>Mogibacterium diversum</i>                                      | LROLK    | 0.41        | 0.0002   | 0.03             |
| <i>Bergeyella</i> sp. HMT322                                       | HROLK    | 0.36        | 1.19E-05 | 0.003            |
| <i>Streptococcus australis infantis</i> clade 431                  | HROLK    | 0.48        | 0.0001   | 0.02             |
| <i>Leptotrichia</i> sp. HMT215                                     | HROLK    | 0.37        | 0.0002   | 0.02             |
| <i>Parvimonas</i> sp. HMT393                                       | OSCC     | 0.71        | 0.002    | 0.12             |
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum</i> hwasookii | OSCC     | 0.61        | 0.003    | 0.12             |

Univariate analysis comparing LROLK with CLN, HROLK with CLN, HROLK with LROLK and OSCC with CLN, LROLK and HROLK was also carried out with a second tool, DESeq2, as previously described (**Chapter 7**). The species associated with LROLK on MaAsLin2 were not associated with LROLK on the DESeq2 analysis comparing LROLK and CLN. Five taxa showed significantly reduced abundance in LROLK on this DESeq2 analysis. However, three of these five taxa showed reduced abundance associated with either site or smoking status, not LROLK, on MaAsLin2 suggesting that the abundance changes seen on DESeq2 were in fact due to either site or smoking status and not LROLK (**Table 64**).

**Table 64.** Results of DESeq2 analysis comparing LROLK and CLN sites showing taxa with significant changes in abundance in LROLK relative to CLN sites (negative log2 fold change values denote reduced abundance).

| Taxa                                    | log2 fold change | p value  | Adjusted p value | Changes in abundance on MaAsLin2 (*= $\text{padj} < 0.1$ )     |
|-----------------------------------------|------------------|----------|------------------|----------------------------------------------------------------|
| <i>Streptococcus oralis infantis</i>    | -1.25            | 4.61E-07 | 0.0003           | NA                                                             |
| <i>Streptococcus</i> sp. HMT423 mitis   | -1.66            | 8.09E-06 | 0.002            | Reduced LROLK, DT/LBT<br>Increased current* and former smokers |
| <i>Gemella haemolysans</i>              | -1.62            | 9.33E-05 | 0.02             | Reduced DT/LBT*, FOM/VT*                                       |
| <i>Neisseria perflava flavescens</i>    | -2.02            | 0.0005   | 0.06             | Reduced current* and former smokers*                           |
| <i>Streptococcus infantis</i> clade 638 | -1.97            | 0.0005   | 0.06             | Increased DT/LBT*                                              |

\*Significant changes in abundance identified on MaAsLin2 analysis also shown. (DT/LBT=dorsum=lateral border tongue, FOM/VT=floor of mouth/ventral tongue, NA=not applicable).

The DESeq2 comparison of HROLK versus CLN sites identified two of the three taxa previously identified as significantly associated with HROLK on MaAsLin2 analysis (**Table 65**); *Bergeyella* sp. HMT322 ( $\text{padj}=0.096$ ) and *Streptococcus australis infantis* clade 431 ( $\text{padj}=1.38\text{E}-07$ ) demonstrated significantly increased abundance associated with HROLK on DESeq2. Several other taxa also showed increased abundance associated with HROLK, although only the adjusted p value for *Streptococcus infantis* clade 431 was significant ( $\text{padj}=0.096$ ). *Saccharibacteria* TM7 G1 *bacterium* HMT349 and *Selenomonas noxia* demonstrated significantly reduced abundance associated with HROLK, a change also seen on MaAsLin2 analysis, although the adjusted p values on the MaAsLin2 analysis were not significant ( $\text{padj}=0.11$  and  $0.14$  respectively). Separate MaAsLin2 analyses identified *Saccharibacteria* TM7 G1 *bacterium* HMT349 as showing significantly increased abundance associated with both LROLK after adjusting for smoking, (**Table 68**) and palatal site (**Table 37**) and *Selenomonas noxia* as showing reduced abundance associated with DT/LBT site ( $p=0.06$ ,  $\text{padj}=0.2$ , data not shown).

**Table 65.** Results of DESeq2 analysis comparing HROLK and CLN sites showing taxa with significant changes in abundance in HROLK relative to CLN sites. (Yellow=increased in HROLK, green=decreased in HROLK)

| Taxa                                                                   | log2 fold change | p value  | Adjusted p value | Changes in abundance on MaAsLin2 (*= $\text{padj} < 0.1$ )        |
|------------------------------------------------------------------------|------------------|----------|------------------|-------------------------------------------------------------------|
| <i>Leptotrichia</i> sp. HMT225                                         | 3.5              | 0.009    | 0.15             | NA                                                                |
| <i>Streptococcus australis infantis</i> clade 431                      | 3.16             | 7.45E-10 | 1.38E-07         | Increased HROLK*                                                  |
| <i>Lachnoanaerobaculum umeaense</i>                                    | 1.79             | 0.02     | 0.24             | Increased LROLK* and HROLK<br>Reduced current* and former smokers |
| <i>Streptococcus infantis</i> clade 431                                | 1.69             | 0.002    | 0.096            | Increased HROLK                                                   |
| <i>Leptotrichia</i> sp. HMT417                                         | 1.57             | 0.005    | 0.14             | Increased PAL<br>Reduced DT/LBT, FOM/VT*                          |
| <i>Fusobacterium periodonticum nucleatum</i> subsp. <i>polymorphum</i> | 1.33             | 0.004    | 0.13             | NA                                                                |
| <i>Bergeyella</i> sp. HMT322                                           | 1.28             | 0.002    | 0.096            | Increased HROLK*                                                  |
| <i>Campylobacter concisus</i>                                          | 0.97             | 0.006    | 0.14             | Reduced FOM/VT<br>Increased OSCC                                  |
| <i>Streptococcus oralis infantis</i>                                   | -0.6             | 0.007    | 0.14             | NA                                                                |
| <i>Streptococcus</i> sp. HMT423 <i>mitis</i>                           | -0.89            | 0.008    | 0.15             | Reduced LROLK, DT/LBT<br>Increased current* and former smokers    |
| <i>Selenomonas noxia</i>                                               | -2.04            | 0.003    | 0.096            | Reduced HROLK, DT/LBT*, PAL                                       |
| <i>Saccharibacteria</i> TM7 G1 <i>bacterium</i> HMT349                 | -2.84            | 0.0005   | 0.04             | Reduced HROLK, DT/LBT, PAL*<br>Increased LROLK*                   |

\*Significant changes in abundance identified on MaAsLin2 analysis also shown. (DT/LBT=dorsum tongue/lateral border tongue, FOM/VT=floor of mouth/ventral tongue, PAL=palate. Negative log2 fold change denotes reduced abundance, NA=not applicable)

Comparing HROLK and LROLK, eight taxa showed abundance changes in common with both DESeq2 and MaAsLin2, demonstrating the consistency of the findings, although only *Streptococcus australis infantis* clade 431 was significantly associated with HROLK in both methods after adjustment ( $\text{padj}=0.06$  and  $0.08$  respectively) (**Table 66**).

**Table 66.** Results of both MaAsLin2 and DESeq2 analysis of HROLK versus LROLK showing taxa with significant changes in abundance in HROLK relative to LROLK. (Yellow=increased in HROLK, green=decreased in HROLK) NA=not detected on analysis

| Taxa                                                                                                                                                                                                | Abundance change MaAsLin2 | p value MaAsLin2 | padj MaAsLin2 | Abundance change DESeq | p value DESeq | padj DESeq |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|------------------|---------------|------------------------|---------------|------------|
| <i>Bergeyella</i> sp. HMT322                                                                                                                                                                        | Increased                 | 0.05             | 0.13          | NA                     | NA            | NA         |
| <i>Capnocytophaga gingivalis</i>                                                                                                                                                                    | Increased                 | 0.10             | 0.22          | NA                     | NA            | NA         |
| <i>Gemella haemolysans</i>                                                                                                                                                                          | Increased                 | 0.11             | 0.23          | Increased              | 0.04          | 0.22       |
| <i>Granulicatella elegans</i>                                                                                                                                                                       | Increased                 | 0.05             | 0.13          | NA                     | NA            | NA         |
| <i>Leptotrichia</i> sp. HMT215                                                                                                                                                                      | Increased                 | 0.12             | 0.25          | NA                     | NA            | NA         |
| <i>Neisseria perflava flavescens</i>                                                                                                                                                                | Increased                 | 0.08             | 0.19          | Increased              | 0.03          | 0.21       |
| <i>Prevotella nanceiensis</i>                                                                                                                                                                       | Increased                 | 0.06             | 0.15          | NA                     | NA            | NA         |
| <i>Rothia aeria dentocariosa</i>                                                                                                                                                                    | Increased                 | 0.07             | 0.17          | NA                     | NA            | NA         |
| <i>Streptococcus australis infantis</i> clade 431                                                                                                                                                   | Increased                 | 0.03             | 0.08          | Increased              | 0.003         | 0.06       |
| <i>Streptococcus infantis</i> clade 431<br><i>infantis</i> clade 638                                                                                                                                | Increased                 | 0.009            | 0.04          | NA                     | NA            | NA         |
| <i>Streptococcus</i> sp. HMT064<br><i>australis oralis</i> subsp. <i>dentisani</i><br>clade 398 sp. HMT423 <i>infantis</i><br>clade 431 <i>infantis</i> clade 638 <i>mitis</i><br><i>pneumoniae</i> | Increased                 | 0.008            | 0.04          | Increased              | 0.02          | 0.15       |
| <i>Streptococcus</i> sp. HMT064 <i>mitis</i>                                                                                                                                                        | Increased                 | 0.05             | 0.13          | Increased              | 0.05          | 0.22       |
| <i>Streptococcus</i> sp. HMT064 sp.<br>HMT423                                                                                                                                                       | Increased                 | 0.11             | 0.22          | NA                     | NA            |            |
| <i>Gemella sanguinis</i>                                                                                                                                                                            | NA                        | NA               | NA            | Increased              | 0.01          | 0.15       |
| <i>Streptococcus infantis</i> clade 638                                                                                                                                                             | NA                        | NA               | NA            | Increased              | 0.003         | 0.06       |
| <i>Streptococcus mitis pneumoniae</i>                                                                                                                                                               | NA                        | NA               | NA            | Increased              | 0.002         | 0.06       |
| <i>Streptococcus</i> sp. HMT423 <i>mitis</i>                                                                                                                                                        | NA                        | NA               | NA            | Increased              | 0.06          | 0.24       |
| <i>Parvimonas micra</i>                                                                                                                                                                             | Reduced                   | 0.0003           | 0.005         | Reduced                | 0.04          | 0.22       |
| <i>Porphyromonas endodontalis</i>                                                                                                                                                                   | Reduced                   | 4.81E-05         | 0.002         | Reduced                | 0.01          | 0.10       |
| <i>Streptococcus vestibularis</i><br><i>salivarius</i>                                                                                                                                              | Reduced                   | 0.14             | 0.03          | Reduced                | 0.02          | 0.15       |

Finally, DESeq2 analyses comparing OSCC with CLN, LROLK and HROLK showed OSCC was associated with a loss of taxa, particularly of two *Haemophilus* spp. and *Fusobacterium* sp. HMT203 when compared with CLN and LROLK, along with a *Prevotella* and *Streptococcus* taxon when compared with HROLK, although the adjusted p values for these latter two taxa were not significant (**Table 67**). There were no similarities with the MaAsLin2 analysis of OSCC, likely reflecting the low numbers of OSCCs and the differences in methodology.

**Table 67.** Results of DESeq2 analysis comparing OSCC with CLN, LROLK and HROLK

| Taxa                                                                           | log2 fold change | p value  | Adjusted p value |
|--------------------------------------------------------------------------------|------------------|----------|------------------|
| <b>OCSS v CLN</b>                                                              |                  |          |                  |
| <i>Haemophilus haemolyticus</i>                                                | -22.57           | 4.91E-21 | 2.81E-18         |
| <i>Haemophilus</i> sp. HMT036                                                  | -22.21           | 7.70E-19 | 2.21E-16         |
| <i>Fusobacterium</i> sp. HMT203                                                | -21.66           | 2.65E-16 | 5.05E-14         |
| <b>OSCC v LROLK</b>                                                            |                  |          |                  |
| <i>Haemophilus haemolyticus</i>                                                | -21.84           | 7.53E-19 | 4.32E-16         |
| <i>Haemophilus</i> sp. HMT036                                                  | -21.42           | 8.30E-17 | 2.28E-14         |
| <i>Fusobacterium</i> sp. HMT203                                                | -22.43           | 3.64E-16 | 6.96E-14         |
| <b>OSCC v HROLK</b>                                                            |                  |          |                  |
| <i>Haemophilus haemolyticus</i>                                                | -22.62           | 1.45E-20 | 8.30E-18         |
| <i>Haemophilus</i> sp. HMT036                                                  | -21.59           | 1.92E-17 | 5.47E-15         |
| <i>Fusobacterium</i> sp. HMT203                                                | -21.11           | 3.35E-15 | 6.40E-13         |
| <i>Prevotella</i> sp. HMT303                                                   | -5.79            | 0.0002   | 0.21             |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 071 <i>sanguinis</i> | -4.59            | 0.0002   | 0.21             |

### 8.3.2 Microbiome associated with LROLK, HROLK, OSCC adjusted for smoking

MaAsLin2 was then used for multivariate analysis to investigate the influence of smoking on the colonisation patterns of LROLK, HROLK and OSCC. Adjusting for smoking resulted in more taxa being associated with LROLK and HROLK. LROLK was associated with significantly increased abundance of *Gemella morbillorum*, *Mogibacterium diversum*, *Peptostreptococcaceae* XI G1 *sulci*, *Porphyromonas* sp. HMT284, *Alloprevotella tannerae*, *Aggregatibacter* sp. HMT458, *Lachnospiraceae* G3 *bacterium* HMT100, *Actinomyces* sp. HMT169, *Bacteroidales* G2 *bacterium* HMT274, *Parvimonas* sp. HMT110, *Porphyromonas endodontalis*, *Peptostreptococcus stomatis*, *Corynebacterium durum*, *Streptococcus gordonii*, *Leptotrichia* sp. HMT392, *Prevotella oris* and *Saccharibacteria* TM7 G1 *bacterium* HMT349. HROLK was associated with significantly increased

abundance of *Bergeyella* sp. HMT322, *Streptococcus australis infantis* clade 431, *Leptotrichia* sp. HMT215 and significantly reduced abundance of *Streptococcus vestibularis salivarius*. Finally, OSCC remained associated with increased abundance of *Parvimonas* sp. HMT393 and *Fusobacterium nucleatum* subsp. *polymorphum hwasookii*, with this association now reaching the level of significance. Although many of these taxa were also associated with smoking, the five previously identified taxa associated with LROLK and HROLK (**Table 63**) were not associated with smoking, with *Gemella morbillorum* and *Bergeyella* sp. HMT322 in fact showing reduced abundance in current smokers, and all five taxa remaining associated with LROLK and HROLK after adjusting for smoking (all  $p < 0.01$ ,  $p_{adj} < 0.1$ ) (**Figure 52, Table 68**).

**Table 68.** Taxa with significant changes in abundance associated with LROLK, HROLK and OSCC after adjusting for smoking habit on MaAsLin2 analysis (smoking results not shown, negative coefficient values denote reduced abundance).

| Taxa                                                                      | Variable | Coefficient | p value  | Adjusted p value |
|---------------------------------------------------------------------------|----------|-------------|----------|------------------|
| <i>Gemella morbillorum</i>                                                | LROLK    | 0.59        | 6.03E-08 | 6.97E-06         |
| <i>Mogibacterium diversum</i>                                             | LROLK    | 0.41        | 0.0002   | 0.006            |
| <i>Aggregatibacter</i> sp. HMT458                                         | LROLK    | 0.28        | 0.001    | 0.02             |
| <i>Lachnospiraceae G 3 bacterium</i> HMT100                               | LROLK    | 0.24        | 0.001    | 0.02             |
| <i>Actinomyces</i> sp. HMT169                                             | LROLK    | 0.27        | 0.002    | 0.04             |
| <i>Bacteroidales G2 bacterium</i> HMT274                                  | LROLK    | 0.25        | 0.002    | 0.04             |
| <i>Peptostreptococcaceae XI G1 sulci</i>                                  | LROLK    | 0.17        | 0.002    | 0.04             |
| <i>Parvimonas</i> sp. HMT110                                              | LROLK    | 0.19        | 0.003    | 0.05             |
| <i>Porphyromonas endodontalis</i>                                         | LROLK    | 0.35        | 0.003    | 0.05             |
| <i>Porphyromonas</i> sp. HMT284                                           | LROLK    | 0.22        | 0.004    | 0.05             |
| <i>Peptostreptococcus stomatis</i>                                        | LROLK    | 0.31        | 0.005    | 0.06             |
| <i>Alloprevotella tannerae</i>                                            | LROLK    | 0.36        | 0.005    | 0.06             |
| <i>Corynebacterium durum</i>                                              | LROLK    | 0.20        | 0.006    | 0.07             |
| <i>Streptococcus gordonii</i>                                             | LROLK    | 0.44        | 0.006    | 0.07             |
| <i>Leptotrichia</i> sp. HMT392                                            | LROLK    | 0.17        | 0.007    | 0.08             |
| <i>Prevotella oris</i>                                                    | LROLK    | 0.35        | 0.008    | 0.08             |
| <i>Saccharibacteria TM7 G1 bacterium</i> HMT349                           | LROLK    | 0.18        | 0.01     | 0.098            |
| <i>Bergeyella</i> sp. HMT322                                              | HROLK    | 0.34        | 2.21E-05 | 0.001            |
| <i>Streptococcus australis infantis</i> clade 431                         | HROLK    | 0.48        | 0.0001   | 0.004            |
| <i>Leptotrichia</i> sp. HMT215                                            | HROLK    | 0.36        | 0.0003   | 0.007            |
| <i>Streptococcus vestibularis salivarius</i>                              | HROLK    | -0.31       | 0.005    | 0.06             |
| <i>Parvimonas</i> sp. HMT393                                              | OSCC     | 0.66        | 0.004    | 0.05             |
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum</i> <i>hwasookii</i> | OSCC     | 0.61        | 0.004    | 0.05             |



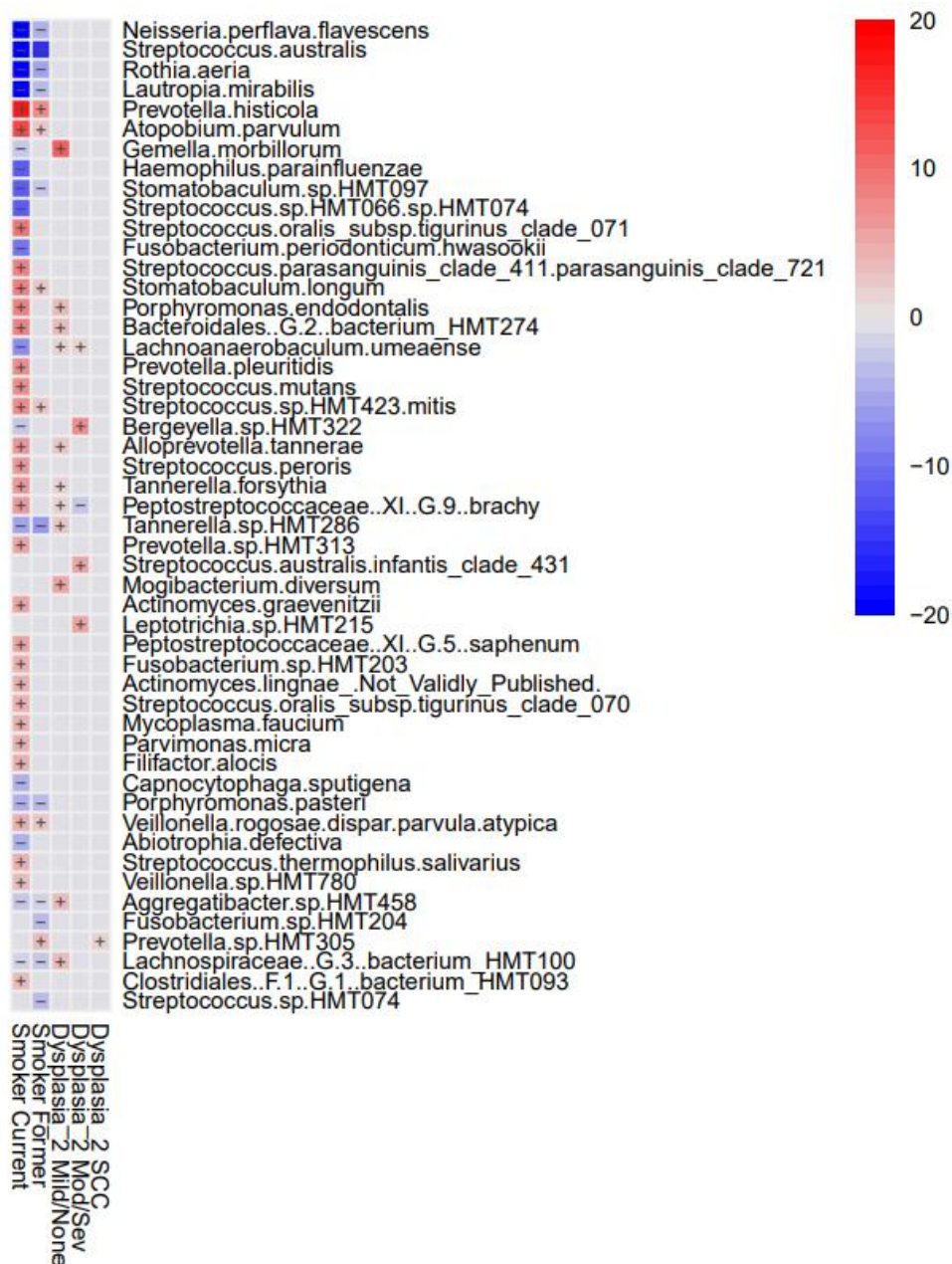


**Figure 52.** Heatmap showing taxa with significant changes in abundance associated with LROLK (Dysplasia\_2 Mild/None), HROLK (Dysplasia\_2 Mod/Sev), OSCC (SCC) and smoking status. Effect size calculated as  $-\log(q\text{-value}) \times (\text{coefficient})$  (red=increased, blue=reduced).

### 8.3.3 Microbiome associated with LROLK, HROLK, OSCC adjusted for alcohol

As for smoking, adjusting for alcohol consumption did not affect the previously identified LROLK and HROLK-associated taxa (**Table 69**); *Gemella morbillorum* and *Mogibacterium diversum* remained associated with LROLK, and *Bergeyella* HMT322, *Streptococcus australis infantis* clade 431 and *Leptotrichia* HMT215 remained associated with HROLK (all  $p < 0.001$  and  $p_{adj} < 0.05$ ) after alcohol consumption was considered. No species was significantly associated with OSCC after adjusting for alcohol consumption. Consuming more than 20 units of alcohol per week was associated with significantly increased abundance of *Leptotrichia* sp. HMT221, *Prevotella salivae*, *Saccharibacteria TM7 G1 bacterium* HMT352, *Megasphaera micronuciformis* and *Actinomyces odontolyticus* as previously described (**Chapter 7, section 7.3.2**).

**Table 69.** Taxa with significant changes in abundance associated with LROLK, HROLK and alcohol consumption on MaAsLin2 analysis.

| Taxa                                              | Variable       | Coefficient | p value  | Adjusted p value |
|---------------------------------------------------|----------------|-------------|----------|------------------|
| <i>Gemella morbillorum</i>                        | LROLK          | 0.57        | 2.43E-07 | 0.0001           |
| <i>Mogibacterium diversum</i>                     | LROLK          | 0.40        | 0.0004   | 0.04             |
| <i>Bergeyella</i> sp. HMT322                      | HROLK          | 0.35        | 1.80E-05 | 0.01             |
| <i>Streptococcus australis infantis</i> clade 431 | HROLK          | 0.48        | 0.0001   | 0.02             |
| <i>Leptotrichia</i> sp. HMT215                    | HROLK          | 0.37        | 0.0002   | 0.03             |
| <i>Leptotrichia</i> sp. HMT221                    | >20 units/week | 0.67        | 1.63E-07 | 0.0001           |
| <i>Prevotella salivae</i>                         | >20 units/week | 0.77        | 2.10E-05 | 0.01             |
| <i>Saccharibacteria TM7 G1 bacterium</i> HMT352   | >20 units/week | 0.41        | 2.40E-05 | 0.01             |
| <i>Megasphaera micronuciformis</i>                | >20 units/week | 0.41        | 0.0001   | 0.02             |
| <i>Actinomyces odontolyticus</i>                  | >20 units/week | 0.46        | 0.0004   | 0.04             |

### 8.3.4 Microbiome associated with LROLK, HROLK, OSCC adjusted for brushing habit

After adjusting for brushing habit, *Gemella morbillorum*, *Mogibacterium diversum*, and *Peptostreptococcaceae XI G1 sulci* showed significantly increased abundance associated with LROLK, while *Bergeyella* sp. HMT322, *Streptococcus australis infantis* clade 431 and *Leptotrichia* sp. HMT215 were significantly associated with HROLK. *Parvimonas* sp. HMT393 showed significantly increased abundance associated with OSCC after adjusting for brushing habit and was also the only OLK/OSCC-associated taxon influenced by brushing habit, being significantly increased in association with brushing once or less per day (**Table 70**).

**Table 70.** Taxa significantly associated with LROLK, HROLK and OSCC after adjusting for brushing habit (per day) on MaAsLin2 analysis (negative coefficient values denote reduced abundance).

| Taxa                                                                            | Variable           | Coefficient | p value  | Adjusted p value |
|---------------------------------------------------------------------------------|--------------------|-------------|----------|------------------|
| <i>Gemella morbillorum</i>                                                      | LROLK              | 0.6         | 8.57E-08 | 6.34E-05         |
| <i>Mogibacterium diversum</i>                                                   | LROLK              | 0.39        | 0.0008   | 0.08             |
| <i>Peptostreptococcaceae XI G1 sulci</i>                                        | LROLK              | 0.19        | 0.001    | 0.08             |
| <i>Bergeyella</i> sp. HMT322                                                    | HROLK              | 0.32        | 0.0001   | 0.05             |
| <i>Streptococcus australis infantis</i> clade 431                               | HROLK              | 0.48        | 0.0002   | 0.05             |
| <i>Leptotrichia</i> sp. HMT215                                                  | HROLK              | 0.34        | 0.0006   | 0.08             |
| <i>Parvimonas</i> sp. HMT393                                                    | OSCC               | 0.79        | 0.001    | 0.08             |
| <i>Gemella haemolysans</i>                                                      | Brushing once/less | 0.45        | 0.001    | 0.08             |
| <i>Parvimonas</i> sp. HMT393                                                    | Brushing once/less | 0.26        | 0.001    | 0.08             |
| <i>Streptococcus</i> sp. HMT064 <i>oralis</i> subsp. <i>dentisani</i> clade 398 | Brushing once/less | -0.22       | 0.001    | 0.08             |

### 8.3.5 Microbiome associated with LROLK, HROLK, OSCC adjusted for oral hygiene

The LROLK-associated species *Gemella morbillorum*, remained significantly associated with LROLK after adjusting for oral hygiene (padj=0.01). The other previously identified taxa associated with LROLK (*Mogibacterium diversum*) and HROLK (*Bergeyella* sp. HMT322, *Streptococcus australis infantis* clade 431 and *Leptotrichia* sp. HMT215) remained associated with LROLK and HROLK after adjusting for oral hygiene, but the adjusted p values did not reach significance ( $p < 0.05$  but padj  $> 0.1$ ) (**Table 71**). The only LROLK or HROLK-associated taxon significantly associated with oral hygiene was *Streptococcus australis infantis* clade 431, which showed significantly increased abundance associated with fair oral hygiene ( $p = 0.001$ , padj=0.06) but not poor oral hygiene, as identified previously (**Chapter 7, section 7.3.5**). Furthermore, it was the only taxon with a significant change in abundance associated with oral hygiene on this model.

**Table 71.** Taxa with significant changes in abundance associated with LROLK, HROLK and oral hygiene category on MaAsLin2 analysis.

| Taxa                                              | Variable          | Coefficient | p value  | Adjusted p value |
|---------------------------------------------------|-------------------|-------------|----------|------------------|
| <i>Gemella morbillorum</i>                        | LROLK             | 0.56        | 6.77E-06 | 0.01             |
| <i>Mogibacterium diversum</i>                     | LROLK             | 0.37        | 0.003    | 0.23             |
| <i>Streptococcus australis infantis</i> clade 431 | HROLK             | 0.50        | 0.001    | 0.11             |
| <i>Leptotrichia</i> sp. HMT215                    | HROLK             | 0.32        | 0.005    | 0.23             |
| <i>Bergeyella</i> sp. HMT322                      | HROLK             | 0.26        | 0.007    | 0.23             |
| <i>Streptococcus australis infantis</i> clade 431 | Fair oral hygiene | 0.56        | 0.0001   | 0.06             |

### 8.3.6 Microbiome associated with LROLK, HROLK, OSCC adjusted for denture wear

Wearing a denture did not influence the colonisation patterns of LROLK and HROLK. LROLK remained associated with increased abundance of *Gemella morbillorum* and *Mogibacterium diversum*, and HROLK with increased abundance of *Bergeyella* sp. HMT322, *Streptococcus australis infantis* clade 431 and *Leptotrichia* sp. HMT215 (all  $p_{adj} < 0.1$ ) when denture wear was considered. OSCC was associated with significantly increased abundance of *Parvimonas* sp. HMT393 on this model. One LROLK- and one HROLK-associated species were influenced by denture wear: *Gemella morbillorum* showed significantly reduced abundance associated with partial denture wear ( $p_{adj} < 0.01$ ), while *Bergeyella* HMT322 showed increased abundance associated with complete denture wear, although the adjusted p value was not significant ( $p = 0.007$ ,  $p_{adj} = 0.15$ ) (**Table 72**).

**Table 72.** Taxa significantly associated with LROLK, HROLK and OSCC after adjusting for denture wear on MaAsLin2 analysis (negative coefficient values denote reduced abundance)

| Taxa                                              | Variable         | Coefficient | p value  | Adjusted p value |
|---------------------------------------------------|------------------|-------------|----------|------------------|
| <i>Gemella morbillorum</i>                        | LROLK            | 0.55        | 3.50E-07 | 0.0002           |
| <i>Mogibacterium diversum</i>                     | LROLK            | 0.41        | 0.0003   | 0.02             |
| <i>Bergeyella</i> sp. HMT322                      | HROLK            | 0.34        | 2.05E-05 | 0.005            |
| <i>Streptococcus australis infantis</i> clade 431 | HROLK            | 0.48        | 0.0001   | 0.02             |
| <i>Leptotrichia</i> sp. HMT215                    | HROLK            | 0.36        | 0.0002   | 0.02             |
| <i>Parvimonas</i> sp. HMT393                      | OSCC             | 0.72        | 0.002    | 0.09             |
| <i>Gemella morbillorum</i>                        | Partial denture  | -0.42       | 8.35E-05 | 0.01             |
| <i>Bergeyella</i> sp. HMT322                      | Complete denture | 0.46        | 0.007    | 0.15             |

### 8.3.7 Microbiome associated with LROLK, HROLK, OSCC adjusted for mouthwash use

Mouthwash use did not influence the microbiome of LROLK or HROLK. Both LROLK-associated species, *Gemella morbillorum* and *Mogibacterium diversum*, and the three HROLK-associated taxa, *Bergeyella* HMT322, *Streptococcus australis infantis* clade 431 and *Leptotrichia* HMT215, demonstrated significantly increased abundance associated with LROLK and HROLK after adjusting for mouthwash use (all  $p < 0.001$ ,  $p_{adj} < 0.05$ ) and no LROLK- or HROLK-associated taxon showed a significant change in abundance associated with mouthwash use (**Table 73**). Previously identified OSCC-associated taxa did not demonstrate significant changes in abundance associated with either OSCC or mouthwash on this model.

**Table 73.** Taxa with significant changes in abundance associated with LROLK and HROLK after adjusting for mouthwash use on MaAsLin2 analysis

| Taxa                                              | Variable | Coefficient | p value  | Adjusted p value |
|---------------------------------------------------|----------|-------------|----------|------------------|
| <i>Gemella morbillorum</i>                        | LROLK    | 0.57        | 1.72E-07 | 0.0001           |
| <i>Mogibacterium diversum</i>                     | LROLK    | 0.41        | 0.0003   | 0.03             |
| <i>Bergeyella</i> sp. HMT322                      | HROLK    | 0.36        | 1.07E-05 | 0.003            |
| <i>Streptococcus australis infantis</i> clade 431 | HROLK    | 0.48        | 0.0001   | 0.02             |
| <i>Leptotrichia</i> sp. HMT215                    | HROLK    | 0.36        | 0.0002   | 0.03             |

### 8.3.8 Microbiome associated with LROLK, HROLK, OSCC adjusted for missing teeth

Eight taxa were identified as associated with LROLK after adjusting for missing teeth. The previously identified LROLK-associated species, *Gemella morbillorum* and *Mogibacterium diversum*, remained associated with LROLK after adjusting for number of missing teeth. *Peptostreptococcaceae XI G1 sulci*, previously identified as associated with LROLK after

adjusting for brushing habit, was also associated with LROLK after adjusting for missing teeth. Five further taxa were identified as associated with LROLK on this model: *Solobacterium moorei*, *Prevotella* sp. HMT396, *Porphyromonas* sp. HMT284, *Alloprevotella tannerae*, and *Aggregatibacter* sp. HMT458. The three previously identified HROLK-associated taxa remained significantly associated with HROLK after adjusting for missing teeth, while the two previously identified OSCC-associated taxa, *Parvimonas* sp. HMT393 and *Fusobacterium nucleatum* subsp. *polymorphum hwasookii*, were also significantly associated with OSCC on this model (**Table 74**). No LROLK-, HROLK- or OSCC-associated taxa were associated with any number of missing teeth.

**Table 74.** Taxa associated with LROLK, HROLK and OSCC after adjusting for missing teeth on MaAsLin2 analysis

| Taxa                                                               | Variable | Coefficient | p value  | Adjusted p value |
|--------------------------------------------------------------------|----------|-------------|----------|------------------|
| <i>Gemella morbillorum</i>                                         | LROLK    | 0.54        | 9.18E-06 | 0.001            |
| <i>Mogibacterium diversum</i>                                      | LROLK    | 0.57        | 1.31E-05 | 0.002            |
| <i>Peptostreptococcaceae XI G1 sulci</i>                           | LROLK    | 0.27        | 8.91E-05 | 0.007            |
| <i>Solobacterium moorei</i>                                        | LROLK    | 0.43        | 0.0006   | 0.03             |
| <i>Prevotella</i> sp. HMT396                                       | LROLK    | 0.35        | 0.0008   | 0.04             |
| <i>Porphyromonas</i> sp. HMT284                                    | LROLK    | 0.33        | 0.0009   | 0.04             |
| <i>Alloprevotella tannerae</i>                                     | LROLK    | 0.48        | 0.002    | 0.06             |
| <i>Aggregatibacter</i> sp. HMT458                                  | LROLK    | 0.30        | 0.003    | 0.08             |
| <i>Bergeyella</i> sp. HMT322                                       | HROLK    | 0.37        | 3.98E-05 | 0.004            |
| <i>Streptococcus australis infantis</i> clade 431                  | HROLK    | 0.50        | 0.0003   | 0.02             |
| <i>Leptotrichia</i> sp. HMT215                                     | HROLK    | 0.33        | 0.002    | 0.06             |
| <i>Parvimonas</i> sp. HMT393                                       | OSCC     | 0.72        | 0.002    | 0.07             |
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum hwasookii</i> | OSCC     | 0.62        | 0.004    | 0.09             |

### 8.3.9 Microbiome associated with LROLK, HROLK, OSCC adjusted for mucosal site

As previously discussed, mucosal site has a significant influence on mucosal microbial colonisation. A multivariate MaAsLin2 analysis was therefore carried out to determine if the observed changes in taxa associated with LROLK, HROLK and OSCC were independent of mucosal site. *Gemella morbillorum* and *Mogibacterium diversum* remained significantly associated with LROLK, and *Bergeyella* sp. HMT322, *Streptococcus australis infantis* clade 431 and *Leptotrichia* sp. HMT215 remained significantly associated with HROLK, after adjusting for site. *Parvimonas* sp. HMT393, previously associated with OSCC after adjusting for smoking, brushing habit, denture wear and number of missing teeth, and *Fusobacterium nucleatum* subsp. *polymorphum hwasookii*, previously significantly associated with OSCC after adjusting for smoking and number of missing teeth, were both also significantly associated with OSCC after adjusting for site. Unsurprisingly, given earlier findings of microbiome site specificity (**Chapter 6, section 6.3.8**), two of these taxa also showed significant changes in abundance associated with mucosal site, *Gemella morbillorum* and *Leptotrichia* sp. HMT215 showing significantly increased abundance associated with PAL sites as well as with LROLK and HROLK respectively (**Table 75**).

**Table 75.** Taxa significantly associated with LROLK, HROLK and OSCC after adjusting for mucosal site on MaAsLin2 analysis. Site associated (if any) also shown with p value and adjusted p value. (Buccal mucosa/alveolus/gingiva used as reference)

| Taxa                                                               | Variables associated | Coefficient  | p value          | Adjusted p value |
|--------------------------------------------------------------------|----------------------|--------------|------------------|------------------|
| <i>Gemella morbillorum</i>                                         | LROLK                | 0.56         | 3.11E-07         | 8.64E-05         |
| <i>Mogibacterium diversum</i>                                      | LROLK<br>PAL         | 0.41<br>0.53 | 0.0003<br>0.0003 | 0.01<br>0.009    |
| <i>Streptococcus australis infantis</i> clade 431                  | HROLK                | 0.51         | 4.88E-05         | 0.005            |
| <i>Bergeyella</i> sp. HMT322                                       | HROLK                | 0.31         | 0.0002           | 0.007            |
| <i>Leptotrichia</i> sp. HMT215                                     | HROLK<br>PAL         | 0.35<br>0.55 | 0.0003<br>0.0001 | 0.01<br>0.006    |
| <i>Parvimonas</i> sp. HMT393                                       | OSCC                 | 0.74         | 0.001            | 0.02             |
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum hwasookii</i> | OSCC                 | 0.63         | 0.003            | 0.05             |



A DESeq2 analysis to identify taxa with significant changes in abundance in HROLK relative to LROLK by mucosal site was also carried out. HROLKs on DT/LBT showed significantly increased abundance of *Leptotrichia* sp. HMT221. Other taxa identified as associated with HROLK on the DT/LBT were not significantly associated after adjustment. While MaAsLin2 also found an association with this *Leptotrichia* species and HROLK, the association was not statistically significant ( $p=0.05$ ,  $padj=0.21$ ) (Table 76).

**Table 76.** Results of DESeq analysis of HROLK versus LROLK for DT/LBT sites showing taxa with significant changes in abundance in HROLK relative to LROLK at these sites. MaAsLin2 associations also shown ( $padj$ =adjusted p value) (negative log2 fold change denotes reduce abundance).  $\ast = padj < 0.1$

| Taxa                                                                                                                                                                      | log2 fold change | p value DESeq | padj DESeq | MaAsLin2 association             | p value MaAsLin2 | padj MaAsLin2 |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|---------------|------------|----------------------------------|------------------|---------------|
| <i>Leptotrichia</i> sp. HMT221                                                                                                                                            | 22.41            | 1.66E-25      | 2.04E-23*  | Increased HROLK<br>Former smoker | 0.05<br>0.06     | 0.21<br>0.22  |
| <i>Prevotella nanceiensis</i>                                                                                                                                             | 6.04             | 0.002         | 0.10       | Increased HROLK                  | 0.04             | 0.18          |
| <i>Fusobacterium periodonticum</i><br><i>hwasookii</i>                                                                                                                    | 4.62             | 0.003         | 0.10       | NA                               | NA               | NA            |
| <i>Streptococcus australis infantis</i> clade 431                                                                                                                         | 3.71             | 0.006         | 0.14       | Increased HROLK*                 | 4.88E-05         | 0.005*        |
| <i>Parvimonas</i> sp. HMT393                                                                                                                                              | 4.43             | 0.005         | 0.14       | Increased HROLK                  | 0.06             | 0.24          |
| <i>Streptococcus</i> sp. HMT064 <i>australis</i> subsp. <i>dentisani</i> clade 398 sp. HMT423 <i>infantis</i> clade 431 <i>infantis</i> clade 638 <i>mitis pneumoniae</i> | 1.53             | 0.008         | 0.15       | NA                               | NA               | NA            |
| <i>Neisseria perflava flavescens</i>                                                                                                                                      | 3.38             | 0.009         | 0.16       | NA                               | NA               | NA            |
| <i>Streptococcus infantis</i> clade 431                                                                                                                                   | 4.29             | 0.01          | 0.17       | Increased HROLK                  | 0.03             | 0.17          |
| <i>Porphyromonas pasteri</i>                                                                                                                                              | 3.80             | 0.01          | 0.17       | NA                               | NA               | NA            |
| <i>Leptotrichia</i> sp. HMT215                                                                                                                                            | 3.35             | 0.01          | 0.17       | Increased HROLK*                 | 0.0003           | 0.01*         |
| <i>Gemella sanguinis</i>                                                                                                                                                  | 1.71             | 0.02          | 0.17       | Increased DT/LBT*                | 0.0004           | 0.01*         |
| <i>Streptococcus</i> sp. HMT064 sp. HMT423 <i>mitis</i>                                                                                                                   | 5.16             | 0.02          | 0.17       | NA                               | NA               | NA            |
| <i>Leptotrichia</i> sp. HMT225                                                                                                                                            | 5.96             | 0.02          | 0.19       | NA                               | NA               | NA            |
| <i>Streptococcus vestibularis salivarius</i>                                                                                                                              | -2.13            | 0.02          | 0.17       | Decreased HROLK                  | 0.03             | 0.15          |

Both DESeq2 and MaAsLin2 identified *Rothia mucilaginosa* as significantly associated with PAL sites, however DESeq2 found this species significantly more abundant on

HROLKs whereas MaAsLin2 did not find an association with either LROLK or HROLK (Table 77).

**Table 77.** Results of DESeq2 analysis of HROLK v LROLK for PAL sites showing taxa with significant changes in abundance in HROLK relative to LROLK at these sites. MaAsLin2 associations also shown (padj=adjusted p value)

| Taxa                       | log2 fold change | p value DESeq | padj DESeq | MaAsLin2 association       | p value MaAsLin2  | padj MaAsLin2    |
|----------------------------|------------------|---------------|------------|----------------------------|-------------------|------------------|
| <i>Rothia mucilaginosa</i> | 2.48             | 0.0009        | 0.07       | Increased<br>PAL<br>DT/LBT | 0.009<br>2.82E-08 | 0.09<br>3.13E-05 |

The DESeq2 FOM/VT analysis identified *Leptotrichia* sp. HMT225 as significantly more abundant on FOM/VT HROLKs compared with FOM/VT LROLKs. This species was not identified on any MaAsLin2 analysis (Table 78).

**Table 78.** Results of DESeq2 analysis of HROLK v LROLK for FOM/VT sites showing taxa with significant changes in abundance in HROLK relative to LROLK at these sites. MaAsLin2 associations also shown (padj=adjusted p value)

| Taxa                           | log2 fold change | p value DESeq | padj DESeq | MaAslin2 association | p value MaAsLin2 | padj MaAsLin2 |
|--------------------------------|------------------|---------------|------------|----------------------|------------------|---------------|
| <i>Leptotrichia</i> sp. HMT225 | 22.7             | 5.46E-20      | 2.06E-17   | NA                   | NA               | NA            |

No significant changes in any taxa were identified in HROLKs at Buc/Alv/Gin sites relative to LROLKs on DESeq2 analysis (all padj>0.1) (Table 79).

**Table 79.** Results of DESeq2 analysis of HROLK v LROLK for Buc/Alv/Gin sites showing taxa with significant changes in abundance in HROLK relative to LROLK at these sites. MaAsLin2 associations also shown (padj=adjusted p value, negative log2 fold change denotes reduced abundance) (\*=padj<0.1)

| Taxa                                                                        | log2 fold change | p value DESeq | padj DESeq | MaAslin2 association                | p value MaAsLin2      | padj MaAsLin2         |
|-----------------------------------------------------------------------------|------------------|---------------|------------|-------------------------------------|-----------------------|-----------------------|
| <i>Rothia mucilaginosa</i>                                                  | 1.58             | 0.002         | 0.18       | Reduced Buc/Alv/Gin* FOM/VT*        | 0.009<br>0.004        | 0.07*<br>0.04*        |
| <i>Granulicatella adiacens</i>                                              | 1.22             | 0.001         | 0.18       | NA                                  | NA                    | NA                    |
| <i>Streptococcus parasanguinis</i> clade 411 <i>parasanguinis</i> clade 721 | 1.59             | 0.006         | 0.23       | Reduced Buc/Alv/Gin FOM/VT*         | 0.03<br>0.01          | 0.14<br>0.09*         |
| <i>Streptococcus mitis pneumoniae</i>                                       | 2.29             | 0.009         | 0.25       | Increased Buc/Alv/Gin* FOM/VT       | 0.003<br>0.02         | 0.04*<br>0.13         |
| <i>Streptococcus infantis</i> clade 638                                     | 2.33             | 0.010         | 0.25       | NA                                  | NA                    | NA                    |
| <i>Veillonella rogosae dispar</i>                                           | 1.62             | 0.01          | 0.28       | Reduced Buc/Alv/Gin* FOM/VT*        | 0.0007<br>5.13E-05    | 0.01*<br>0.002*       |
| <i>Capnocytophaga</i> sp.HMT326                                             | -4.30            | 0.006         | 0.23       | Increased Buc/Alv/Gin               | 0.03                  | 0.16                  |
| <i>Saccharibacteria</i> TM7 G1 <i>bacterium</i> HMT348                      | -3.64            | 0.01          | 0.29       | Increased LROLK                     | 0.04                  | 0.19                  |
| <i>Aggregatibacter</i> sp. HMT458                                           | -3.59            | 0.003         | 0.18       | Increased LROLK* Buc/Alv/Gin FOM/VT | 0.003<br>0.03<br>0.03 | 0.04*<br>0.16<br>0.11 |

### 8.3.10 Microbiome associated with LROLK, HROLK, OSCC adjusted for clinical appearance of OLK

*Bergeyella* sp. HMT322 was the only taxon significantly associated with HROLK after adjusting for the clinical appearance, homogenous or non-homogenous, of the OLK (padj=0.03).

### 8.3.11 Microbiome associated with LROLK, HROLK, OSCC adjusted for multiple measures of oral hygiene

A MaAsLin2 model assessing the influence of multiple measures of oral hygiene and health (brushing habit, oral hygiene score, denture wear, mouthwash use and number of missing teeth) on LROLK and HROLK colonisation was then constructed. The two previously identified LROLK- associated taxa, *Gemella morbillorum* and *Mogibacterium diversum*, along with *Peptostreptococcaceae XI G1 sulci*, were significantly associated with LROLK after adjusting for multiple measures of oral hygiene and health (all padj<0.07). Interestingly, *Leptotrichia* HMT215, previously identified as associated with HROLK, appeared to be associated with LROLK on this model, although the adjusted p value was not significant. The other two previously identified HROLK-associated taxa, *Bergeyella* sp. HMT322, and *Streptococcus australis infantis* clade 431, remained associated with OLK on this model, although the adjusted p values failed to reach significance (**Table 80**).

**Table 80.** LROLK- and HROLK-associated taxa showing significant changes in abundance on a MaAslin2 model adjusted for multiple measures of oral hygiene

| Taxa                                              | Variable | Coefficient | p value  | Adjusted p value |
|---------------------------------------------------|----------|-------------|----------|------------------|
| <i>Gemella morbillorum</i>                        | LROLK    | 0.61        | 7.40E-06 | 0.008            |
| <i>Peptostreptococcaceae XI G1 sulci</i>          | LROLK    | 0.28        | 0.0001   | 0.04             |
| <i>Mogibacterium diversum</i>                     | LROLK    | 0.50        | 0.0005   | 0.06             |
| <i>Leptotrichia</i> sp. HMT215                    | LROLK    | 0.41        | 0.004    | 0.16             |
| <i>Streptococcus australis infantis</i> clade 431 | HROLK    | 0.50        | 0.001    | 0.12             |
| <i>Bergeyella</i> sp. HMT322                      | HROLK    | 0.31        | 0.002    | 0.13             |

### 8.3.12 Microbiome associated with LROLK, HROLK, OSCC adjusted for site and smoking model

Finally, with site and smoking previously demonstrated to be the most significant influences on the oral microbiome outside of the presence of an OLK, a MaAsLin2 model comprising LROLK, HROLK, OSCC, site and smoking was created. All taxa previously associated with LROLK, HROLK and OSCC remained associated on this model, demonstrating that the associations with LROLK, HROLK and OSCC were independent of mucosal site and smoking status (**Table 79**). Sixteen further taxa demonstrated increased abundance associated with LROLK, with the majority of these also being associated with LROLK on the MaAsLin2 smoking model (**Table 68**). *Saccharibacteria* TM7 G3 *bacterium* HMT351 demonstrated reduced abundance associated with HROLK, an association not previously noted. Finally, *Prevotella* sp. HMT305 demonstrated increased abundance associated with OSCC, an association also seen on the MaAslin2 smoking model (**Table 68**).

**Table 79.** Results of MaAsLin2 analysis showing taxa significantly associated with LROLK, HROLK and OSCC after adjusting for mucosal site and smoking. (Only LROLK, HROLK and OSCC results shown, negative coefficient values denote reduced abundance).

| Taxa                                               | Variable | Coefficient | p value  | Adjusted p value |
|----------------------------------------------------|----------|-------------|----------|------------------|
| <i>Gemella morbillorum</i>                         | LROLK    | 0.57        | 1.75E-07 | 2.15E-05         |
| <i>Mogibacterium diversum</i>                      | LROLK    | 0.41        | 0.0003   | 0.006            |
| <i>Aggregatibacter</i> sp. HMT458                  | LROLK    | 0.27        | 0.002    | 0.03             |
| <i>Lachnospiraceae</i> G 3 <i>bacterium</i> HMT100 | LROLK    | 0.23        | 0.003    | 0.03             |
| <i>Porphyromonas endodontalis</i>                  | LROLK    | 0.35        | 0.004    | 0.04             |
| <i>Porphyromonas</i> sp. HMT284                    | LROLK    | 0.23        | 0.004    | 0.04             |
| <i>Parvimonas</i> sp. HMT110                       | LROLK    | 0.19        | 0.004    | 0.04             |
| <i>Bacteroidales</i> G2 <i>bacterium</i> HMT274    | LROLK    | 0.23        | 0.005    | 0.05             |
| <i>Actinomyces</i> sp. HMT169                      | LROLK    | 0.25        | 0.005    | 0.05             |
| <i>Peptostreptococcus stomatis</i>                 | LROLK    | 0.31        | 0.006    | 0.05             |
| <i>Alloprevotella tannerae</i>                     | LROLK    | 0.36        | 0.006    | 0.06             |
| <i>Peptostreptococcaceae</i> XI G1 <i>sulci</i>    | LROLK    | 0.15        | 0.008    | 0.07             |
| <i>Prevotella oris</i>                             | LROLK    | 0.35        | 0.008    | 0.07             |
| <i>Lachnoanaerobaculum umeaense</i>                | LROLK    | 0.21        | 0.01     | 0.08             |
| <i>Streptococcus constellatus intermedius</i>      | LROLK    | 0.35        | 0.01     | 0.08             |
| <i>Streptococcus gordonii</i>                      | LROLK    | 0.41        | 0.01     | 0.08             |

|                                                                    |       |       |          |       |
|--------------------------------------------------------------------|-------|-------|----------|-------|
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum hwasookii</i> | LROLK | 0.21  | 0.01     | 0.08  |
| <i>Prevotella</i> sp. HMT314                                       | LROLK | 0.18  | 0.01     | 0.09  |
| <i>Streptococcus australis infantis</i> clade 431                  | HROLK | 0.52  | 4.10E-05 | 0.002 |
| <i>Bergeyella</i> sp. HMT322                                       | HROLK | 0.29  | 0.0004   | 0.008 |
| <i>Leptotrichia</i> sp. HMT215                                     | HROLK | 0.34  | 0.0006   | 0.01  |
| <i>Saccharibacteria</i> TM7 G3 <i>bacterium</i> HMT351             | HROLK | -0.19 | 0.01     | 0.08  |
| <i>Parvimonas</i> sp. HMT393                                       | OSCC  | 0.69  | 0.003    | 0.03  |
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum hwasookii</i> | OSCC  | 0.62  | 0.003    | 0.03  |
| <i>Prevotella</i> sp. HMT305                                       | OSCC  | 0.37  | 0.008    | 0.07  |

## 8.4 Microbiome of dysplastic OLK graded by the 3-tier WHO classification

Acknowledging that dysplasia is currently graded according to the 3-tier WHO 2017 classification (mild, moderate and severe)<sup>114</sup>, further MaAsLin2 analysis was carried out with dysplasia categorised in this manner (**Table 60**). Analysis was carried out for all variables as in **section 8.3 (8.2.1 - 8.2.10)** but only the results of models of multiple measures of oral hygiene and site and smoking are presented below.

### 8.4.1 Microbiome associated with dysplasia graded by the 3-tier WHO classification adjusted for multiple measures of oral hygiene

A MaAsLin2 model assessing the influence of multiple measures of oral hygiene and health (brushing habit, oral hygiene score, denture wear, mouthwash use and number of missing teeth) on the colonisation patterns of OLKs with no, mild, moderate and severe dysplasia and OSCC was constructed. The findings were similar to the model with dysplasia classified in a binary manner (**Table 80**), with *Peptostreptococcaceae XI G1 sulci* associated with OLKs with no dysplasia while *Gemella morbillorum* and *Mogibacterium diversum* showed increased abundance associated with mildly dysplastic OLKs. Three other taxa, *Peptostreptococcus stomatis*, *Ruminococcaceae G2 bacterium* HMT085 and *Prevotella pallens* were also significantly associated with OLKs which showed no dysplasia on biopsy. *Bergeyella* HMT322 and *Streptococcus australis infantis* clade 431, previously associated with HROLK, showed significantly increased abundance associated with severe dysplasia on this model. *Saccharibacteria TM7 G6 bacterium* HMT870, which had not previously been identified as significantly associated with OLK, was also significantly associated with

severe dysplasia on this model. No taxa were associated with moderate dysplasia in this model.

**Table 80.** Taxa showing significant changes in abundance associated with degree of dysplasia on a MaAslin2 model adjusted for multiple measures of oral hygiene

| Taxa                                                   | Variable         | Coefficient | p value  | Adjusted p value |
|--------------------------------------------------------|------------------|-------------|----------|------------------|
| <i>Peptostreptococcaceae</i> XI G1 <i>sulci</i>        | No dysplasia     | 0.93        | 3.02E-07 | 0.0007           |
| <i>Peptostreptococcus stomatis</i>                     | No dysplasia     | 1.47        | 3.16E-05 | 0.02             |
| <i>Ruminococcaceae</i> G2 <i>bacterium</i> HMT085      | No dysplasia     | 0.93        | 0.0001   | 0.03             |
| <i>Prevotella pallens</i>                              | No dysplasia     | 1.31        | 0.0004   | 0.05             |
| <i>Gemella morbillorum</i>                             | Mild dysplasia   | 0.59        | 3.10E-05 | 0.02             |
| <i>Mogibacterium diversum</i>                          | Mild dysplasia   | 0.49        | 0.001    | 0.099            |
| <i>Bergeyella</i> sp. HMT322                           | Severe dysplasia | 0.61        | 6.24E-06 | 0.005            |
| <i>Streptococcus australis infantis</i> clade 431      | Severe dysplasia | 0.74        | 0.0004   | 0.05             |
| <i>Saccharibacteria</i> TM7 G6 <i>bacterium</i> HMT870 | Severe dysplasia | 0.28        | 0.0007   | 0.08             |

As can be seen on the heatmap (**Figure 53**), very few OLK-associated taxa were affected by measures of oral hygiene, with only *Streptococcus australis infantis* clade 431 showing a significant change ( $p=8.01E-05$ ,  $p_{adj}=0.03$  (data not shown)), increased abundance, associated with fair oral hygiene, as previously demonstrated (**Table 71**).



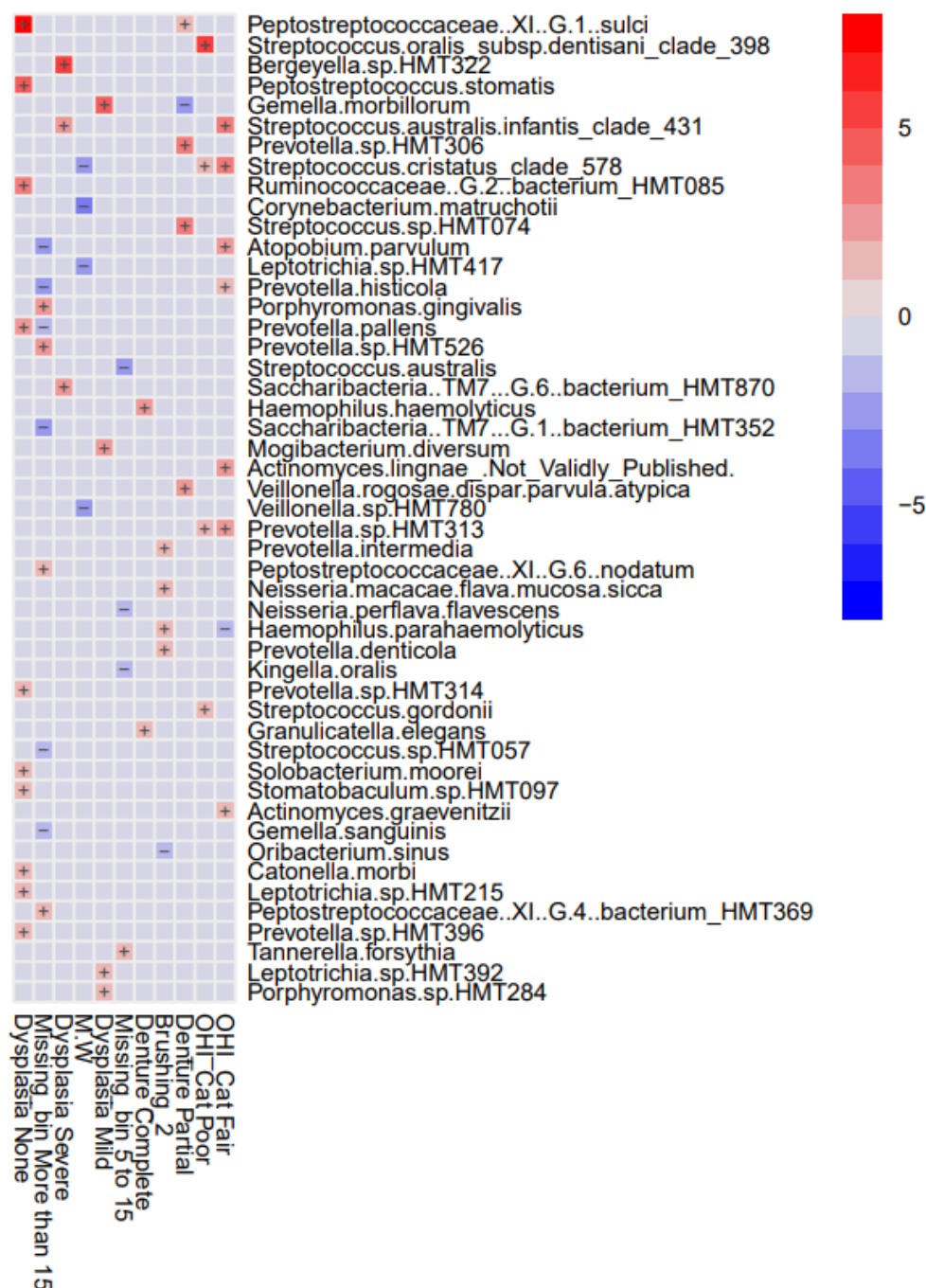


**Figure 53.** Heatmap showing the top 50 taxa with significant changes in abundance associated with no, mild and severely dysplastic OLK and multiple measures of oral hygiene. (Effect size calculated as  $-\log(q\text{-value}) \times (\text{coefficient})$ . Red=increased, blue=reduced. Missing\_bin=number of missing teeth, Brushing\_2=brushing once or less/day, M.W=mouthwash use, OHI\_Cat=oral hygiene category)

### 8.4.2 Microbiome associated with dysplasia graded by the WHO 3-tier classification adjusted for site and smoking

A MaAsLin2 model comprising site, smoking and degree of dysplasia was also constructed and unsurprisingly showed more taxa with associations with no, mild, moderate and severe dysplasia (**Table 84**). Importantly however, taxa previously identified as associated with no and mild dysplasia (and LROLK), severe dysplasia (and HROLK) and OSCC remained associated after adjusting for site and smoking. Four taxa were associated with moderate dysplasia on this model, including *Leptotrichia* HMT215 which was previously associated with HROLK, and *Streptococcus australis infantis* clade 431 which was previously associated with both severe dysplasia and HROLK.

**Table 84.** Results of MaAsLin2 analysis showing taxa significantly associated with no, mild, moderate and severe dysplasia and OSCC after adjusting for mucosal site and smoking. (Only dysplasia and OSCC results shown. Negative coefficient values denote reduced abundance)

| Taxa                                       | Variable       | Coefficient | p value  | Adjusted p value |
|--------------------------------------------|----------------|-------------|----------|------------------|
| <i>Gemella morbillorum</i>                 | No dysplasia   | 1.16        | 1.02E-05 | 0.0006           |
| <i>Alloprevotella</i> sp. HMT914           | No dysplasia   | 0.63        | 0.0002   | 0.005            |
| <i>Prevotella</i> sp. HMT314               | No dysplasia   | 0.59        | 0.0008   | 0.01             |
| <i>Prevotella pallens</i>                  | No dysplasia   | 0.88        | 0.0010   | 0.02             |
| <i>Stomatobaculum</i> sp. HMT097           | No dysplasia   | 0.63        | 0.002    | 0.02             |
| <i>Peptostreptococcus stomatis</i>         | No dysplasia   | 0.83        | 0.002    | 0.03             |
| <i>Prevotella nanceiensis</i>              | No dysplasia   | 0.65        | 0.002    | 0.03             |
| <i>Peptostreptococcaceae</i> XI G1 sulci   | No dysplasia   | 0.41        | 0.002    | 0.03             |
| <i>Porphyromonas</i> sp. HMT284            | No dysplasia   | 0.57        | 0.003    | 0.03             |
| <i>Leptotrichia</i> sp. HMT215             | No dysplasia   | 0.77        | 0.003    | 0.04             |
| <i>Capnocytophaga sputigena</i>            | No dysplasia   | 0.56        | 0.004    | 0.04             |
| <i>Catonella morbi</i>                     | No dysplasia   | 0.81        | 0.004    | 0.05             |
| <i>Porphyromonas</i> sp. HMT277            | No dysplasia   | 0.55        | 0.006    | 0.06             |
| <i>Actinomyces</i> sp. HMT175              | No dysplasia   | 0.72        | 0.006    | 0.06             |
| <i>Solobacterium moorei</i>                | No dysplasia   | 0.63        | 0.010    | 0.08             |
| <i>Gemella morbillorum</i>                 | Mild dysplasia | 0.48        | 2.27E-05 | 0.001            |
| <i>Mogibacterium diversum</i>              | Mild dysplasia | 0.42        | 0.0004   | 0.008            |
| <i>Lachnospiraceae</i> G3 bacterium HMT100 | Mild dysplasia | 0.25        | 0.002    | 0.03             |
| <i>Bacteroidales</i> G2 bacterium HMT274   | Mild dysplasia | 0.26        | 0.003    | 0.03             |
| <i>Porphyromonas endodontalis</i>          | Mild dysplasia | 0.36        | 0.005    | 0.05             |
| <i>Actinomyces</i> sp. HMT169              | Mild dysplasia | 0.27        | 0.005    | 0.05             |
| <i>Parvimonas</i> sp. HMT110               | Mild dysplasia | 0.19        | 0.005    | 0.05             |

|                                                                                                    |                    |       |          |       |
|----------------------------------------------------------------------------------------------------|--------------------|-------|----------|-------|
| <i>Tannerella forsythia</i>                                                                        | Mild dysplasia     | 0.26  | 0.006    | 0.06  |
| <i>Alloprevotella tannerae</i>                                                                     | Mild dysplasia     | 0.36  | 0.008    | 0.07  |
| <i>Leptotrichia</i> sp. HMT392                                                                     | Mild dysplasia     | 0.17  | 0.010    | 0.08  |
| <i>Streptococcus australis infantis</i> clade 431                                                  | Moderate dysplasia | 0.43  | 0.004    | 0.04  |
| <i>Leptotrichia</i> sp. HMT215                                                                     | Moderate dysplasia | 0.33  | 0.004    | 0.05  |
| <i>Streptococcus infantis</i> clade 431                                                            | Moderate dysplasia | 0.44  | 0.006    | 0.06  |
| <i>Rothia aerea dentocariosa</i>                                                                   | Moderate dysplasia | 0.37  | 0.008    | 0.07  |
| <i>Bergeyella</i> sp. HMT322                                                                       | Severe dysplasia   | 0.47  | 3.29E-05 | 0.002 |
| <i>Streptococcus australis infantis</i> clade 431                                                  | Severe dysplasia   | 0.65  | 0.0003   | 0.006 |
| <i>Capnocytophaga leadbetteri</i>                                                                  | Severe dysplasia   | 0.48  | 0.0003   | 0.006 |
| <i>Prevotella nanceiensis</i>                                                                      | Severe dysplasia   | 0.40  | 0.0005   | 0.009 |
| <i>Capnocytophaga gingivalis</i>                                                                   | Severe dysplasia   | 0.38  | 0.0007   | 0.01  |
| <i>Parvimonas</i> sp. HMT393                                                                       | Severe dysplasia   | 0.34  | 0.003    | 0.04  |
| <i>Fusobacterium periodonticum nucleatum</i> subsp. <i>polymorphum</i> sp. HMT204 <i>hwasookii</i> | Severe dysplasia   | 0.54  | 0.006    | 0.06  |
| <i>Granulicatella elegans</i>                                                                      | Severe dysplasia   | 0.38  | 0.007    | 0.06  |
| <i>Leptotrichia</i> sp. HMT215                                                                     | Severe dysplasia   | 0.35  | 0.01     | 0.09  |
| <i>Lachnoanaerobaculum umeaense</i>                                                                | Severe dysplasia   | 0.26  | 0.01     | 0.09  |
| <i>Saccharibacteria</i> TM7 G6 <i>bacterium</i> HMT870                                             | Severe dysplasia   | 0.15  | 0.01     | 0.09  |
| <i>Streptococcus parasanguinis</i> clade 411                                                       | Severe dysplasia   | -0.52 | 0.004    | 0.05  |
| <i>Streptococcus parasanguinis</i> clade 411 <i>parasanguinis</i> clade 721                        | Severe dysplasia   | -0.58 | 0.0008   | 0.01  |
| <i>Parvimonas</i> sp. HMT393                                                                       | OSCC               | 0.71  | 0.002    | 0.02  |
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum hwasookii</i>                                 | OSCC               | 0.63  | 0.003    | 0.03  |

## 8.5 Discussion

As previously discussed, there is agreement in the literature that the microbiome of OLK differs from that of normal mucosa and of OSCC. However, to my knowledge, no research has been published to date on the microbial colonisation patterns of OLKs with different degrees of dysplasia. I have demonstrated for the first time that low-risk (no/mildly dysplastic) OLKs have different colonisation patterns to high-risk (moderate/severely dysplastic) OLKs. The LROLKs in the sample cohort were characterised by significantly increased abundances of *Gemella morbillorum* and *Mogibacterium diversum* (Table 85). *Gemella morbillorum* is a facultatively anaerobic Gram-positive coccus commonly found in

the oropharynx. An uncommon cause of infective endocarditis and bacteraemia<sup>244</sup>, it has recently been associated with colorectal carcinoma<sup>245</sup>. Kwong *et al.*<sup>245</sup> carried out a 10-year retrospective review on 13,096 adults in Hong Kong hospitalised with bacteraemia (defined by positive blood culture with previously identified colorectal carcinoma-associated microbes of genera *Bacteroides*, *Clostridium*, *Filifactor*, *Fusobacterium*, *Gemella*, *Granulicatella*, *Parvimonas*, *Peptostreptococcus*, *Prevotella*, *Solobacterium*, and *Streptococcus*) and compared the incidence of colorectal carcinoma (CRC) in this group with a group of age, gender and co-morbidity matched controls with negative blood cultures. The incidence of CRC was significantly higher in the positive blood culture group ( $p=2.6 \times 10^{-7}$ ). Species level analysis demonstrated that *Gemella morbillorum*, along with other species including *Fusobacterium nucleatum*, was significantly associated with CRC (HR 15.2, 95% CI: 1.54-150,  $p=0.02$ ). However, whether CRC is simply a conducive environment to these species, or these species have role to play in the causation of CRC requires further investigation.

*Mogibacterium diversum* is an anaerobic, rod-shaped, Gram-positive bacteria commonly isolated from the oral cavity. *Peptostreptococcaceae XI GI sulci* was also associated with LROLK after adjustment for smoking, brushing habit and number of missing teeth. Fourteen additional taxa were associated with LROLK after adjusting for smoking, with three of these, and two additional taxa, associated with LROLK after adjusting for number of missing teeth (**Table 85**).

**Table 85.** Summary table showing taxa significantly associated with LROLK alone and associated with LROLK after adjusting for various potential confounding variables (X denotes significant association where  $p_{adj} < 0.1$ )

| Model adjusted for (LROLK + variable)           | LROLK | Smoking | Alcohol use | Brushing habit | Oral hygiene | Denture wear | Mouth-wash | Missing teeth | Mucosal site |
|-------------------------------------------------|-------|---------|-------------|----------------|--------------|--------------|------------|---------------|--------------|
| <b>LROLK associated taxa</b>                    |       |         |             |                |              |              |            |               |              |
| <i>Gemella morbillorum</i>                      | X     | X       | X           | X              | X            | X            | X          | X             | X            |
| <i>Mogibacterium diversum</i>                   | X     | X       | X           | X              |              | X            | X          | X             | X            |
| <i>Peptostreptococcaceae</i> XI G1 sulci        |       | X       |             | X              |              |              |            | X             |              |
| <i>Aggregatibacter</i> sp. HMT458               |       | X       |             |                |              |              |            | X             |              |
| <i>Lachnospiraceae</i> G3 bacterium HMT100      |       | X       |             |                |              |              |            |               |              |
| <i>Actinomyces</i> sp. HMT169                   |       | X       |             |                |              |              |            |               |              |
| <i>Bacteroidales</i> G2 bacterium HMT274        |       | X       |             |                |              |              |            |               |              |
| <i>Parvimonas</i> sp. HMT110                    |       | X       |             |                |              |              |            |               |              |
| <i>Porphyromonas endodontalis</i>               |       | X       |             |                |              |              |            |               |              |
| <i>Porphyromonas</i> sp. HMT284                 |       | X       |             |                |              |              |            | X             |              |
| <i>Peptostreptococcus stomatis</i>              |       | X       |             |                |              |              |            |               |              |
| <i>Alloprevotella tannerae</i>                  |       | X       |             |                |              |              |            | X             |              |
| <i>Corynebacterium durum</i>                    |       | X       |             |                |              |              |            |               |              |
| <i>Streptococcus gordonii</i>                   |       | X       |             |                |              |              |            |               |              |
| <i>Leptotrichia</i> sp. HMT392                  |       | X       |             |                |              |              |            |               |              |
| <i>Prevotella oris</i>                          |       | X       |             |                |              |              |            |               |              |
| <i>Saccharibacteria</i> TM7 G1 bacterium HMT349 |       | X       |             |                |              |              |            |               |              |
| <i>Solobacterium moorei</i>                     |       |         |             |                |              |              |            | X             |              |
| <i>Prevotella</i> sp. HMT396                    |       |         |             |                |              |              |            | X             |              |

HROLKs were characterised by increased abundance of *Bergeyella* sp. HMT322, *Streptococcus australis infantis* clade 431 and *Leptotrichia* sp. HMT215 (**Table 85**). When dysplasia was classified according to the 3-tier WHO classification, *Bergeyella* sp. HMT322 was associated with severe dysplasia, *Streptococcus australis infantis* clade 431 with both moderate and severe dysplasia and *Leptotrichia* sp. HMT215 with moderate dysplasia. *Bergeyella* spp. are aerobic Gram-negative rod-shaped bacteria which have been associated with infective endocarditis<sup>246</sup> and more recently with oesophageal squamous cell

carcinoma<sup>247</sup>. Peters *et al.*<sup>247</sup> identified *Bergeyella* HMT322 as significantly associated (OR 1.81; 95% CI 1.03-3.17, p=0.03) with oesophageal squamous cell carcinoma in a prospective study nested in two cohorts taken from two large prospective US cancer screening and prevention studies. They assessed the relationship of oral bacteria with oesophageal adenocarcinoma (81 cases, 160 matched controls) and oesophageal squamous cell carcinoma (25 cases, 50 matched controls) via mouthwash samples and analysis of the V4 region of the 16S rRNA gene. Notably, the same *Bergeyella* taxon I identified as associated with HROLK and severe dysplasia, HMT322, was one of the taxa identified, along with *Treponema vincentii*, *Prevotella nanceiensis* and *Neisseria weaveri*, as significantly associated with oesophageal squamous cell carcinoma. The genus *Streptococcus* consists of mostly aerobic Gram-positive cocci and comprises the dominant colonisers of the oral cavity. Most OLK and OSCC microbiome studies report decreased abundance of streptococci in OLK and OSCC and my genera level analysis agrees with this (**Figure 47, Table 62**) although it should also be noted that the majority of previous studies do not identify taxa to species level. To date, only one other small biopsy-based study has identified an increase in abundance of *Streptococcus* spp. in OLK<sup>93</sup>. Finally, *Leptotrichia* are anaerobic Gram-negative bacilli commonly present in the oral cavity, particularly in bacterial biofilms. All species of *Leptotrichia* ferment carbohydrates producing lactic acid and can therefore cause tooth decay.

**Table 86.** Summary table showing taxa with significant changes in abundance associated with HROLK alone and associated with HROLK after adjusting for various potential confounding variables (X denotes significant association where  $p_{adj} < 0.1$ , \*denotes reduced abundance).

| Model adjusted for (HROLK + variable)        | HROLK | Smoking | Alcohol use | Brushing habit | Oral hygiene | Denture wear | Mouth-wash | Missing teeth | Mucosal site |
|----------------------------------------------|-------|---------|-------------|----------------|--------------|--------------|------------|---------------|--------------|
| <b>OLK associated taxa</b>                   |       |         |             |                |              |              |            |               |              |
| <i>Bergeyella</i> sp. HMT322                 | X     | X       | X           | X              |              | X            | X          | X             | X            |
| <i>Streptococcus australis</i> clade 431     | X     | X       | X           | X              |              | X            | X          | X             | X            |
| <i>Leptotrichia</i> sp. HMT215               | X     | X       | X           | X              |              | X            | X          | X             | X            |
| <i>Streptococcus vestibularis salivarius</i> |       | X*      |             |                |              |              |            |               |              |

Finally, I found that OSCC was significantly associated with increased abundance of *Parvimonas* sp. HMT393 and *Fusobacterium nucleatum* subsp. *polymorphum hwasookii* (Table 87) in agreement with previous studies that species from genera *Parvimonas* and particularly *Fusobacterium* are frequently associated with OSCC<sup>77,79,80,82,87,88,97,111</sup>. Both *Parvimonas* spp. and *Fusobacterium* spp. are common periodontal pathogens<sup>236</sup>, *Parvimonas* spp. being Gram-positive anaerobic cocci while *Fusobacterium* spp. are Gram-negative anaerobes. *Fusobacterium nucleatum* is also increasingly being associated with colorectal, oesophageal and breast cancer<sup>15–17,19,21,22</sup>.

**Table 87.** Summary table showing taxa significantly associated with OSCC alone and associated with OSCC after adjusting for various potentially influencing variables (X denotes significant association where  $p_{adj} < 0.1$ ).

| Model adjusted for (OSCC + variable)                               | OSCC | Smoking | Alcohol use | Brushing habit | Oral hygiene | Denture wear | Mouth-wash | Missing teeth | Mucosal site |
|--------------------------------------------------------------------|------|---------|-------------|----------------|--------------|--------------|------------|---------------|--------------|
| <b>Taxa associated with OSCC</b>                                   |      |         |             |                |              |              |            |               |              |
| <i>Parvimonas</i> sp. HMT393                                       | X    | X       |             | X              |              | X            |            | X             | X            |
| <i>Fusobacterium nucleatum</i> subsp. <i>polymorphum hwasookii</i> |      | X       |             |                |              |              |            | X             | X            |

## 8.6 Conclusion

I have identified taxa consistently and independently significantly associated with LROLK and HROLK. *Gemella morbillorum* and *Mogibacterium diversum* were associated with LROLK and *Bergeyella* sp. HMT322, *Streptococcus australis infantis* clade 431 and *Leptotrichia* sp. HMT215, associated with HROLK, both alone and after adjusting for all variables except oral hygiene, measured in the sample cohort. Further analysis demonstrated that *Bergeyella* sp. HMT322 appears to be an independent marker for severe dysplasia. *Parvimonas* sp. HMT393 and *Fusobacterium nucleatum* subsp. *polymorphum hwasookii* were also significantly associated with OSCC, although the associations with OSCC were somewhat dependent on variable adjustment.

In summary, with regards to the aims of this chapter:

- The microbial colonisation of OLK was shown to differ with differing degrees of dysplasia raising the potential for microbial colonisation patterns to be used to discriminate sites at lower or higher risk of malignant transformation.



## 9 Predictive modelling

### 9.1 Introduction

The final element of data analysis in this study comprised predictive modelling using machine learning to investigate the ability of, firstly clinical and patient factors and secondly, the microbiome, with and without clinical and patient factors, to predict whether an OLK was likely to be LROLK (no/mild dysplasia) or HROLK (moderate/severe dysplasia). Although huge strides have been made in microbiome research in recent years, there is no definitive evidence as yet that the microbiome can predict the likelihood of development of cancer or pre-cancer from normal tissue. Ongoing research aims to answer this question which could, in time, enable the development of screening tests for cancer or pre-cancer, which in the case of the mouth would have the significant advantage over many other sites of being a non-invasive test. To such end, analysis of microbiome data needs to be able to identify the changes in microbial community composition associated with disease states, in this case OLK, and ideally differentiate high-risk from low-risk OLK, while also accounting for variables known to influence the microbiome, such as smoking and oral site, with the aim of being able to predict the disease state based on microbial taxonomy.

Machine learning (ML), a form of artificial intelligence whereby large datasets can be analysed, and different statistical models trained and evaluated to recognise, classify and predict patterns from data<sup>248</sup>, is needed for such analysis as ML is able to infer relationships between variables and discover patterns that traditional statistical methods cannot. The R packages LASSO<sup>249</sup> (least absolute shrinkage and selection operator) and randomForest<sup>221</sup> (RF) are two of the most commonly used supervised ML methods at the present time - supervised approaches relying on specific labels to train the models and map input features to specified output labels<sup>250</sup>. LASSO modelling is based on logistic regression while RF

uses an ensemble method to make a complex model through combining several simple decision trees. RF does this by using bootstrap resampling (random sampling with replacement) on the dataset to develop multiple individual decision trees each using a single bootstrap set. Majority voting on these individual decision trees is then used to obtain the final output<sup>251</sup>. Over-fitting is limited by cross-validation built into the model<sup>251</sup>. Some authors have used LASSO to identify the best features to include in the model and then applied RF modelling on the selected features<sup>252</sup>. However, others report that, although RF is a more powerful modelling method, LASSO can demonstrate greater performance for moderate sample sets with less predictors which are mainly binary in nature<sup>253</sup>.

## 9.2 Methods

Modelling was first carried out using the clinical and patient factors and the R packages `rpart`<sup>220</sup> and `randomForest`<sup>221</sup> to establish what, if any, combination of features could be used to predict if an OLK was likely to be low-risk or high-risk. All 216 OLKs, and their degree of dysplasia, site and clinical appearance, were included in the dataset along with patient demographics, smoking status, alcohol consumption, denture wear, brushing habit and mouthwash use. The number of missing teeth and the OHI-S could not be included in the analysis as 20-25% of data was missing in these categories (due to these variables being added to the study after it had commenced), and `randomForest`<sup>221</sup> cannot account for this volume of missing data. Any text variables were converted to factors and minimum values of 20 observations per node before a new split and 10 observations per terminal node along with a maximum depth of five were set. Up to 10,000 iterations of Classification trees using random subsets of 90% of samples for each iteration were carried out. 90% of samples within each Classification tree were used to ensure cross validation and avoid over-

estimation of predictive values. Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV) and model accuracies with CIs, along with receiver operator characteristic (ROC) curves and Gini indices, measures of the impurity of the dataset, were generated for each Classification tree.

The R package SIAMCAT<sup>219</sup>, was then used for an initial assessment of the ability of the microbiome with and without clinical and patient factors, to differentiate HROLK from LROLK, and to identify the most predictive taxa. SIAMCAT carries out statistical association testing and statistical modelling, and plots and analyses ROC curves (plot of sensitivity versus 1 – specificity) and precision-recall (PR) curves (plot of PPV versus sensitivity). A total of 628 taxa from 216 OLK samples were included. First the samples were split into cases (HROLK n=125) and controls (LROLK n=91) and low abundance taxa (abundance <0.1%) removed. Associations between microbial features (i.e. taxa) and the cases/controls using a significance level of 0.1 and false discovery rate correction were then computed using default linear modelling, and visualised. Features were normalised using log transformation of the data and z-score (number of standard deviations [SD] above or below the mean) standardisation. Different metadata features, including clinical appearance, site, gender, smoking habit, alcohol consumption and brushing habit were then added manually, at first singly and then in various combinations to ascertain the best model. In preparation for modelling, the data were split into five cross-validation folds, where four fifths of the data were used to train the algorithms and one fifth of the data for the actual testing, based on the best predictions from the training dataset. Testing of various combinations of metadata was carried out using two repetitions (training runs) of LASSO for cross-validation. ROC and PR curves were generated for each model tested and the two best models were chosen based on these values. Ten and 100 iterations with random forest

and LASSO respectively were then performed on these two best models and bar plots showing feature weights and robustness, heatmaps showing z-scores of the metagenomic features and metadata categories, and a boxplot with the proportion of weight for each feature in the model, generated. Finally, confounding factors within the metadata were checked for and a correlation heatmap and quantitative bar plots generated.

To further refine the results and ensure that the results of the clinical/patient factor modelling, and microbiome modelling could be compared accurately via the same methodology, the z-scores of the top 50 microbial features (i.e. taxa) generated by the RF modelling in SIAMCAT were used for additional predictive modelling analysis using the R packages `rpart`<sup>220</sup> and `randomForest`<sup>221</sup>. Up to 10,000 iterations of Classification trees using random subsets of 90% of samples were carried out on the microbiome data to estimate sensitivity, specificity, PPV, NPV and classification accuracy with corresponding CIs, ROC curves and Gini indices. This was followed by running a RF with 500 trees with 25 random variables used for each tree. Finally, up to 10,000 iterations of Classification trees using random subsets of 90% of samples were carried out on the combined clinical/patient and microbiome data to estimate sensitivity, specificity, PPV, NPV and classification accuracy with corresponding CIs, along with ROC curves and Gini indices. This was then followed by running a random forest with 500 trees with 30 random variables used for each tree.

## 9.3 Results

### 9.3.1 Modelling with clinical and patient factors

Predictive modelling with RF using only the clinical and patient factors was able to predict HROLK from LROLK with a sensitivity of 76.3% (95%CI: 68.2%-84.3%) and specificity

of 66.1% (95%CI: 55.8%-76.4%). The PPV of the model was 75.5% (95%CI: 67.3%-83.8%), the NPV was 68.3% (95%CI: 57.6%-78.9%) and the accuracy 71.2% (95%CI: 64.6%-77.8%). Finally, the ROC value, as defined by the area under the curve (AUC) was 78.5% (**Figure 54**) and the estimated error rate was 31%.

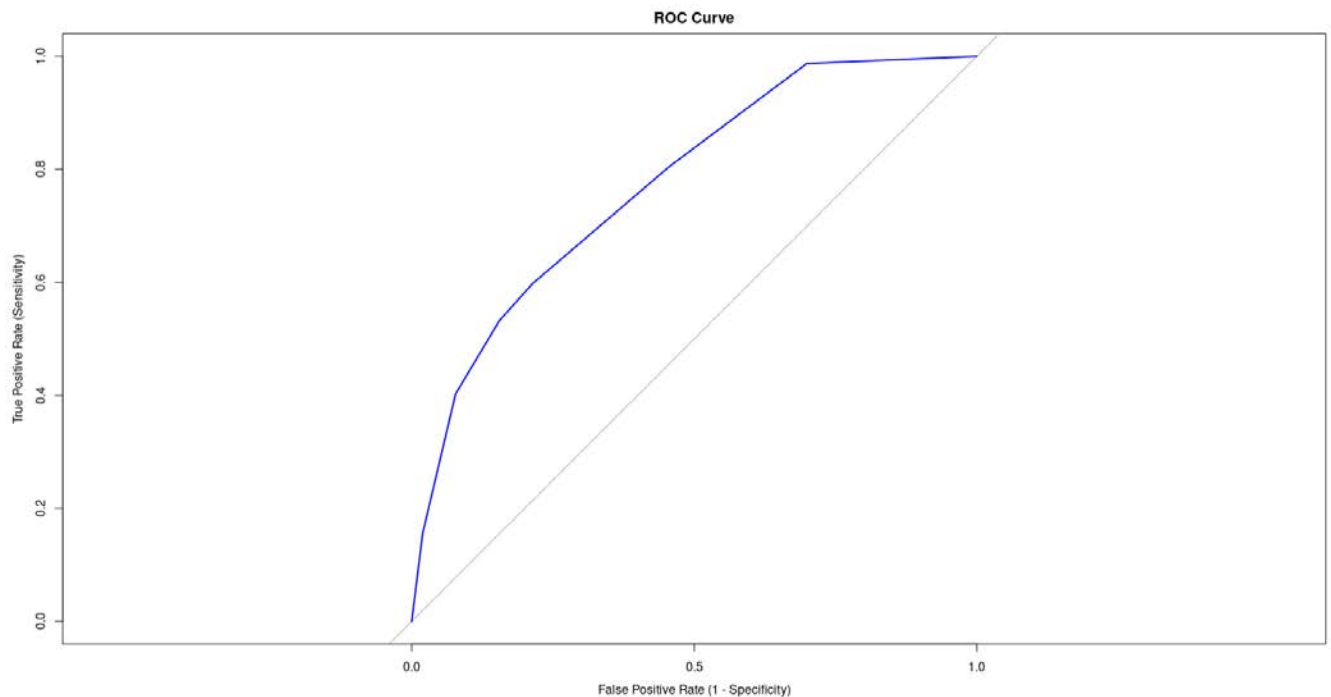


**Figure 54.** Receiver operating characteristic curve for RF modelling using clinical and patient factors (AUC=78.5%)

An illustrative Classification tree (among 10,000 generated) is presented in **Figure 55**. An OLK with a non-homogenous clinical appearance was almost always associated with HROLK (node 13). Working down the decision tree, if a white patch was homogenous then the next parameter to be considered was the patient age. A homogenous OLK in a patient under 40 years of age was predictive of LROLK in approximately 90% of cases (node 3). In those over the age of 40 years, brushing habit was the next most important variable to consider, with brushing once or less per day predictive of HROLK in approximately 65% of cases (node 12). Where brushing occurred more than once per day, site was the next variable

to examine, leading then to mouthwash use for Buc/Alv/Gin OLKs and a further breakdown of age for other sites. Interestingly, following the Classification tree illustrated, age greater than 60 years was predictive of HROLK in approximately 65% of cases of homogenous OLKs on the floor of mouth/ventral tongue, dorsum/lateral tongue or palate in those who brushed once or less per day (node 11).

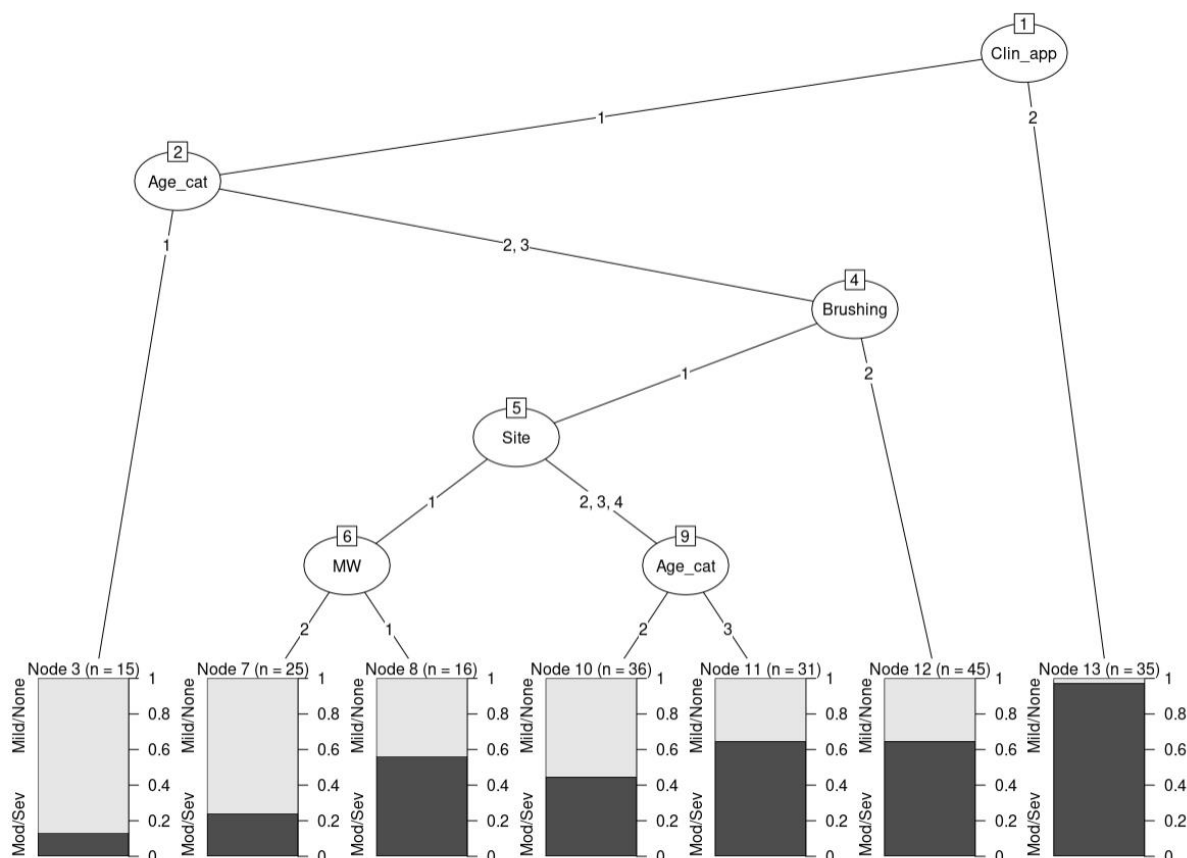


**Figure 55.** Graphic of a Classification tree from RF modelling of clinical and patient factors. (Clin\_app = clinical appearance, 1 = homogenous, 2 = non-homogenous; Age\_cat = age category, 1 = <40 years, 2 = 40-60 years, 3 = >60 years, Brushing: 1 = More than once/day, 2 = once or less/day, Site 1 = Buc/Alv/Gin, 2 = FOM/VT, 3 = DT/LBT, 4 = PAL, MW = mouthwash use, 1 = Yes, 2 = No)

### 9.3.2 Initial modelling with microbiome data

An initial analysis was carried out using the R package SIAMCAT<sup>219</sup> to identify the most discriminatory features for inclusion in further models using R packages rpart<sup>220</sup> and randomForest<sup>221</sup>. Models using the microbial signatures for all samples along with various combinations of clinical and patient features were constructed and tested for their predictive ability to differentiate low and high-risk dysplasia using ROC and precision recall (PR) curves. The best models generated in SIAMCAT all included the clinical appearance of the OLK along with variations of age, gender, OLK site and alcohol consumption (**Table 88**). The predictive values, as defined by the AUC of the ROC and PR curves, of the models ranged from 73.1% to 79.3% with LASSO and from 73.3% to 76.4% with random forest. As mentioned, some authors report better performance of LASSO for smaller datasets with binary variables<sup>253</sup>, which may explain the higher PR curve values with LASSO compared with RF modelling in this dataset. The AUC values for the PR curves were higher than ROC curves in both the LASSO and RF modelling, likely due to the significant difference between the number of homogenous (n=181) and non-homogenous OLKs (n=35) and the unequal number of LROLK (n=91) and HROLK (n=125) in the dataset.

**Table 88.** Results of best models with LASSO (100 iterations) and random forest (10 iterations) (AUC = area under the curve)

| Model   | Microbiome +                                    | AUC for ROC/PR curves | LASSO x 100    | Random Forest x 10 |
|---------|-------------------------------------------------|-----------------------|----------------|--------------------|
| Model 1 | Clinical appearance, age, site, gender, alcohol | ROC<br>PR             | 0.731<br>0.793 | 0.742<br>0.764     |
| Model 2 | Clinical appearance, age, gender                | ROC<br>PR             | 0.730<br>0.793 | 0.733<br>0.755     |

*Streptococcus oralis infantis*, *Mogibacterium diversum*, *Gemella morbillorum*, *Porphyromonas endodontalis*, *Saccharibacteria* TM7 G1 *bacterium* HMT349, *Streptococcus vestibularis salivarius*, *Prevotella melaninogenica*, *Alloprevotella tanneriae* and homogenous clinical appearance represented the nine best predictors of presence of LROLK (or absence of HROLK) on the best random forest model (model 1 **Table 88**) (**Figure 56**, **Figure 57**). Age ranked 18<sup>th</sup>, while site, gender and alcohol consumption ranked outside the top 50 features. These findings agreed with the MaAsLin2 multivariate analysis in Chapter 8 where *Gemella morbillorum* and *Mogibacterium diversum* were consistently significantly associated with LROLK while *Porphyromonas endodontalis*, *Saccharibacteria* TM7 G1 *bacterium* HMT349 and *Alloprevotella tanneriae* (4<sup>th</sup>, 5<sup>th</sup> and 8<sup>th</sup> respectively model 1) and *Actinomyces* sp. HMT169 and *Prevotella oris*, (15<sup>th</sup> and 17<sup>th</sup> respectively in model 1), were also associated with LROLK after adjusting for smoking (**section 8.3.2**). *Leptotrichia* sp. HMT215 and *Streptococcus australis infantis* HMT431, taxa consistently significantly associated with high-risk dysplasia (**Chapter 8**), were ranked 35<sup>th</sup> and 49<sup>th</sup> respectively on this model. The same taxa, along with clinical appearance, were also present amongst the top 50 features on the LASSO analysis of model 1 (**Figure 58**, **Figure 59**), although the order was different, and age was not present. The complete LASSO analysis included fewer features than the random forest analysis because, as mentioned, it is a less powerful tool. Clinical appearance ranked as the best predictive feature on the LASSO analysis. Features associated with high-risk dysplasia appeared to have more weight on



LASSO analysis with the HROLK-associated taxa, *Leptotrichia* sp. HMT215 and *Streptococcus australis infantis* HMT431 ranked 5<sup>th</sup> and 13<sup>th</sup> respectively, and the LROLK-associated taxa *Mogibacterium diversum* and *Gemella morbillorum* ranked 29<sup>th</sup> and 49<sup>th</sup> respectively, on the LASSO model.

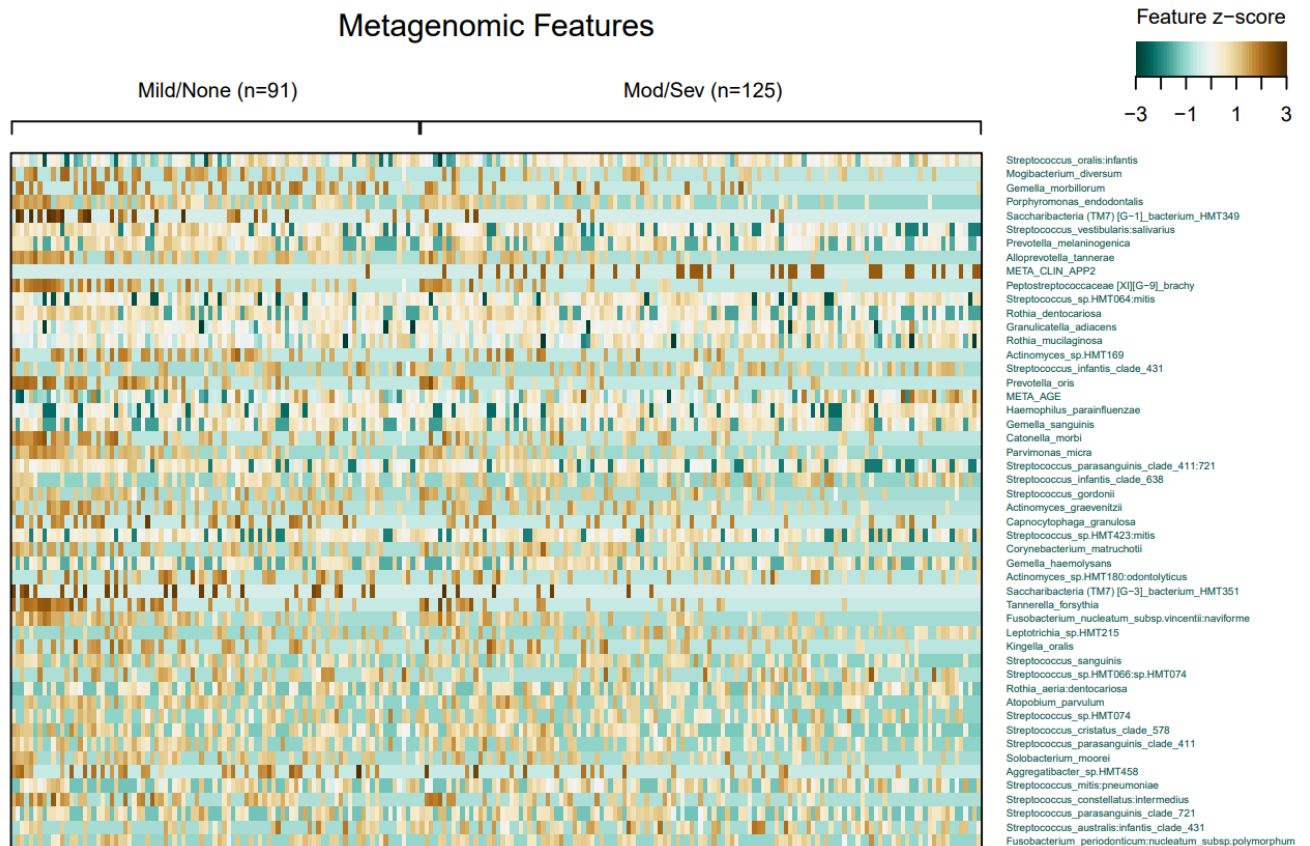


**Figure 56.** Plot showing the top 50 features, i.e. including bacterial taxa and clinical/patient features, from the best random forest model (Model 1: microbiome, clinical appearance, age, site, gender and alcohol consumption). Mild/None = mild/no dysplasia (LROLK). Mod/Sev = moderate/severe dysplasia (HROLK). Brown = increased, Green = decreased where colour intensity relates to the z-score (standard deviations from the mean).

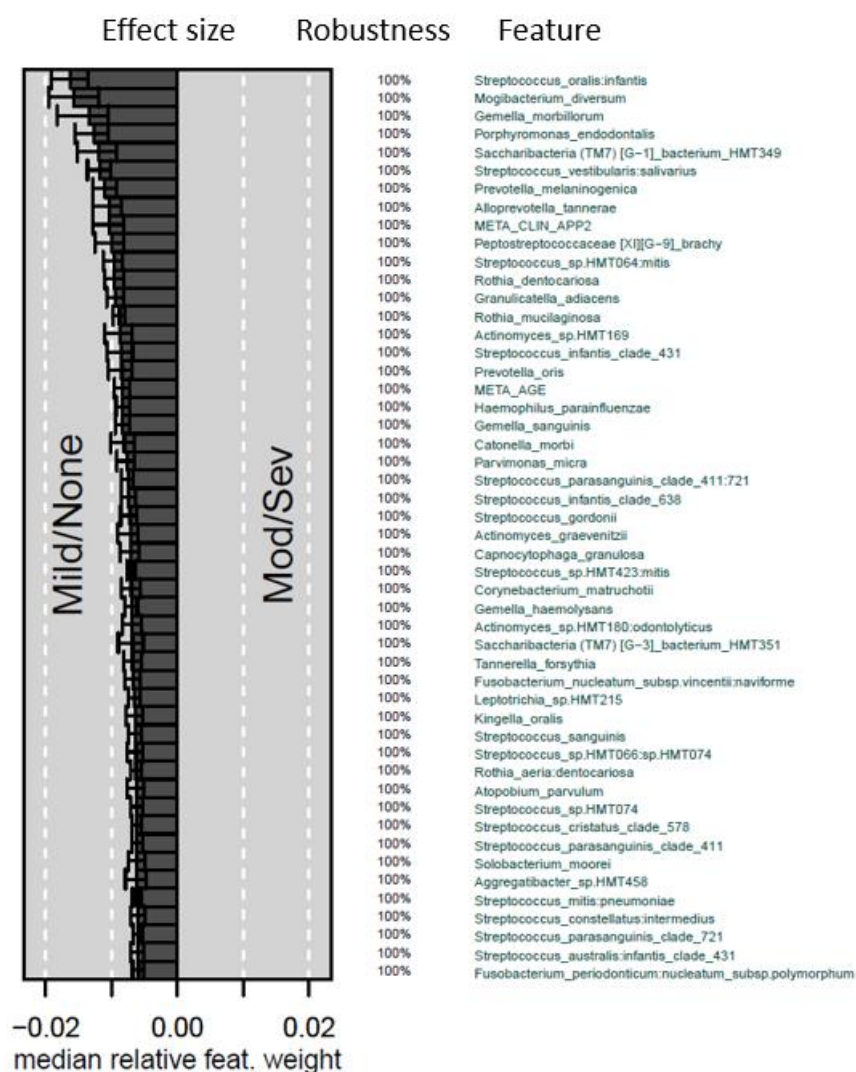


**Figure 57.** Effect size, relative feature weight and robustness of each feature in **Figure 56**.

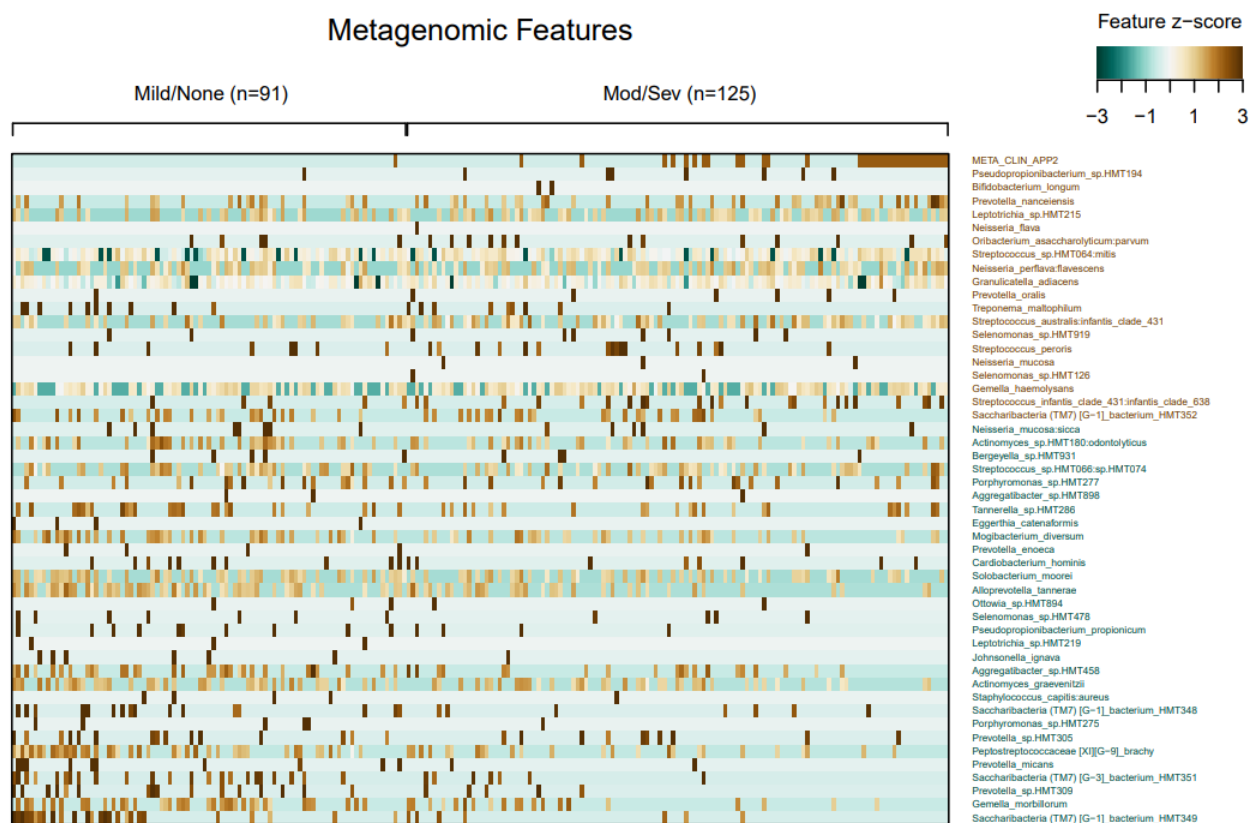


**Figure 58.** Plot of the top 50 features using LASSO modelling of the same model (microbiome, clinical appearance, age, site, gender and alcohol consumption) as **Figure 56**. Mild/None = mild/no dysplasia (LROLK). Mod/Sev = moderate/severe dysplasia (HROLK). Brown=increased, green=decreased, colour intensity relates to z-score (standard deviations from the mean).

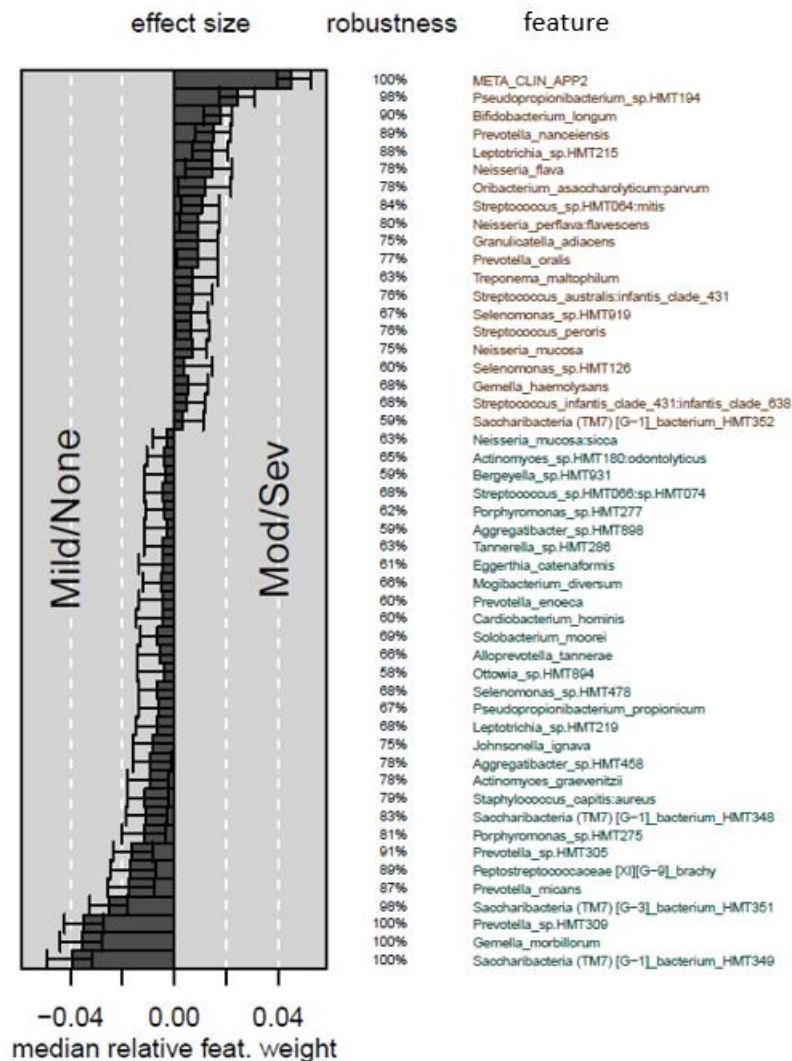


**Figure 59.** Effect size, relative feature weight and robustness of each feature in **Figure 58**

### 9.3.3 Further modelling with microbiome data

Next, the z-scores of the top 50 microbial features generated from the RF modelling in SIAMCAT were used in a further modelling analysis in an attempt to further refine the discriminatory ability of the microbiome alone with regards to differentiating LROLK from HROLK using the R packages rpart<sup>220</sup> and randomForest<sup>221</sup>. This analysis found that these top 50 microbial features could predict HROLK from LROLK with a sensitivity of 85.5% (95%CI: 79.0%-0.92.0%), specificity of 76.5% (95%CI: 67.3%-85.7%), PPV of 83.5%

(95%CI: 76.7%-90.3%), NPV of 79.7% (95%CI: 70.8%-88.7%) and accuracy of 81.0% (95%CI: 75.4%-0.86.6), results which surprisingly were superior to those of the clinical/patient feature modelling. The overall estimated error rate was 27.8%, although this was higher for LROLK at 35.2%, than HROLK at 22.4%. The ROC, with an AUC of 78.8%, for this model is demonstrated in **Figure 60**.

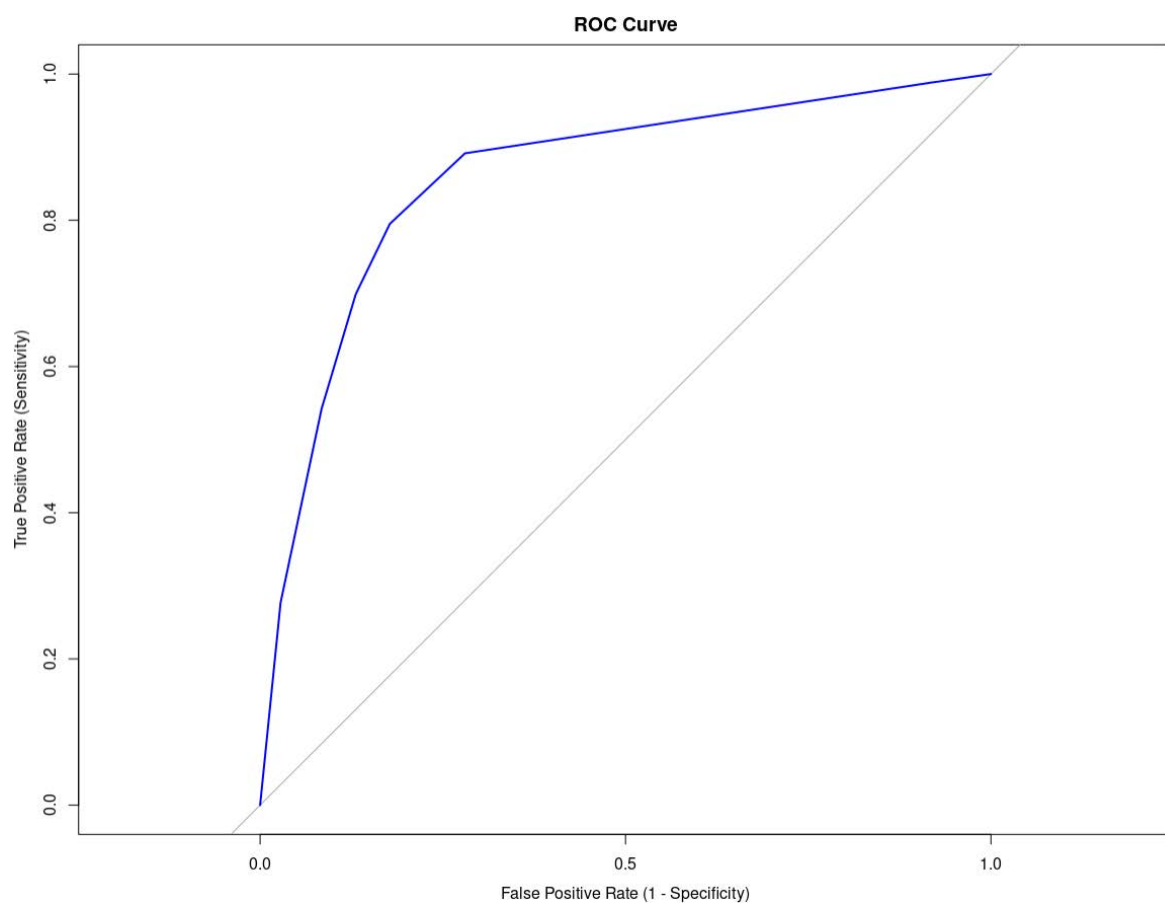


**Figure 60.** Receiver operating characteristic curve for RF modelling using the z-scores of the top 50 microbial features from initial SIAMCAT analysis (AUC=78.8%).

**Figure 61** is an illustrative Classification tree (from 500 generated) from the RF microbiome modelling using the SIAMCAT generated z-scores. *Mogibacterium diversum* was the most significant variable in this analysis as shown by its position at the top of both the Classification tree and Gini index (**Figure 62**). 100% of OLKs with increased relative

abundance of *Mogibacterium diversum* followed by increased relative abundance of *Prevotella oris* were LROLK (node 3 in **Figure 61**). If the abundance of *Mogibacterium diversum* was reduced then the relative abundance of *Porphyromonas endodontalis* was considered, with increased abundance of it being associated with LROLK in 80% of cases (node 8). On the other hand, if the relative abundance of *Porphyromonas endodontalis* was reduced then the abundance of *Gemella morbillorum* was next to be considered, reduced abundance of this species leading then to examination of the abundance of *Saccharibacteria* TM7 G1 *bacterium* HMT349, reduced abundance of this taxa being associated with HROLK in almost 90% of cases (node 15, **Figure 61**). Other variations of the algorithm are shown in **Figure 61**.

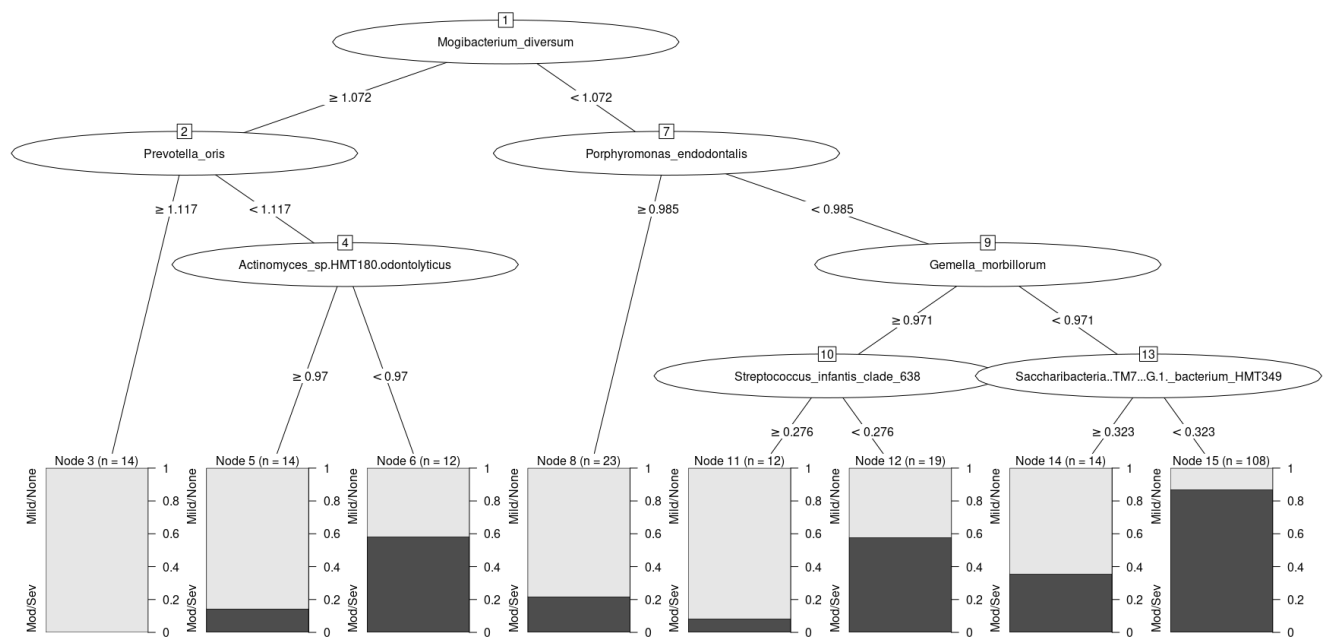


**Figure 61.** Graphic of a Classification tree from RF modelling using the z-scores of the top 50 microbial features from initial SIAMCAT analysis.

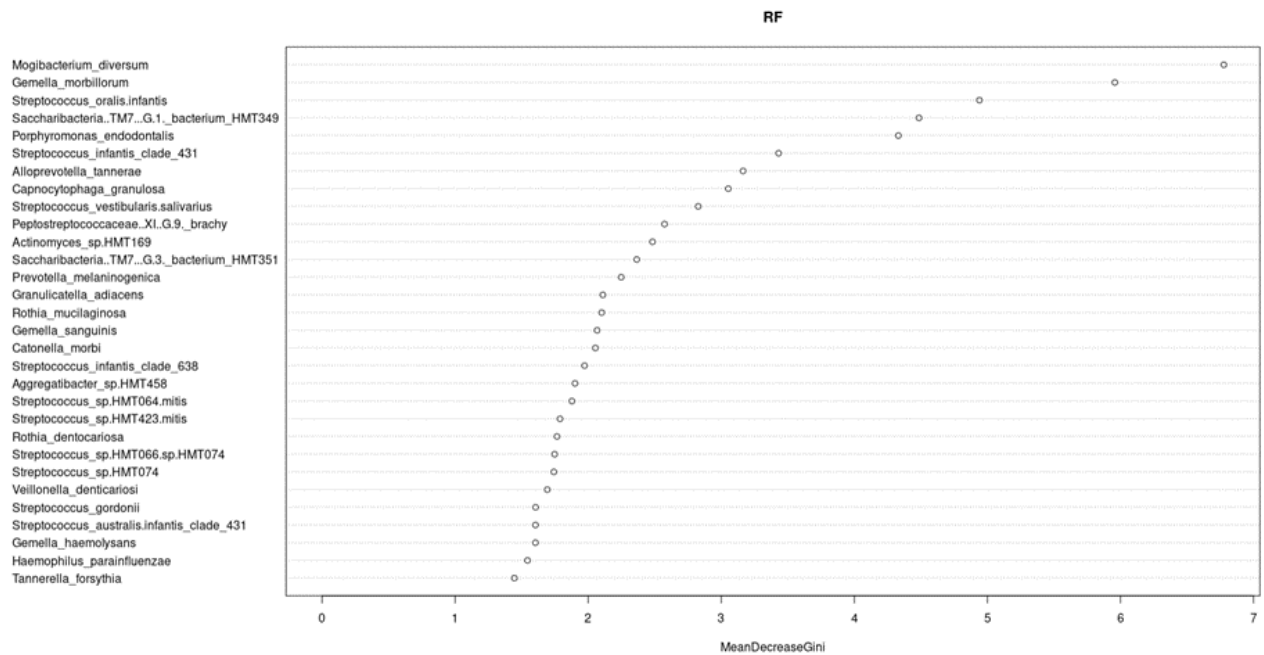


**Figure 62.** Graph of the Gini index, a measure of variable impurity, for the microbial z-score RF analysis. The higher the score on the index the greater the discriminatory value of the variable.

### 9.3.4 Modelling with combined data

Finally, predictive modelling was carried out using both the clinical/patient data detailed in section 9.3.1 and the microbial z-score data from the SIAMCAT analysis using the R packages rpart<sup>220</sup> and randomForest<sup>221</sup>. Modelling using these combined data could differentiate LROLK from HROLK with a sensitivity of 87.4% (95%CI: 80.0%-93.6%) and specificity of 76.5% (95%CI: 67.2%-85.8%). The PPV of the model was 83.4% (95%CI: 76.5%-90.4%), the NPV was 82.2% (95%CI: 73.5%-91.0%) and the accuracy 81.9% (95%CI: 76.3-87.5%). The AUC was 88% (**Figure 63**) and the estimated error rate was 27.6%. As for the microbiome modelling (**section 9.3.2**), the error rate was much lower for predicting HROLK (20.5%) than for LROLK (35.6%).

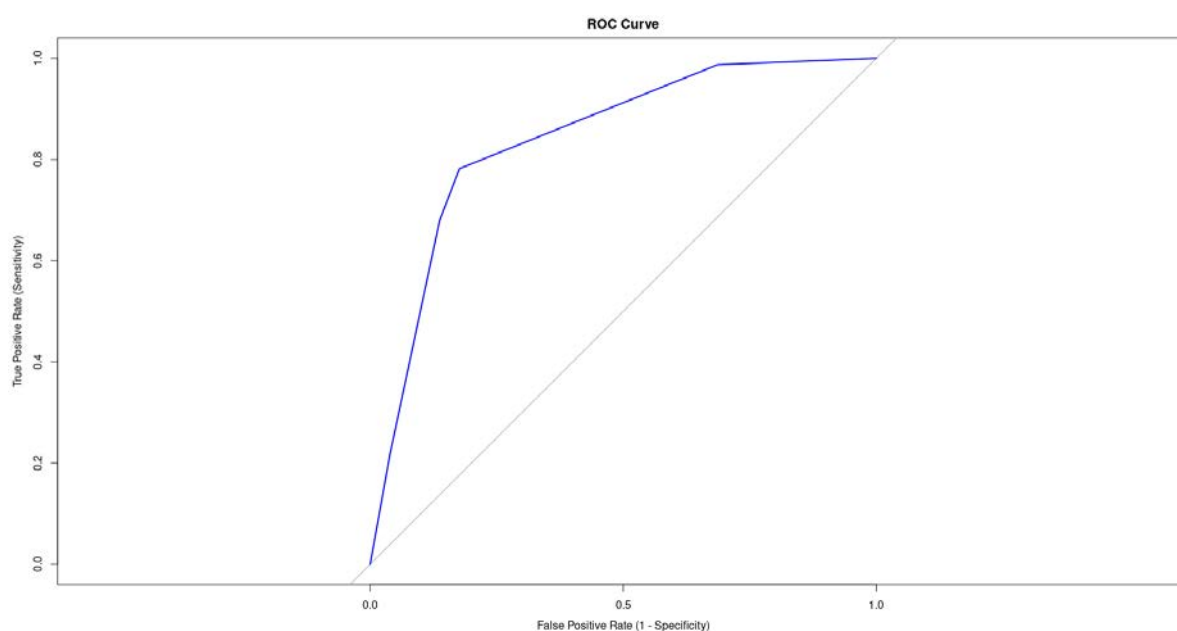


**Figure 63.** ROC curve for RF modelling using the clinical/patient and microbial (z-scores) features (AUC=88%)

An illustrative Classification tree (from 500 generated) is shown in **Figure 64**. As in section **9.3.1** non-homogeneity was almost always associated with HROLK (node 9 in **Figure 64**). If the OLK was homogenous however, then the microbial features were next to be considered. In the Classification tree shown homogenous OLKs with increased abundance of *Mogibacterium diversum* were associated with LROLK in approximately 90% of cases (node 3). On the other hand, if the abundance of *Mogibacterium diversum* was reduced, the abundance of *Porphyromonas endodontalis* was next to be considered, increased abundance of it being associated with LROLK in approximately 90% of cases (node 5). If the abundance of *Porphyromonas endodontalis* was reduced, then the abundance of *Gemella morbillorum* was examined, with reduced abundance being associated with HROLK in approximately 75% of cases (node 8). The Gini index in **Figure 65** shows the most discriminatory features in this analysis; *Streptococcus oralis infantis*, *Gemella morbillorum*,



*Porphyromonas endodontalis*, *Actinomyces* sp. HMT169, clinical appearance of the OLK and *Mogibacterium diversum*, comprising the top six features.

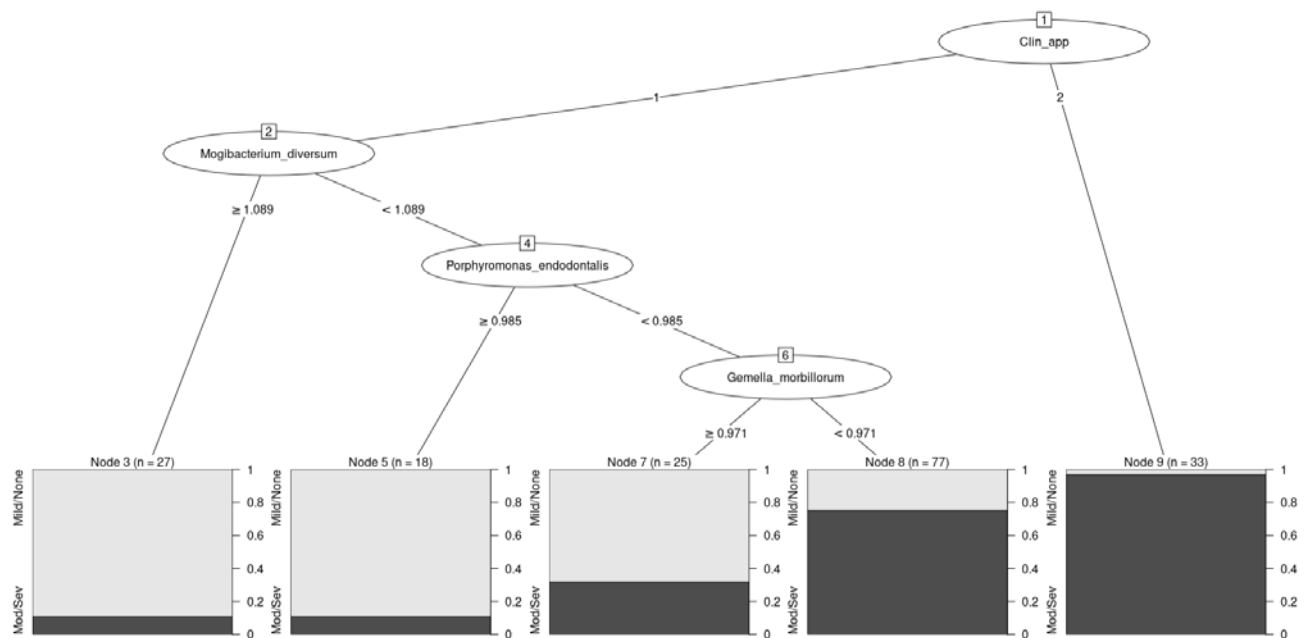


**Figure 64.** Graphic of a Classification tree from RF modelling of clinical/patient factors and microbial z-scores. (Clin\_app = clinical appearance, 1 = homogenous, 2 = non-homogenous)

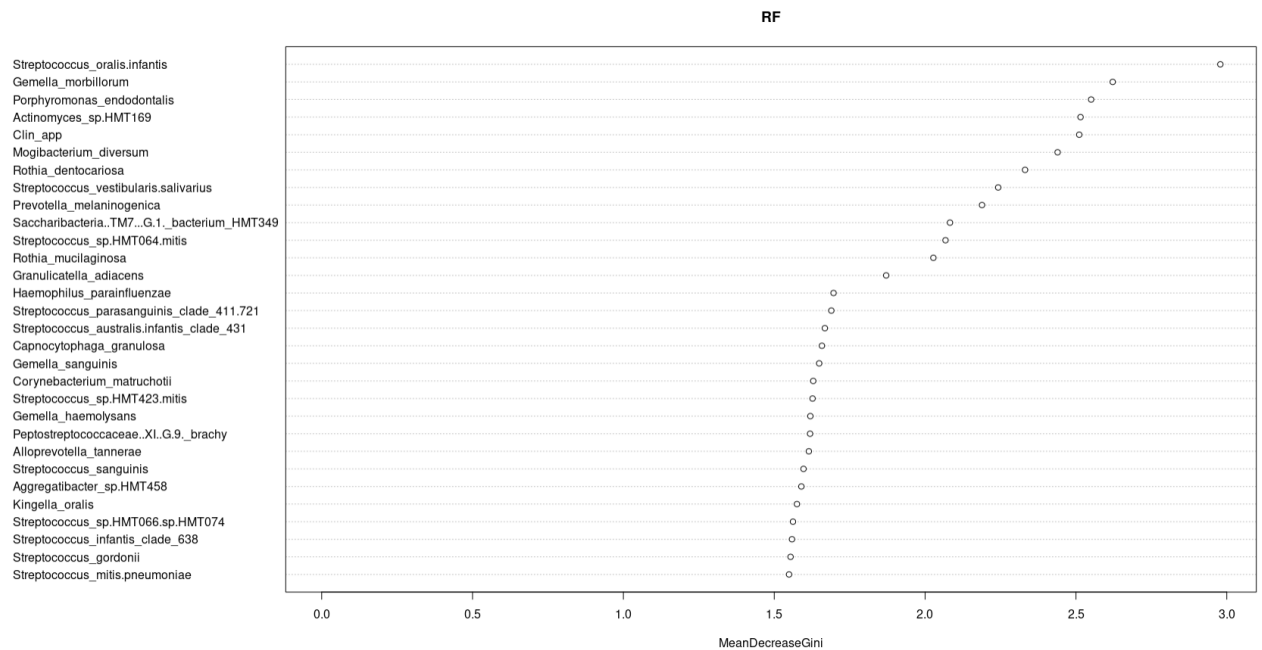


**Figure 65.** Graph of the Gini index of the RF analysis of clinical/patient and microbial features. The higher the score on the index the greater the discriminatory value of the variable.

## 9.4 Discussion

To my knowledge no attempt to date has been made to use predictive modelling to develop an algorithm to predict if an OLK is likely to be low-risk or high-risk. Using a dataset of 216 well-defined OLKs, clinical and patient factors, and random forest modelling, HROLK could be predicted with an overall sensitivity of 76.4% and specificity of 66.1%. However, an algorithm using clinical appearance, age, brushing habit and oral site could predict approximately 95% of HROLK. As non-homogenous OLKs are known to be at increased risk for malignant transformation and are always considered high risk<sup>3,116,127</sup>, the most useful aspect of the algorithm was in predicting HROLK in homogenous OLKs which are typically considered low risk. Tongue and floor of mouth OLKs are also considered high risk locations for malignant transformation of OLK<sup>3,40,129,130,133–136,139,140</sup>, however the literature

does not consider the palate to be a high-risk site. Nonetheless my data indicates that the palate may be a site at increased risk of having high risk dysplasia in certain cohorts. Of note, hard and soft palate OLKs were combined for this analysis due to the low numbers of OLKs at these sites.

Patient age was the second most important feature to be considered after the clinical appearance of the OLK. Increased age is known to be associated with increased risk of malignant transformation of OLK, although the literature varies with regards to the reported age thresholds<sup>3,134,136</sup>. Analysis of my data found that age over 60 years was a predictor of HROLK for homogenous OLKs on the tongue, floor of mouth or palate in those who brushed once or less per day but perhaps more significantly, age under 40 years was associated with approximately 90% of homogenous OLKs being LROLK.

With further development, this simple algorithm has the potential to be used as a screening tool in primary care to guide urgency of referral to tertiary services. While the clinical appearance and site of the OLK and the age of the patient should be standard details included in any referral, brushing habit and mouthwash use may also be variables to consider. Unfortunately, due to missing data, tooth number and oral hygiene score could not be included in this analysis, however they may be variables to consider in future modelling.

No studies to date have investigated the potential of using ML analysis of the oral microbiome to predict the histological status of OLK. Three studies to date have reported on using RF analysis of tissue or salivary microbial profiles to differentiate OSCC or HNSCC from normal tissue<sup>90,254,255</sup>. Shin *et al.*<sup>254</sup> sampled tissue from primary HNSCC (n=24), metastatic HNSCC (n=10) and normal adjacent mucosa (n=24) and sequenced the V4

region of the 16S rRNA gene. They found that RF analysis could identify 76% of HNSCC (either primary or metastatic) but only 59% of normal samples. The distance separating the normal tissue samples from HNSCC samples was not specified and so it is possible that adjacent clinically normal mucosa was partly colonised by pathogens associated with HNSCC perhaps contributing to the poor predictive values for normal tissue. Eun *et al.*<sup>255</sup> found that RF ML could differentiate OSCC with metastasis (n=20) from OSCC without metastasis (n=34) using the salivary microbiome sequenced using the V3-V4 region of the 16S rRNA gene with a predictive accuracy of 86.3%. Finally, Zhou *et al.*<sup>90</sup> demonstrated that they could predict OSCC (n=47) from healthy controls (n=46) with 95.7% accuracy using RF modelling based on microbial signatures of saliva sequenced using the V4-V5 region of the 16S rRNA gene. They noted that the predictive accuracy increased as the training sample size increased and recommended that 80% be used as a minimum training set size.

I also examined, for the first time, the ability of the microbiome to predict low or high-risk dysplasia. Various combinations of clinical and patient features were included in multiple predictive LASSO models to ascertain the best combination(s) for final modelling. Clinical appearance featured in all models, ranking as the most important feature in LASSO, and the 9<sup>th</sup> most important in RF modelling. Combinations of age, gender, site and alcohol consumption, but not smoking status, also featured in the best models (**Table 88**), which predicted low-risk dysplasia (or absence of high-risk dysplasia) in up 79.4% with LASSO and 76.4% with random forest. However, further analysis using the z-scores of the top 50 microbial features from the RF SIAMCAT modelling remarkably produced a considerably more accurate algorithm.

Modelling using the top 50 microbial features could differentiate HROLK from LROLK with a sensitivity of 85.5% and specificity of 76.5%, while a combined model using both clinical/patient and microbial features could predict LROLK/HROLK with a sensitivity of 87.4% and specificity of 76.5%. Interestingly, although the PPVs were similar, the NPV of the combined model was higher than the microbial only model (82.2% versus 79.7%). The accuracy of both models was comparable.

I have therefore shown for the first time that the microbiome, with and without clinical/patient factors, in this case the clinical appearance of the OLK, has the ability to differentiate LROLK from HROLK with a sensitivity of more than 85%. Notably, the sensitivities, specificities, PPVs, NPVs and accuracies of both the microbial and microbial/clinical models were considerably higher than the model using clinical/patient features only. While non-homogenous OLKs should always be considered high risk, the addition of the microbial features to the modelling enabled more accurate separation of homogenous OLKs into LROLK and HROLK than modelling with only the clinical/patient features. Interestingly, using the microbial features alone provided similar results to the combined model, emphasising the value of the microbiome in differentiating LROLK from HROLK.

With refinement and further development of this modelling there is the potential to develop a screening tool, which could be used in primary care or low resource areas, to differentiate low risk from high risk OLK based on microbiome analysis of OLK swab samples. Microbial only modelling has the benefit that differentiation of homogenous from non-homogenous OLK, which requires clinical expertise and at times can be contentious, is not needed. Combined modelling, taking the clinical appearance of the OLK into consideration,

may have more potential for use in secondary or tertiary centres for monitoring of OLKs in the absence of clinical change. A shift in the microbiome of an OLK to a ‘high-risk profile’ could inform the need for repeat biopsy in these settings.

## 9.5 Conclusion

I have demonstrated that clinical and patient related factors, including the clinical appearance of the OLK and patient age, can be used in an algorithm to predict the risk of an OLK having low risk or high-risk dysplasia. However, more importantly, I have demonstrated for the first time that the OLK microbiome, either alone or when combined with clinical and patient-related factors, can also predict whether an OLK is likely to have low or high-risk dysplasia. Indeed, the OLK microbiome can make this prediction more accurately than clinical/patient factors alone. While these algorithms need external validation in another large data set, they have shown that it is possible to predict the likelihood of an OLK having low risk or high-risk dysplasia based on microbial features only or a combination of clinical and microbiological features which could be in future be used as a simple screening tool in primary care to guide urgency of referral to tertiary services. Furthermore, in time, these algorithms may also have the ability to be used, along with clinical monitoring, for non-invasive monitoring of OLKs in the absence of clinical change.

In summary:

- It was shown that microbial colonisation patterns could discriminate low risk from high-risk OLK.

## 10 Summary and Conclusions

The primary aims of this study were to establish if oral leukoplakia (OLK) was associated with changes in the microbiome and whether these changes differed with differing degrees of dysplasia such that they could be used to discriminate OLKs at lower or higher risk of malignant transformation. I also aimed to establish how other factors such as smoking and oral site affect the microbiome of healthy mucosa and of OLK. Finally, I wished to establish which clinical and patient factors affect the degree of dysplasia of OLK found on histological analysis.

It was estimated that a sample size of 197 OLK was needed to examine the association of clinical and patient factors with the degree of dysplasia, while a sample size of 212 OLK was needed to adequately investigate the association of any microbiome changes with the degree of dysplasia (if graded as none, mild, moderate and severe). Patients were recruited and sampled, and the samples then processed as detailed in Chapter 4.

### 10.1 Influence of clinical and patient-related factors on the degree of dysplasia

Tobacco smoking and alcohol consumption are known to be the most significant risk factors for the development of OLK in Western populations<sup>117-121</sup>. 81.2% of my patient cohort reported either current (52.7%) or past (28.5%) tobacco use, while 79% consumed alcohol, including almost 14% who consumed more than 20 units of alcohol per week. Males and smokers were more likely to engage in heavy alcohol consumption, which is concerning given that these cohorts are also known to have higher incidences of OLK and OSCC<sup>2,5,118</sup>. The association between smoking and heavy alcohol consumption, particularly amongst

males, needs be borne in mind when assessing patients' risks for OLK and OSCC. It also validates the idea of targeted public health messaging to specifically encourage these higher risk groups to attend for regular dental review with the aim of identifying OLK, so risk factors can be addressed, and patients referred to tertiary care.

The most common site for OLK was the floor of the mouth/ventral tongue, followed by the buccal mucosa and lateral border of the tongue. Although worldwide, the buccal mucosa is considered the most common site for OLK<sup>3,129,140</sup>, populations with high proportions of smokers tend to have higher incidences of floor of mouth/ventral tongue OLKs<sup>129,227,228</sup>. With 81.2% of my patient cohort being either current or former smokers it is not surprising that the FOM/VT was the most commonly identified site for OLK. Regression analysis confirmed the association of OLK site with smoking habit, with current smokers six times more likely, and former smokers nine times more likely, to have OLKs on the FOM/VT. It should therefore be emphasised when teaching mucosal examination technique to dentists, hygienists, dental science and hygiene students and other health care professionals, that this site needs to be carefully examined in all patients, but particularly in those with a history of smoking.

It is accepted that the degree of dysplasia on biopsy is the most accurate predictor of malignant transformation risk<sup>3,127,131,134,136</sup>, while OLKs on the floor of mouth or tongue<sup>3,40,129,130,133–136,139,140</sup> and with areas of redness or ulceration<sup>3,128–132,136,138,141</sup> are also at increased risk. My analysis agreed with this, with floor of mouth/tongue OLKs and non-homogenous and ulcerated OLKs significantly more likely to be severely dysplastic. 84% of the OLKs with severe dysplasia (42 of 50 OLKs with severe dysplasia) were located on the tongue or floor of mouth compared with just 6% on the buccal mucosa, 4% on the



alveolus/gingiva, and 6% on the palate. Regression analysis confirmed this association, demonstrating that OLKs on the floor of mouth or tongue had a 51.3 times increased relative risk of having severe dysplasia compared with BM OLKs. Non-homogeneity was also associated with an increased risk of severe dysplasia; 66.7% of NHOLKs and 85.7% of ulcerated OLKs showed severe dysplasia on biopsy compared with 12.9% of homogenous OLKs, while only 2.6% of NHOLKs but 44.1% of homogenous OLKs were mildly dysplastic. Finally, OLKs on the lateral tongue and floor of mouth/ventral tongue were 4.8 times and 3.3 times respectively more likely to be non-homogenous, which is consistent with my data on the frequency of severe dysplasia at these sites. These data add to the literature on OLK, the malignant transformation of which remains poorly understood, and should be considered when risk-assessing patients with OLK. These associations should also inform dental education programmes to ensure prompt and adequately detailed referrals to oral medicine or oral and maxillofacial surgery units of patients with OLK to ensure that those with OLKs on the tongue or FOM or with areas of non-homogeneity receive priority appointments in tertiary care.

## 10.2 Microbiome of clinically normal mucosa and the influence of patient factors on microbial colonisation

Almost all variables measured had some influence on the oral microbiome, the most significant of which were oral site and smoking habit. Poor oral hygiene was associated with increased abundance of several plaque and periodontal disease-associated species on healthy mucosal surfaces, while brushing once or less per day was associated with reduced abundance of several commensals. Tooth number also affected the microbiome of mucosal surfaces. Missing more than 15 teeth was associated with significantly reduced abundance

of several plaque-associated species, reflecting the lack of availability of tooth surfaces for plaque colonisers. The same trend of reduced abundance of plaque-associated species, including some of the same species, was found in complete and partial denture wearers. On the other hand, the presence of more teeth (missing five to 15 teeth group) was associated with increased abundance of several early plaque-associated species on mucosal surfaces.

These findings concur with the published literature that poor oral hygiene, reduced toothbrushing and periodontal disease are characterised by a switch to a Gram-negative, anaerobic, and pro-inflammatory microbiome<sup>256,257</sup>. Epidemiological studies and several meta-analyses have found associations between poor oral hygiene, tooth loss and periodontal disease and OSCC<sup>29,44–53</sup>. There are several proposed theories explaining these associations, including smoking being a risk factor for both periodontal disease and OSCC. However, it is also possible that the microbial dysbiosis and resultant pro-inflammatory environment associated with poor oral hygiene, tooth loss and periodontal disease may play a role in carcinogenesis, inflammation being recognised as a fundamental component of malignant processes<sup>180</sup>. It is therefore also possible that this switch to a more Gram-negative, anaerobic microbiome could be involved in the pathogenesis of OLK, the most common recognised precursor of OSCC.

Interestingly age and gender were also found to influence the oral microbiome, with those in the over 60 years age group having increased abundances of an *Actinomyces* and *Streptococcus* species and loss of a *Veillonella* species. Surprisingly, increased abundances of several periodontal pathogens were found in men, independent of smoking, providing further support for a potential association between microbial dysbiosis and OLK and OSCC, both of which are known to be more prevalent in males.

Site-specificity of the oral microbiome has been reported by several authors<sup>207,238,240,241</sup> and my findings concur that different mucosal surfaces represent distinct microbial niches. Dorsal and lateral tongue sites were characterised by increased abundances of *Rothia*, *Veillonella* and *Actinomyces* spp., and species level variance in streptococcal abundances. Palatal sites were characterised particularly by increased abundances of *Prevotella* spp., along with several *Actinomyces* and *Streptococcus* species. Finally, buccal mucosa and floor of mouth/ventral tongue sites displayed remarkably similar microbiomes, both dominated by streptococci.

Previous culture-based studies have shown that the tongue is the main habitat of *Rothia mucilaginosa*<sup>258</sup> while a more recent NGS study found *Rothia* to be enriched on tongue OLKs<sup>61</sup>. *Rothia* is a potent producer of the carcinogen acetaldehyde<sup>176</sup> and so its presence on clinically normal tongue mucosa raises the question as to whether it may play a role in the pathogenesis of OLK. The lateral tongue is one of the most common sites for OLK and considered a high-risk site for malignant transformation<sup>3,40,129,130,133–136,139,140</sup>. I have also shown that dorsal/lateral tongue OLKs have a 38.6 times increased relative risk of being severely dysplastic compared with buccal mucosa OLKs. The presence of an ACH-producing organism at this site may therefore be relevant with regards to the development and progression of OLK and warrants further investigation.

Smoking had arguably the most significant effect on the oral microbiome. There are many studies documenting the influence of smoking on oral bacteria, notably its profound inhibitory effect on *Neisseria* and its potentiation of a switch to a facultatively anaerobic or frankly anaerobic Gram-negative heavy microbiome, rich in periodontal

pathogens<sup>59,60,62,63,65,68–71</sup>, similar to that of OSCC<sup>9,77,79,80,82–84,87–91,94–104,110,111,113</sup>. My findings agree, with current smoking primarily associated with reduced abundance of *Neisseria* spp. and *Streptococcus* spp., along with increased abundance of plaque and periodontal disease-associated pathogens of genera *Prevotella*, *Actinomyces*, *Porphyromonas*, *Fusobacterium*, *Tannerella*, *Parvimonas*, *Filifactor*, *Clostridiales* and *Peptostreptococcus*. Interestingly, smoking appeared to affect palatal colonisation more than other sites, with ten species showing changes common to both current smoking and palatal site. Smoking is the most significant risk factor for OLK and OSCC and while tobacco contains multiple carcinogens, as mentioned, smoking is also associated with a microbial shift similar to that of periodontal disease and OSCC, again posing the question as to the role of the microbiome in both OLK and OSCC. If this changing microbial profile does in fact contribute to carcinogenesis, it is as yet unclear as to whether it is the pro-inflammatory dysbiosis or a specific organism that is involved. *Fusobacterium nucleatum*, associated with smoking, periodontal disease, OLK and OSCC has been proposed as a potential oral carcinogen and has been shown to induce carcinogenesis in a mouse model through its FadA adhesin facilitating epithelial change. However, this species is not always isolated from OLK or OSCC and, although the trends of microbial change in OLK and OSCC are similar between studies, the specific species and genera differ. It is likely therefore that if the oral microbiome is involved in the pathogenesis of OLK and OSCC, it is a multi- rather than single microbe effect.

### 10.3 The microbiome of OLK

To date only four studies have investigated the microbiome of OLK or OPMD through swab sampling, the largest of which sampled 36 OLKs and was the only study to adjust for potential confounding variables<sup>61,79,81,152</sup>. There is general agreement that the microbiome of

OLK differs from that of healthy mucosa, and although the studies differ as to what the differences are, there appears to be a trend towards increased abundance of *Fusobacterium* spp. and reduced abundance of *Streptococcus* spp. associated with OLK. This study is by far the largest to date to use direct swab sampling of well-characterised OLKs and contralateral normal sites to investigate the microbiome of OLK. OLKs consistently demonstrated increased abundance of five taxa, *Bergeyella* sp. HMT322, *Leptotrichia* sp. HMT215, *Streptococcus australis infantis* clade 431, *Gemella morbillorum*, and *Leptotrichia* sp. HMT392, even when adjusting for multiple variables. Several other taxa were associated with OLK when other variables such as missing teeth, smoking, and site were considered (**Table 58**).

Given the significance of site on the oral microbiome it was unsurprising that several of the OLK-associated species also showed site-specific changes in the patient cohort. *Bergeyella* sp. HMT322 was significantly enriched at floor of mouth/ventral tongue sites, which may relate to its association with high-risk dysplasia, while *Leptotrichia* sp. HMT215 was significantly enriched at palatal sites. Two other OLK-associated taxa, *Gemella morbillorum* and *Leptotrichia* sp. HMT392, showed significantly reduced abundance at palatal and dorsum/lateral border tongue, and dorsum/lateral border tongue sites respectively.

Interestingly, both *Bergeyella* sp. HMT322 and *Gemella morbillorum* displayed reduced abundance associated with smoking on the analysis of clinically normal mucosa (**Chapter 6**) confirming that their association with OLK is independent of smoking and suggesting that there are other, perhaps OLK-specific, factors driving their enrichment on OLKs. *Bergeyella* HMT322 has recently been associated with oesophageal squamous cell

carcinoma<sup>247</sup> and *Gemella morbillorum* with colorectal carcinoma<sup>244,245</sup> and so it is possible that these bacteria possess as yet unidentified adherence factors which facilitate binding to the unique surfaces of OLK and cancerous tissues.

*Streptococcus australis infantis* clade 431 was the only OLK-associated taxa influenced by oral hygiene, showing increased abundance associated with fair but not poor oral hygiene. However, this taxa remained associated with OLK even after adjusting for oral hygiene. Only one other small biopsy based OLK/OSCC microbiome study<sup>93</sup>, which included nine homogenous OLK, has to date identified an increase in a streptococcal species, in OLK. Most OLK and OSCC microbiome studies describe reduced abundance of *Streptococcus* spp. associated with OLK and OSCC<sup>9,79,81–83,87,88,90,92–98,100,101,103,104,152–154</sup>, although it should be noted that most previous studies only carried out analysis to genera level, and with which my genera level analysis concurs. Interestingly and perhaps of relevance, *Streptococcus infantis* was recently identified in the metagenome of breast cancer<sup>22</sup>.

While I have shown that the microbial community of OLK differs to that of normal mucosa, it should be noted that this study was not designed to investigate causation. The epithelial surface of OLK differs clinically (white with or without redness/ulceration), histologically (showing keratinisation and dysplasia to varying degrees) and biochemically (demonstrating loss of E-cadherin, and upregulation of N-cadherin, vimentin and SNAIL)<sup>157,158</sup> from normal mucosa. It could therefore be hypothesised that these surface changes facilitate a different pattern of microbial colonisation through as yet unrecognised adhesins. Further research in this area is needed in order to establish cause and effect with regards to the relationship between the oral microbiome and OLK.

## 10.4 The microbiome of OLKs of different degrees of dysplasia and OSCC

OLKs were divided into low (no and mild dysplasia) and high risk (moderate and severe dysplasia) for further analysis to ascertain if there were changes in the microbiome that were specific to different degrees of dysplasia. To my knowledge, this is the first time that this level of analysis on the microbiome of OLK has been carried out. Low-risk OLKs were consistently associated with significantly increased abundances of *Gemella morbillorum* and *Mogibacterium diversum*, even after adjusting for multiple variables. Several other taxa were also associated with LROLK after adjusting for smoking and number of missing teeth (**Table 84**). High-risk OLKs were associated with significantly increased abundances of three taxa, *Bergeyella* sp. HMT322, *Streptococcus australis infantis* clade 431 and *Leptotrichia* sp. HMT215. Further analysis, with OLKs divided according to the WHO classification of dysplasia (no, mild, moderate, severe dysplasia)<sup>114</sup>, demonstrated that of these three taxa, *Bergeyella* sp. HMT322 was specifically associated with severe dysplasia, *Streptococcus australis infantis* clade 431 was associated with both moderate and severe dysplasia and *Leptotrichia* sp. HMT215 with moderate dysplasia. As mentioned, *Streptococcus* spp. tend to show reduced abundance associated with OLK and OSCC, but I have identified for the first time a streptococcal species that appears to be specific to high-risk OLK.

Finally, in agreement with previous studies that species from the genera *Parvimonas* and particularly *Fusobacterium* are frequently associated with OSCC<sup>9,77,79,80,82,83,87–90,96,97,100,103,111</sup>, *Fusobacterium nucleatum* subsp. *polymorphum hwasookii* and *Parvimonas* HMT393 showed significantly increased abundance associated with OSCC after adjusting for smoking, missing teeth and mucosal site, with *Parvimonas* HMT393 also being associated with OSCC after adjusting for denture wear and brushing habit. As discussed, the OSCC

findings are somewhat limited as there were only 10 OSCCs in the study, all of which presented clinically as non-homogenous (some of which were ulcerated) OLKs.

## 10.5 Predictive modelling

Random forest predictive modelling using clinical and patient factors could predict high risk dysplasia with an overall sensitivity of 76.4% and specificity of 66% using a predictive algorithm which included the clinical appearance of the OLK (homogenous/non-homogenous), patient age, site of OLK, brushing habit and mouthwash use. The most significant finding from this modelling, however, was that for homogenous OLKs, being aged under 40 years was predictive of LROLK in 90% of cases. To our knowledge this is the first time such an algorithm has been proposed. Validation of such an algorithm would have significant benefit in primary care, where OLKs are typically identified before being referred to secondary or tertiary care, to guide urgency of referrals for those at highest risk. Other variables such as tooth number and oral hygiene could also be considered in future algorithm development.

Finally, predictive modelling using the top 50 microbial features from a preliminary modelling analysis using SIAMCAT<sup>219</sup>, both with and without clinical/patient features, could differentiate LROLK from HROLK with sensitivities of 87.4% (with clinical/patient features) and 85.5% (without clinical/patient features) and a specificity of 76.5% (both with and without clinical/patient features). Although a biopsy is always needed to establish a histological diagnosis, my findings demonstrate for the first time that there is the potential to develop a simple screening tool based on the microbial analysis of OLK swab samples which could differentiate low risk from high-risk OLK more accurately than clinical/patient



factors alone. A tool based on microbial-only modelling would have significant benefit in primary care as differentiation of homogenous from non-homogenous OLK, which requires clinical expertise, would not be needed. Moreover, while a change in the clinical appearance of an OLK will usually necessitate a repeat biopsy, combined modelling, using both microbial and clinical/patient features, has the potential in future to be used in secondary or tertiary care for monitoring OLKs without clinical change, where a shift to ‘high-risk’ in a model could inform the need for repeat biopsy.

## 10.6 The future

Amplification and sequencing of a variable region, in this case the V1-V3 region, of the 16S rRNA gene, an established method of bacterial identification and still the method of choice in current literature, was used in this study. However, next generation sequencing methods are constantly evolving and metagenomic methods, which sequence the entire genome and so provide taxonomy along with the metabolic pathways and functional profile of the entire microbial community sampled, are becoming available<sup>259,260</sup>. Although these methods are currently limited by cost and accessibility, it is reasonable to expect that they will supersede marker gene sequencing in time. Incorporating metagenomic sequencing into our methods could improve taxonomic identification, while examining the functional and metabolic profiles of those taxa showing increased abundance on OLK could provide insight into their role in the pathogenesis of OLK, which is purely speculative at present.

Very few studies to date have examined the mycobiome of oral potentially malignant disorders<sup>103,190,194</sup> and further investigation of its role is warranted, given that *Candida*, frequently identified in OSCC, is known to produce acetaldehyde<sup>174</sup>. There are no reports

as yet on the virome of OLK. HPV is of particular interest in view of its established role in the aetiology of oropharyngeal cancer<sup>13</sup>, and a recent meta-analysis found that HPV DNA is detected in up to 27% of oral epithelial dysplasia<sup>261</sup>. Future studies should ideally examine the entire metagenome (bacteria, fungi and viruses) of OLK, which may help to further our understanding of the pathogenesis and drivers of malignant transformation of this common condition.

## 10.7 Limitations

Although this study is one of the largest to date on the microbiome of OLK and the first to adjust for multiple variables and examine differences in the microbiome associated with different degrees of dysplasia, it must be noted that all participants were patients of a single institution and so the findings may not be applicable to other geographical areas. Several authors have previously demonstrated that the oral microbiome varies between geographical regions even when differing diets and habits are considered<sup>262,263</sup>.

Secondly, I was unable to recruit adequately matched controls due to the higher prevalence of smoking, alcohol consumption, poor oral hygiene and tooth loss amongst the OLK cohort. Although I have shown that contralateral normal sites are appropriate controls for OLK microbiome studies as they ensure that all samples are matched for potential confounders, matched controls would negate any potential changes on contralateral normal sites due to the presence of OLK. Some of the patient cohort also had widespread changes with no available contralateral normal site with the result that there were more OLK than CLN sites and differences between the number of OLK and CLN samples at different oral sites.

Thirdly the time between swab sample and incisional biopsy varied amongst participants as the swab samples were carried out in Oral Medicine or Dysplasia clinics while most of the biopsies were carried out in a dedicated biopsy clinic. More than half (51.1%) of swab samples were carried out within one month of the biopsy, with a further 14.6% carried out within one to three months of the biopsy. Photographic comparisons of OLKs were made at the time of biopsy and swab sampling and if a significant change was noted a repeat swab was taken and this was used for microbiome analysis.

Finally, as in all other previous OLK microbiome studies to date, participants were sampled at a single point in time and so temporal variability was not accounted for. However, as mentioned, (**section 4.1**), it has been shown that the oral microbiome is generally stable over time and so single time point sampling is the accepted methodology for these types of research studies. Furthermore, by excluding factors likely to affect the oral microbiome temporarily but significantly, in particular the use of topical or systemic antibiotics or steroids in the 2 months prior to sampling or current respiratory tract infection, temporal variations should have been minimal.

## 10.8 Conclusion

I have conclusively shown that firstly the microbiome of OLK differs to that of clinically normal mucosa and secondly that OLKs with different degrees of dysplasia also differ in their microbial communities, such that these differences could potentially be used in future to differentiate OLKs with low and high-risk dysplasia. The development of OLK is associated with a distinct switch from an aerobic, streptococcal-dominated microbiome to an anaerobic/facultatively anaerobic one. Only two of the 11 taxa associated with OLK, and

two of the 19 taxa associated with LROLK, were aerobic. Although smoking can facilitate such a switch, the changes identified were independent of smoking. The microbial colonisation pattern of HROLK however, was different, with two of the three taxa with increased abundance in HROLK being aerobic. Perhaps significantly, one of these, *Bergeyella* HMT322, has recently been associated with oesophageal cancer<sup>247</sup>. *Streptococcus australis infantis* clade 431 was the other aerobe associated with HROLK, and although more detailed taxonomic identification, ideally through metagenomic sequencing, is needed to fully identify this species it is possible that this taxa is related to *Streptococcus infantis* which has been associated with breast cancer<sup>22</sup>. Further research is now needed to elucidate the significance of these changes and whether they may play a role in the malignant transformation of OLK. From a clinical point of view, awareness of the association between smoking and heavy alcohol consumption, particularly in males, should guide lifestyle advice, while non-homogenous white patches, particularly if ulcerated, or located on the tongue or floor of mouth should be considered high risk for severe dysplasia and prioritised for investigation and definitive management. Finally, I have demonstrated that predictive modelling using the microbiome, both with and without clinical and patient factors, can differentiate low risk from high-risk dysplasia. Microbial modelling therefore has the potential to be developed into a simple screening tool for use in primary care or low resource settings. This modelling may also have the ability to be used, along with clinical monitoring, in tertiary care for non-invasive monitoring of OLKs in the absence of clinical change, where a shift in the microbiome to that of a ‘high-risk profile’ could direct the need for repeat biopsy.

# 11 Bibliography

1. Warnakulasuriya, S., Johnson, Newell. W. & Waal, I. V. D. Nomenclature and classification of potentially malignant disorders of the oral mucosa. *J. Oral Pathol. Med.* **36**, 575–580 (2007).
2. Mello, F. W. *et al.* Prevalence of oral potentially malignant disorders: a systematic review and meta-analysis. *J Oral Pathology Medicine Official Publ Int Assoc Oral Pathologists Am Acad Oral Pathology* **47**, 633–640 (2018).
3. Aguirre-Urizar, J. M., Mendoza, I. L. de & Warnakulasuriya, S. Malignant transformation of oral leukoplakia: Systematic review and meta-analysis of the last 5 years. *Oral Dis* (2021) doi:10.1111/odi.13810.
4. Bray, F. *et al.* Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *Ca Cancer J Clin* **68**, 394–424 (2018).
5. Scully, C. & Bagan, J. Oral squamous cell carcinoma: overview of current understanding of aetiopathogenesis and clinical implications. *Oral Dis* **15**, 388–399 (2009).
6. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clin.* **71**, 209–249 (2021).
7. Warnakulasuriya, S. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol* **45**, 309–316 (2009).
8. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci.* **87**, 4576–4579 (1990).
9. Zhu, H. *et al.* Convergent dysbiosis of upper aerodigestive microbiota between patients with esophageal and oral cavity squamous cell carcinoma. *Int. J. Cancer* (2023) doi:10.1002/ijc.34460.
10. Martel, C. de, Georges, D., Bray, F., Ferlay, J. & Clifford, G. M. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Global Heal* **8**, e180–e190 (2019).
11. Martel, C. de *et al.* Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncol* **13**, 607–615 (2012).
12. Garrett, W. S. Cancer and the microbiota. *Science* **348**, 80–86 (2015).

13. Carbone, A. *et al.* Biological agents. Volume 100 B. A review of human carcinogens. *Iarc Monogr Evaluation Carcinog Risks Humans World Heal Organization Int Agency Res Cancer* **100**, 1–441 (2012).
14. McCoy, A. N. *et al.* Fusobacterium Is Associated with Colorectal Adenomas. *Plos One* **8**, e53653 (2013).
15. Castellarin, M. *et al.* Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res* **22**, 299–306 (2012).
16. Kostic, A. D. *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res* **22**, 292–298 (2012).
17. Mima, K. *et al.* Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. *Gut* **65**, 1973 (2016).
18. Bullman, S. *et al.* Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. *Science* **358**, eaal5240 (2017).
19. Zhang, N. *et al.* Clinical Significance of Fusobacterium nucleatum Infection and Regulatory T Cell Enrichment in Esophageal Squamous Cell Carcinoma. *Pathol. Oncol. Res.* **27**, 1609846 (2021).
20. Yu, T. *et al.* Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell* **170**, 548–563.e16 (2017).
21. Parhi, L. *et al.* Breast cancer colonization by Fusobacterium nucleatum accelerates tumor growth and metastatic progression. *Nat. Commun.* **11**, 3259 (2020).
22. Nejman, D. *et al.* The human tumor microbiome is composed of tumor type–specific intracellular bacteria. *Science* **368**, 973–980 (2020).
23. Dewhirst, F. E. *et al.* The human oral microbiome. *J Bacteriol* **192**, 5002–17 (2010).
24. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–14 (2012).
25. Health, N. I. of. Surveillance, Epidemiology and End Results Programme, NIH. <https://seer.cancer.gov/statfacts/html>.
26. Hashibe, M. *et al.* Interaction between Tobacco and Alcohol Use and the Risk of Head and Neck Cancer: Pooled Analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiology Prev Biomarkers* **18**, 541–550 (2009).
27. Macfarlane, G. J. *et al.* Alcohol, tobacco, diet and the risk of oral cancer: a pooled analysis of three case-control studies. *European J Cancer Part B Oral Oncol* **31**, 181–187 (1995).
28. Vecchia, C. L. *et al.* Epidemiology and prevention of oral cancer. *Oral Oncol* **33**, 302–312 (1997).

29. Zheng, T. *et al.* Dentition, oral hygiene, and risk of oral cancer: a case-control study in Beijing, People's Republic of China. *Cancer Cause Control* **1**, 235–241 (1990).
30. Radoï, L. *et al.* Tobacco smoking, alcohol drinking and risk of oral cavity cancer by subsite. *Eur J Cancer Prev* **22**, 268–276 (2013).
31. Llewellyn, C. D., Linklater, K., Bell, J., Johnson, N. W. & Warnakulasuriya, K. A. A. S. Squamous cell carcinoma of the oral cavity in patients aged 45 years and under: a descriptive analysis of 116 cases diagnosed in the South East of England from 1990 to 1997. *Oral Oncol* **39**, 106–114 (2003).
32. Montero, P. H. *et al.* Changing Trends in Smoking and Alcohol Consumption in Patients With Oral Cancer Treated at Memorial Sloan-Kettering Cancer Center From 1985 to 2009. *Archives Otolaryngology Head Neck Surg* **138**, 817–822 (2012).
33. Centers for Disease Control and Prevention. Data and Statistics. Atlanta (GA): Office of Smoking and Health, National Center for Chronic Disease Prevention and Health Promotion; 2017. [https://www.cdc.gov/tobacco/data\\_statistics/index.htm](https://www.cdc.gov/tobacco/data_statistics/index.htm) (2021).
34. Tota, J. E. *et al.* Rising incidence of oral tongue cancer among white men and women in the United States, 1973–2012. *Oral Oncol* **67**, 146–152 (2017).
35. Koch, W. M., Lango, M., Sewell, D., Zahurak, M. & Sidransky, D. Head and Neck Cancer in CNonsmokers: A Distinct Clinical and Molecular Entity. *Laryngoscope* **109**, 1544–1551 (1999).
36. Durr, M. L., Li, D. & Wang, S. J. Oral cavity squamous cell carcinoma in never smokers: Analysis of clinicopathologic characteristics and survival. *Am J Otolaryng* **34**, 388–393 (2013).
37. Dediol, E. *et al.* HPV prevalence and p16INKa overexpression in non-smoking non-drinking oral cavity cancer patients. *Oral Dis.* **22**, 517–522 (2016).
38. Li, R. *et al.* Clinical, genomic, and metagenomic characterization of oral tongue squamous cell carcinoma in patients who do not smoke. *Head Neck* **37**, 1642–1649 (2015).
39. Deneuve, S. *et al.* Oral tongue squamous cell carcinomas in young patients according to their smoking status: a GETTEC study. *Eur Arch Oto-rhino-l* 1–10 (2021) doi:10.1007/s00405-021-06793-7.
40. Rock, L. D. *et al.* Characterization of epithelial oral dysplasia in non-smokers: First steps towards precision medicine. *Oral Oncol* **78**, 119–125 (2018).
41. Rodriguez, T. *et al.* Risk factors for oral and pharyngeal cancer in young adults. *Oral Oncol* **40**, 207–213 (2004).
42. Llewellyn, C. D., Linklater, K., Bell, J., Johnson, N. W. & Warnakulasuriya, S. An analysis of risk factors for oral cancer in young people: a case-control study. *Oral Oncol* **40**, 304–313 (2004).

43. Macfarlane, T. V. *et al.* The aetiology of upper aerodigestive tract cancers among young adults in Europe: the ARCAGE study. *Cancer Cause Control* **21**, 2213–2221 (2010).
44. Gupta, B., Bray, F., Kumar, N. & Johnson, N. W. Associations between oral hygiene habits, diet, tobacco and alcohol and risk of oral cancer: A case–control study from India. *Cancer Epidemiol* **51**, 7–14 (2017).
45. Yao, Q.-W., Zhou, D.-S., Peng, H.-J., Ji, P. & Liu, D.-S. Association of periodontal disease with oral cancer: a meta-analysis. *Tumor Biol* **35**, 7073–7077 (2014).
46. Sato, F. *et al.* Inverse association between toothbrushing and upper aerodigestive tract cancer risk in a Japanese population. *Head Neck* **33**, 1628–1637 (2011).
47. Moreno-López, L. A. *et al.* Risk of oral cancer associated with tobacco smoking, alcohol consumption and oral hygiene: a case-control study in Madrid, Spain. *Oral Oncol* **36**, 170–174 (2000).
48. Chang, J. S. *et al.* Investigating the association between oral hygiene and head and neck cancer. *Oral Oncol* **49**, 1010–1017 (2013).
49. Maier, H., Zöller, J., Herrmann, A., Kreiss, M. & Heller, W.-D. Dental Status and Oral Hygiene in Patients with Head and Neck Cancer. *Otolaryngology Head Neck Surg* **108**, 655–661 (1993).
50. Tezal, M. *et al.* Chronic periodontitis and the incidence of head and neck squamous cell carcinoma. *Cancer Epidemiology Biomarkers Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* **18**, 2406–12 (2009).
51. Ahrens, W. *et al.* Oral health, dental care and mouthwash associated with upper aerodigestive tract cancer risk in Europe: The ARCAGE study. *Oral Oncol* **50**, 616–625 (2014).
52. Javed, F. & Warnakulasuriya, S. Is there a relationship between periodontal disease and oral cancer? A systematic review of currently available evidence. *Crit Rev Oncol Hemat* **97**, 197–205 (2016).
53. Hashim, D. *et al.* The role of oral hygiene in head and neck cancer: results from International Head and Neck Cancer Epidemiology (INHANCE) consortium. *Ann Oncol* **27**, 1619–1625 (2016).
54. Tasoulas, J. *et al.* Poor oral health influences head and neck cancer patient survival: an International Head and Neck Cancer Epidemiology Consortium pooled analysis. *JNCI: J. Natl. Cancer Inst.* djad156 (2023) doi:10.1093/jnci/djad156.
55. Liu, J. *et al.* Tooth count, untreated caries and mortality in US adults: a population-based cohort study. *Int. J. Epidemiology* **51**, 1291–1303 (2022).
56. Baek, O., Zhu, W., Kim, H. C. & Lee, S.-W. Effects of nicotine on the growth and protein expression of *Porphyromonas gingivalis*. *J Microbiol* **50**, 143–148 (2012).



57. Bardell, D. Viability of six species of normal oropharyngeal bacteria after exposure to cigarette smoke in vitro. *Microbios* **32**, 7–13 (1981).
58. Ertel, A., Eng, R. & Smith, S. M. The Differential Effect of Cigarette Smoke on the Growth of Bacteria Found in Humans. *Chest* **100**, 628–630 (1991).
59. Charlson, E. S. *et al.* Disordered Microbial Communities in the Upper Respiratory Tract of Cigarette Smokers. *Plos One* **5**, e15216 (2010).
60. Thomas, A. M. *et al.* Alcohol and tobacco consumption affects bacterial richness in oral cavity mucosa biofilms. *Bmc Microbiol* **14**, 250 (2014).
61. Amer, A., Galvin, S., Healy, C. M. & Moran, G. P. The Microbiome of Potentially Malignant Oral Leukoplakia Exhibits Enrichment for *Fusobacterium*, *Leptotrichia*, *Campylobacter*, and *Rothia* Species. *Front Microbiol* **8**, 2391 (2017).
62. Gopinath, D. *et al.* Compositional profile of mucosal bacteriome of smokers and smokeless tobacco users. *Clin Oral Invest* 1–10 (2021) doi:10.1007/s00784-021-04137-7.
63. Morris, A. *et al.* Comparison of the Respiratory Microbiome in Healthy Nonsmokers and Smokers. *Am J Resp Crit Care* **187**, 1067–1075 (2013).
64. Belstrøm, D. *et al.* Bacterial profiles of saliva in relation to diet, lifestyle factors, and socioeconomic status. *J Oral Microbiol* **6**, 23609 (2014).
65. Wu, J. *et al.* Cigarette smoking and the oral microbiome in a large study of American adults. *Isme J* **10**, ismej201637 (2016).
66. Calle, E. E. *et al.* The American Cancer Society Cancer Prevention Study II Nutrition Cohort. *Cancer* **94**, 2490–2501 (2002).
67. Hayes, R. B. *et al.* Etiologic and early marker studies in the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Control Clin Trials* **21**, 349S-355S (2000).
68. Hsiao, J.-R. *et al.* The interplay between oral microbiome, lifestyle factors and genetic polymorphisms in the risk of oral squamous cell carcinoma. *Carcinogenesis* **39**, 778–787 (2018).
69. Wirth, R. *et al.* A case study of salivary microbiome in smokers and non-smokers in Hungary: analysis by shotgun metagenome sequencing. *J Oral Microbiol* **12**, 1773067 (2020).
70. Moon, J. -H., Lee, J. -H. & Lee, J. -Y. Subgingival microbiome in smokers and non-smokers in Korean chronic periodontitis patients. *Mol Oral Microbiol* **30**, 227–241 (2015).
71. Mason, M. R. *et al.* The subgingival microbiome of clinically healthy current and never smokers. *Isme J* **9**, 268–272 (2015).
72. Fan, X. *et al.* Drinking alcohol is associated with variation in the human oral microbiome in a large study of American adults. *Microbiome* **6**, 59 (2018).

73. Priyanka, K., Sudhir, K. M., Reddy, V. C. S., Kumar, R. K. & Srinivasulu, G. Impact of Alcohol Dependency on Oral Health – A Cross-sectional Comparative Study. *J Clin Diagnostic Res* **11**, ZC43–ZC46 (2017).
74. Manicone, P. F. *et al.* Dental health in patients affected by alcohol use disorders: a cross-sectional study. *Eur Rev Med Pharmacol* **21**, 5021–5027 (2017).
75. Lages, E. J. P. *et al.* Alcohol Consumption and Periodontitis: Quantification of Periodontal Pathogens and Cytokines. *J Periodontol* **86**, 1058–1068 (2015).
76. Bolz, J., Dosá, E., Schubert, J. & Eckert, A. W. Bacterial colonization of microbial biofilms in oral squamous cell carcinoma. *Clin Oral Invest* **18**, 409–414 (2014).
77. Nagy, K. N., Sonkodi, I., Szöke, I., Nagy, E. & Newman, H. N. The microflora associated with human oral carcinomas. *Oral Oncol* **34**, 304–308 (1998).
78. Hamady, M. & Knight, R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res* **19**, 1141–1152 (2009).
79. Schmidt, B. L. *et al.* Changes in Abundance of Oral Microbiota Associated with Oral Cancer. *Plos One* **9**, e98741 (2014).
80. Zhao, H. *et al.* Variations in oral microbiota associated with oral cancer. *Sci Rep-uk* **7**, 11773 (2017).
81. Mok, S. F. *et al.* The oral microbiome community variations associated with normal, potentially malignant disorders and malignant lesions of the oral cavity. *Malays J Pathology* **39**, 1–15 (2017).
82. Zhang, L., Liu, Y., Zheng, H. J. & Zhang, C. P. The Oral Microbiota May Have Influence on Oral Cancer. *Front Cell Infect Mi* **9**, 476 (2019).
83. Su, S.-C. *et al.* Oral microbial dysbiosis and its performance in predicting oral cancer. *Carcinogenesis* (2020) doi:10.1093/carcin/bgaa062.
84. Pushalkar, S. *et al.* Microbial diversity in saliva of oral squamous cell carcinoma. *Fems Immunol Medical Microbiol* **61**, 269–277 (2011).
85. Pushalkar, S. *et al.* Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. *Bmc Microbiol* **12**, 144 (2012).
86. Perera, M. *et al.* A dysbiotic mycobiome dominated by *Candida albicans* is identified within oral squamous-cell carcinomas. *J Oral Microbiol* **9**, 1385369 (2017).
87. Perera, M. *et al.* Inflammatory Bacteriome and Oral Squamous Cell Carcinoma. *J Dent Res* **97**, 725–732 (2018).
88. Al-hebshi, N. N. *et al.* Inflammatory bacteriome featuring *Fusobacterium nucleatum* and *Pseudomonas aeruginosa* identified in association with oral squamous cell carcinoma. *Scientific Reports* **7**, 1834 (2017).

89. Zhou, J. *et al.* Signatures of Mucosal Microbiome in Oral Squamous Cell Carcinoma Identified Using a Random Forest Model. *Cancer Management Res* **12**, 5353–5363 (2020).
90. Zhou, X. *et al.* The Clinical Potential of Oral Microbiota as a Screening Tool for Oral Squamous Cell Carcinomas. *Front Cell Infect Mi* **11**, 728933 (2021).
91. Sarkar, P. *et al.* Dysbiosis of Oral Microbiota During Oral Squamous Cell Carcinoma Development. *Frontiers Oncol* **11**, 614448 (2021).
92. Ye, P. *et al.* Microbial community alteration in tongue squamous cell carcinoma. *Appl Microbiol Biotechnol* **105**, 8457–8467 (2021).
93. Herreros-Pomares, A. *et al.* On the Oral Microbiome of Oral Potentially Malignant and Malignant Disorders: Dysbiosis, Loss of Diversity, and Pathogens Enrichment. *Int J Mol Sci* **24**, 3466 (2023).
94. Hu, X., Zhang, Q., Hua, H. & Chen, F. Changes in the salivary microbiota of oral leukoplakia and oral cancer. *Oral Oncol* **56**, e6–e8 (2016).
95. Hashimoto, K. *et al.* Changes in oral microbial profiles associated with oral squamous cell carcinoma vs leukoplakia. *J Investigative Clin Dent* **10**, e12445 (2019).
96. Ganly, I. *et al.* Periodontal pathogens are a risk factor of oral cavity squamous cell carcinoma, independent of tobacco and alcohol and human papillomavirus. *Int J Cancer* **145**, 775–784 (2019).
97. Ganly, I. *et al.* Oral Microbiome in Nonsmoker Patients with Oral Cavity Squamous Cell Carcinoma, Defined by Metagenomic Shotgun Sequencing. *Cancers* **14**, 6096 (2022).
98. Gopinath, D. *et al.* Salivary bacterial shifts in oral leukoplakia resemble the dysbiotic oral cancer bacteriome. *J Oral Microbiol* **13**, 1857998 (2021).
99. Lee, W.-H. *et al.* Bacterial alterations in salivary microbiota and their association in oral cancer. *Sci Rep-uk* **7**, 16540 (2017).
100. Chen, J.-W. *et al.* Taxonomic and Functional Dysregulation in Salivary Microbiomes During Oral Carcinogenesis. *Front Cell Infect Mi* **11**, 663068 (2021).
101. Sawant, S., Dugad, J., Parikh, D., Srinivasan, S. & Singh, H. Identification & correlation of bacterial diversity in oral cancer and long-term tobacco chewers- A case-control pilot study. *J Med Microbiol* **70**, (2021).
102. Guerrero-Preston, R. *et al.* 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment. *Oncotarget* **7**, 51320–51334 (2016).
103. Heng, W. *et al.* Oral Bacteriome and Mycobiome across Stages of Oral Carcinogenesis. *Microbiol Spectr* **10**, e02737-22 (2022).

104. Yang, S.-F. *et al.* Compositional and functional variations of oral microbiota associated with the mutational changes in oral cancer. *Oral Oncol* **77**, 1–8 (2018).
105. Yu, X. *et al.* Microbial dysbiosis in oral squamous cell carcinoma: A systematic review and meta-analysis. *Heliyon* **9**, e13198 (2023).
106. Zhang, S. *et al.* Fusobacterium nucleatum promotes epithelial-mesenchymal transition through regulation of the lncRNA MIR4435-2HG/miR-296-5p/Akt2/SNAI1 signaling pathway. *Febs J* **287**, 4032–4047 (2020).
107. Hooper, S. J. *et al.* A molecular analysis of the bacteria present within oral squamous cell carcinoma. *J Med Microbiol* **56**, 1651–1659 (2007).
108. Hooper, S. J. *et al.* Viable Bacteria Present within Oral Squamous Cell Carcinoma Tissue. *J Clin Microbiol* **44**, 1719–1725 (2006).
109. Wade, W. G. & Prosdocimi, E. M. Profiling of Oral Bacterial Communities. *J Dent Res* **99**, 621–629 (2020).
110. Mager, D. *et al.* The salivary microbiota as a diagnostic indicator of oral cancer: A descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. *J Transl Med* **3**, 27 (2005).
111. Takahashi, Y. *et al.* Analysis of oral microbiota in Japanese oral cancer patients using 16S rRNA sequencing. *J Oral Biosci* (2019) doi:10.1016/j.job.2019.03.003.
112. Yang, C.-Y. *et al.* Oral Microbiota Community Dynamics Associated With Oral Squamous Cell Carcinoma Staging. *Frontiers in Microbiology* **9**, 862 (2018).
113. Hayes, R. B. *et al.* Association of Oral Microbiome With Risk for Incident Head and Neck Squamous Cell Cancer. *JAMA Oncology* (2018) doi:10.1001/jamaoncol.2017.4777.
114. El-Naggar, A. K., Chan, J. K. C., Takata, T., Grandis, J. R. & Slootweg, P. J. The fourth edition of the head and neck World Health Organization blue book: editors' perspectives. *Hum Pathol* **66**, 10–12 (2017).
115. Lee, J. *et al.* Human Primary Epithelial Cells Acquire an Epithelial-Mesenchymal-Transition Phenotype during Long-Term Infection by the Oral Opportunistic Pathogen, Porphyromonas gingivalis. *Front Cell Infect Mi* **7**, 493 (2017).
116. Warnakulasuriya, S. *et al.* Oral potentially malignant disorders: nomenclature and classification. *Oral Dis* (2020) doi:10.1111/odi.13704.
117. Hashibe, M. *et al.* Alcohol drinking, body mass index and the risk of oral leukoplakia in an Indian population. *Int J Cancer* **88**, 129–134 (2000).
118. Warnakulasuriya, S. Clinical features and presentation of oral potentially malignant disorders. *Oral Surg Oral Medicine Oral Pathology Oral Radiology* **125**, 582–590 (2018).

119. Gueiros, L. A., Leão, J. C., Fedele, S. & Porter, S. Risk factors and aetiopathogenesis of potentially premalignant oral epithelial lesions. *Oral Surg Oral Medicine Oral Pathology Oral Radiology* **125**, 603–611 (2018).
120. Morse, D. E. *et al.* Smoking and drinking in relation to oral epithelial dysplasia. *Cancer Epidemiology Biomarkers Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* **5**, 769–77 (1996).
121. Maserejian, N. N., Joshipura, K. J., Rosner, B. A., Giovannucci, E. & Zavras, A. I. Prospective Study of Alcohol Consumption and Risk of Oral Premalignant Lesions in Men. *Cancer Epidemiology Biomarkers Prev* **15**, 774–781 (2006).
122. Speight, P. M. *et al.* Interobserver agreement in dysplasia grading: toward an enhanced gold standard for clinical pathology trials. *Oral Surg Oral Medicine Oral Pathology Oral Radiology* **120**, 474–482.e2 (2015).
123. Odell, E., Kujan, O., Warnakulasuriya, S. & Sloan, P. Oral epithelial dysplasia: recognition, grading and clinical significance. *Oral Dis* (2021) doi:10.1111/odi.13993.
124. Küffer, R. & Lombardi, T. Premalignant lesions of the oral mucosa. A discussion about the place of oral intraepithelial neoplasia (OIN). *Oral Oncol* **38**, 125–130 (2002).
125. Kujan, O. *et al.* Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. *Oral Oncol* **42**, 987–993 (2006).
126. Reibel, J. *et al.* Oral potentially malignant disorders and oral epithelial dysplasia. in *WHO classification of head and neck tumours* vol. 9 112–113 (2017).
127. Iocca, O. *et al.* Potentially malignant disorders of the oral cavity and oral dysplasia: A systematic review and meta-analysis of malignant transformation rate by subtype. *Head Neck* **42**, 539–555 (2020).
128. Guan, J. *et al.* Malignant transformation rate of oral leukoplakia in the past 20 years: A systematic review and meta-analysis. *J Oral Pathol Med* (2023) doi:10.1111/jop.13440.
129. Warnakulasuriya, S. & Ariyawardana, A. Malignant transformation of oral leukoplakia: a systematic review of observational studies. *J Oral Pathol Med* **45**, 155–166 (2016).
130. Paglioni, M. de P. *et al.* Clinical predictors of malignant transformation and recurrence in oral potentially malignant disorders: a systematic review and meta-analysis. *Oral Surg Oral Medicine Oral Pathology Oral Radiology* (2022) doi:10.1016/j.oooo.2022.07.006.
131. Schepman, K. P., Meij, E. H. van der, Smeele, L. E. & Waal, I. van der. Malignant transformation of oral leukoplakia: a follow-up study of a hospital-based population of 166 patients with oral leukoplakia from The Netherlands. *Oral Oncol* **34**, 270–275 (1998).
132. Jr, S. S., Gorsky, M. & Lozada, D., Ms Francina. Oral leukoplakia and malignant transformation. A follow-up study of 257 patients. *Cancer* **53**, 563–568 (1984).

133. Ho, M. W. *et al.* The clinical determinants of malignant transformation in oral epithelial dysplasia. *Oral Oncol* **48**, 969–976 (2012).
134. Warnakulasuriya, S. *et al.* Factors predicting malignant transformation in oral potentially malignant disorders among patients accrued over a 10-year period in South East England. *J Oral Pathol Med* **40**, 677–683 (2011).
135. Cai, X. *et al.* Development and validation of a nomogram prediction model for malignant transformation of oral potentially malignant disorders. *Oral Oncol* **123**, 105619 (2021).
136. Liu, W. *et al.* Oral Cancer Development in Patients with Leukoplakia – Clinicopathological Factors Affecting Outcome. *Plos One* **7**, e34773 (2012).
137. Holmstrup, P., Vedtofte, P., Reibel, J. & Stoltze, K. Long-term treatment outcome of oral premalignant lesions. *Oral Oncol* **42**, 461–474 (2006).
138. Napier, S. S. & Speight, P. M. Natural history of potentially malignant oral lesions and conditions: an overview of the literature. *J Oral Pathol Med* **37**, 1–10 (2008).
139. Chiu, S.-F. *et al.* Malignant transformation of oral potentially malignant disorders in Taiwan: An observational nationwide population database study. *Medicine* **100**, e24934 (2021).
140. Pinto, A. C. *et al.* Malignant transformation rate of oral leukoplakia—systematic review. *Oral Surg Oral Medicine Oral Pathology Oral Radiology* (2020) doi:10.1016/j.oooo.2020.02.017.
141. Arduino, P. G. *et al.* Outcome of oral dysplasia: a retrospective hospital-based study of 207 patients with a long follow-up. *J Oral Pathol Med* **38**, 540–544 (2009).
142. Mehanna, H. M., Rattay, T., Smith, J. & McConkey, C. C. Treatment and follow-up of oral dysplasia — A systematic review and meta-analysis. *Head Neck* **31**, 1600–1609 (2009).
143. Kierce, J. *et al.* Identification of specific clinical risk factors associated with the malignant transformation of oral epithelial dysplasia. *Head Neck* (2021) doi:10.1002/hed.26851.
144. Gilvetti, C., Soneji, C., Bisase, B. & Barrett, A. W. Recurrence and malignant transformation rates of high grade oral epithelial dysplasia over a 10 year follow up period and the influence of surgical intervention, size of excision biopsy and marginal clearance in a UK regional maxillofacial surgery unit. *Oral Oncol* **121**, 105462 (2021).
145. Sperandio, M. *et al.* Predictive Value of Dysplasia Grading and DNA Ploidy in Malignant Transformation of Oral Potentially Malignant Disorders. *Cancer Prev Res* **6**, 822–831 (2013).
146. Lodi, G. & Porter, S. Management of potentially malignant disorders: evidence and critique. *J Oral Pathol Med* **37**, 63–69 (2008).

147. Zhang, L., Lubpairee, T., Laronde, D. M. & Rosin, M. P. Should severe epithelial dysplasia be treated? *Oral Oncol* **60**, 125–129 (2016).
148. Roosaar, A. *et al.* A long-term follow-up study on the natural course of oral leukoplakia in a Swedish population-based sample. *J Oral Pathol Med* **36**, 78–82 (2007).
149. Roed-Petersen, B. Effect on oral leukoplakia of reducing or ceasing tobacco smoking. *Acta Derm-venereol* **62**, 164–7 (1982).
150. Burris, J. L., Studts, J. L., DeRosa, A. P. & Ostroff, J. S. Systematic Review of Tobacco Use after Lung or Head/Neck Cancer Diagnosis: Results and Recommendations for Future Research. *Cancer Epidemiology Prev Biomarkers* **24**, 1450–1461 (2015).
151. Vladimirov, B. S. & Schiodt, M. The effect of quitting smoking on the risk of unfavorable events after surgical treatment of oral potentially malignant lesions. *Int J Oral Max Surg* **38**, 1188–1193 (2009).
152. Radaic, A. *et al.* Specific Oral Microbial Differences in Proteobacteria and Bacteroidetes Are Associated with Distinct Sites When Moving from Healthy Mucosa to Oral Dysplasia—A Microbiome and Gene Profiling Study and Focused Review. *Microorganisms* **11**, 2250 (2023).
153. Decsi, G. *et al.* Chicken or the Egg: Microbial Alterations in Biopsy Samples of Patients with Oral Potentially Malignant Disorders. *Pathol Oncol Res* 1–11 (2018) doi:10.1007/s12253-018-0457-x.
154. Herreros-Pomares, A. *et al.* Oral microbiome in Proliferative Verrucous Leukoplakia exhibits loss of diversity and enrichment of pathogens. *Oral Oncol* **120**, 105404 (2021).
155. Shen, X., Zhang, Y., Zhu, J. & Xu, B. Oral dysbiosis in the onset and carcinogenesis of oral epithelial dysplasia: A systematic review. *Arch Oral Biol* **147**, 105630 (2023).
156. Gallimidi, A. B. *et al.* Periodontal pathogens *Porphyromonas gingivalis* and *Fusobacterium nucleatum* promote tumor progression in an oral-specific chemical carcinogenesis model. *Oncotarget* **6**, 22613–22623 (2015).
157. Abdalla, Z., Walsh, T., Thakker, N. & Ward, C. M. Loss of epithelial markers is an early event in oral dysplasia and is observed within the safety margin of dysplastic and T1 OSCC biopsies. *Plos One* **12**, e0187449 (2017).
158. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J Clin Invest* **119**, 1420–1428 (2009).
159. Yu, H. *et al.* *Helicobacter pylori* Promotes Epithelial–Mesenchymal Transition in Gastric Cancer by Downregulating Programmed Cell Death Protein 4 (PDCD4). *Plos One* **9**, e105306 (2014).
160. Pan, C., Xu, X., Tan, L., Lin, L. & Pan, Y. The effects of *Porphyromonas gingivalis* on the cell cycle progression of human gingival epithelial cells. *Oral Dis.* **20**, 100–108 (2014).

161. Rubinstein, M. R. *et al.* Fusobacterium nucleatum promotes colorectal cancer by inducing Wnt/ $\beta$ -catenin modulator Annexin A1. *Embo Rep* **20**, (2019).
162. Nosova, T. *et al.* Aldehyde dehydrogenase activity and acetate production by aerobic bacteria representing the normal oral flora of human large intestine. *Alcohol Alcoholism* **31**, 555–564 (1996).
163. Baan, R. *et al.* Carcinogenicity of alcoholic beverages. *Lancet Oncol* **8**, 292–293 (2007).
164. Cancer., I. A. for R. on. Acetaldehyde. IARC monographs on the evaluations of carcinogenic risks to humans. **71**, 319–335 (1999).
165. Hooper, S. J., Wilson, M. J. & Crean, S. J. Exploring the link between microorganisms and oral cancer: A systematic review of the literature. *Head Neck* **31**, 1228–1239 (2009).
166. Salaspuro, M. P. Acetaldehyde, Microbes, and Cancer of the Digestive Tract. *Crit Rev Cl Lab Sci* **40**, 183–208 (2003).
167. Marttila, E. *et al.* Acetaldehyde production and microbial colonization in oral squamous cell carcinoma and oral lichenoid disease. *Oral Surg Oral Medicine Oral Pathology Oral Radiology* **116**, 61–8 (2013).
168. Homann, N. *et al.* Poor dental status increases acetaldehyde production from ethanol in saliva: a possible link to increased oral cancer risk among heavy drinkers. *Oral Oncol* **37**, 153–158 (2001).
169. Brooks, P. J. & Theruvathu, J. A. Acetaldehyde-DNA Adducts: Implications for the Molecular Mechanism of Alcohol-Related Carcinogenesis. *Alcohol* 78–94 (2006) doi:10.1159/000095017.
170. Seitz, H. K. & Stickel, F. Molecular mechanisms of alcohol-mediated carcinogenesis. *Nat Rev Cancer* **7**, 599–612 (2007).
171. Meurman, J. H. Infectious and dietary risk factors of oral cancer. *Oral Oncol* **46**, 411–413 (2010).
172. Homann, N., Jousimies-Somer, H., Jokelainen, K., Heine, R. & Salaspuro, M. High acetaldehyde levels in saliva after ethanol consumption: methodological aspects and pathogenetic implications. *Carcinogenesis* **18**, 1739–1743 (1997).
173. Homann, N. *et al.* Increased salivary acetaldehyde levels in heavy drinkers and smokers: a microbiological approach to oral cavity cancer. *Carcinogenesis* **21**, 663–668 (2000).
174. Alnuaimi, A. *et al.* Candida virulence and ethanol-derived acetaldehyde production in oral cancer and non-cancer subjects. *Oral Dis* **22**, 805–814 (2016).



175. Muto, M. *et al.* Acetaldehyde production by non-pathogenic *Neisseria* in human oral microflora: Implications for carcinogenesis in upper aerodigestive tract. *Int J Cancer* **88**, 342–350 (2000).
176. Moritani, K. *et al.* Acetaldehyde production by major oral microbes. *Oral Dis* **21**, 748–754 (2015).
177. Yokoyama, S. *et al.* Characterization of oral microbiota and acetaldehyde production. *J Oral Microbiol* **10**, 1492316 (2018).
178. Amer, A., Whelan, A., Al-Hebshi, N. N., Healy, C. M. & Moran, G. P. Acetaldehyde production by *Rothia mucilaginosa* isolates from patients with oral leukoplakia. *J Oral Microbiol* **12**, 1743066 (2020).
179. Kurkivuori, J. *et al.* Acetaldehyde production from ethanol by oral streptococci. *Oral Oncol* **43**, 181–186 (2007).
180. Francescone, R., Hou, V. & Grivennikov, S. I. Microbiome, Inflammation, and Cancer. *Cancer J* **20**, 181–189 (2014).
181. O’Grady, I., Anderson, A. & O’Sullivan, J. The interplay of the oral microbiome and alcohol consumption in oral squamous cell carcinomas. *Oral Oncol* **110**, 105011 (2020).
182. Han, Y. W. *et al.* Interactions between Periodontal Bacteria and Human Oral Epithelial Cells: *Fusobacterium nucleatum* Adheres to and Invades Epithelial Cells. *Infect Immun* **68**, 3140–3146 (2000).
183. Xu, M. *et al.* FadA from *Fusobacterium nucleatum* Utilizes both Secreted and Nonsecreted Forms for Functional Oligomerization for Attachment and Invasion of Host Cells\*. *J Biol Chem* **282**, 25000–25009 (2007).
184. Marsh *et al.* *Oral Microbiology*. vol. 6 (Edinburgh: Elsevier 2016, 2016).
185. Moyes, D. L. *et al.* Candidalysin is a fungal peptide toxin critical for mucosal infection. *Nature* **532**, 64–68 (2016).
186. Satiman, E. A. F. E. N. *et al.* The role of *Candida albicans* candidalysin ECE1 gene in oral carcinogenesis. *J Oral Pathol Med* **49**, 835–841 (2020).
187. Vadovics, M. *et al.* *Candida albicans* Enhances the Progression of Oral Squamous Cell Carcinoma In Vitro and In Vivo. *mBio* **13**, e03144-21 (2022).
188. Abdulrahim, M. H., McManus, B. A., Flint, S. R. & Coleman, D. C. Genotyping *Candida albicans* from *Candida* Leukoplakia and Non-*Candida* Leukoplakia Shows No Enrichment of Multilocus Sequence Typing Clades but Enrichment of ABC Genotype C in *Candida* Leukoplakia. *Plos One* **8**, e73738 (2013).
189. Saigal, S., Bhargava, A., Mehra, S. & Dakwala, F. Identification of *Candida albicans* by using different culture medias and its association in potentially malignant and malignant lesions. *Contemp Clin Dent* **2**, 188–193 (2011).

190. Roy, S., Astekar, M., Sapra, G., Chitlangia, R. & Raj, N. Evaluation of candidal species among individuals with oral potentially malignant disorders and oral squamous cell carcinoma. *J Oral Maxillofac Pathology* **23**, 302 (2019).
191. Hulimane, S. *et al.* Speciation of *Candida* using CHROMagar in cases with oral epithelial dysplasia and squamous cell carcinoma. *J Clin Exp Dent* **10**, e657–e660 (2018).
192. Bakri, M. M., Hussaini, H. M., Holmes, A. R., Cannon, R. D. & Rich, A. M. Revisiting the association between candidal infection and carcinoma, particularly oral squamous cell carcinoma. *J Oral Microbiol* **2**, (2010).
193. Warnakulasuriya, S. Oral potentially malignant disorders: A comprehensive review on clinical aspects and management. *Oral Oncol* **102**, 104550 (2020).
194. McCullough, M. *et al.* Oral yeast carriage correlates with presence of oral epithelial dysplasia. *Oral Oncol* **38**, 391–393 (2002).
195. Gupta, V. *et al.* Identification of *Candida albicans* using different culture media and its association in leukoplakia and oral squamous cell carcinoma. *J Oral Maxillofac Pathology Jomfp* **23**, 28–35 (2019).
196. Hebbar, P. B., Pai, A. & D, S. Mycological and histological associations of *Candida* in oral mucosal lesions. *J Oral Sci* **55**, 157–160 (2013).
197. Barrett, A., Kingsmill, V. & Speight, P. The frequency of fungal infection in biopsies of oral mucosal lesions. *Oral Dis* **4**, 26–31 (1998).
198. Chiang, W. *et al.* Malignant development in patients with oral potentially malignant disorders detected through nationwide screening: Outcomes of 5-year follow-up at a single hospital. *Head Neck* (2019) doi:10.1002/hed.25973.
199. Alnuaimi, A. D., Wiesenfeld, D., O'Brien-Simpson, N. M., Reynolds, E. C. & McCullough, M. J. Oral *Candida* colonization in oral cancer patients and its relationship with traditional risk factors of oral cancer: A matched case-control study. *Oral Oncol* **51**, 139–145 (2015).
200. Tillonen, J., Homann, N., Rautio, M., Jousimies-Somer, H. & Salaspuro, M. Role of Yeasts in the Salivary Acetaldehyde Production From Ethanol Among Risk Groups for Ethanol-Associated Oral Cavity Cancer. *Alcohol Clin Exp Res* **23**, 1409–1411 (1999).
201. Gainza-Cirauqui, M. L. *et al.* Production of carcinogenic acetaldehyde by *Candida albicans* from patients with potentially malignant oral mucosal disorders. *J Oral Pathol Med* **42**, 243–249 (2013).
202. Nieminen, M. T., Uttamo, J., Salaspuro, M. & Rautemaa, R. Acetaldehyde production from ethanol and glucose by non-*Candida albicans* yeasts in vitro. *Oral Oncol* **45**, e245–e248 (2009).

203. Peduzzi, P., Concato, J., Kemper, E., Holford, T. R. & Feinstein, A. R. A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiology* **49**, 1373–1379 (1996).
204. Belstrøm, D. *et al.* Temporal Stability of the Salivary Microbiota in Oral Health. *PLoS ONE* **11**, e0147472 (2016).
205. Greene, J. G. & Vermillion, J. R. The Simplified Oral Hygiene Index. *J Am Dent Assoc* **68**, 7–13 (1964).
206. Frank, J. A. *et al.* Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl. Environ. Microbiol.* **74**, 2461–70 (2008).
207. Eren, A. M., Borisy, G. G., Huse, S. M. & Welch, J. L. M. Oligotyping analysis of the human oral microbiome. *Proc National Acad Sci* **111**, E2875–E2884 (2014).
208. Kim, M., Morrison, M. & Yu, Z. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J Microbiol Meth* **84**, 81–87 (2011).
209. Zheng, W. *et al.* An accurate and efficient experimental approach for characterization of the complex oral microbiota. *Microbiome* **3**, 48 (2015).
210. Comeau, A. M., Douglas, G. M. & Langille, M. G. I. Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. *mSystems* **2**, e00127-16 (2016).
211. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).
212. Escapa, I. F. *et al.* New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. *mSystems* **3**, e00187-18 (2018).
213. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **8**, e61217 (2013).
214. Xu, S. *et al.* MicrobiotaProcess: A comprehensive R package for managing and analyzing microbiome and other ecological data within the tidy framework. (2022) doi:10.21203/rs.3.rs-1284357/v1.
215. Barnett, D., Arts, I. & Penders, J. microViz: an R package for microbiome data visualization and statistics. *J Open Source Softw* **6**, 3201 (2021).
216. Mallick, H. *et al.* Multivariable Association Discovery in Population-scale Meta-omics Studies. *Biorxiv* 2021.01.20.427420 (2021) doi:10.1101/2021.01.20.427420.
217. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Royal Statistical Soc Ser B Methodol* **57**, 289–300 (1995).

218. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
219. Zych, K. *et al.* SIAMCAT by Bioconductor. Preprint at <https://www.bioconductor.org/packages/devel/bioc/manuals/SIAMCAT/man/SIAMCAT.pdf> (2023).
220. Therneau, T., Atkinson, B. & Ripley, B. rpart: Recursive Partitioning and Regression Trees. <https://cran.r-project.org/package=rpart> (2023).
221. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* **3**, 18–22 (2002).
222. Stevenson *et al.* EpiR: An R package for the analysis of epidemiological data. Preprint at <https://mvs.unimelb.edu.au/research/groups/veterinary-epidemiology-melbourne> (2023).
223. Fox, J. *et al.* Package ‘car.’ Preprint at <https://CRAN.R-project.org/package=car> (2023).
224. Ripley, B. & Venables, W. Package “nnet.” Preprint at <http://www.stats.ox.ac.uk/pub/MASS4/> (2016).
225. Warnakulasuriya, S., Reibel, J., Bouquot, J. & Dabelsteen, E. Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement: Oral epithelial dysplasia classification systems. *J Oral Pathol Med* **37**, 127–133 (2008).
226. Speight, P. M., Khurram, S. A. & Kujan, O. Oral potentially malignant disorders: risk of progression to malignancy. *Oral Surg Oral Medicine Oral Pathology Oral Radiology* **125**, 612–627 (2018).
227. Schepman, K., Bezemer, P., Meij, E. V. D., Smeele, L. & Waal, I. V. D. Tobacco usage in relation to the anatomical site of oral leukoplakia. *Oral Dis.* **7**, 25–27 (2001).
228. Pentenero, M. *et al.* Evidence for a possible anatomical subsite-mediated effect of tobacco in oral potentially malignant disorders and carcinoma. *J. Oral Pathol. Med.* **40**, 214–217 (2011).
229. Embaló, B., Miguel, A. F. P., Konrath, A. C., Modolo, F. & Rivero, E. R. C. Evaluation of two classification systems for oral epithelial dysplasia. *Oral Dis* (2021) doi:10.1111/odi.13867.
230. Willis, A. D. Rarefaction, Alpha Diversity, and Statistics. *Front Microbiol* **10**, 2407 (2019).
231. Lynch, D. Alpha Diversity. <https://docs.onecodex.com/en/articles/4136553-alpha-diversity>.
232. Lynch, D. Beta Diversity. <https://docs.onecodex.com/en/articles/4150649-beta-diversity>.

233. Bray, J. R. & Curtis, J. T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
234. Wen, B.-W. *et al.* Cancer risk among gingivitis and periodontitis patients: a nationwide cohort study. *Qjm Int J Medicine* **107**, 283–290 (2014).
235. Teughels, W., Quirynen, M. & Jakubovics, N. Carranza's Clinical Periodontology. *Part 4 Etiology Periodontal Dis* 232–270 (2012) doi:10.1016/b978-1-4377-0416-7.00023-8.
236. Socransky, S. S. & Huffajee, A. D. Periodontal microbial ecology. *Periodontology* **2000** **38**, 135–187 (2005).
237. Liu, X. *et al.* Sex differences in the oral microbiome, host traits, and their causal relationships. *iScience* **26**, 105839 (2023).
238. Mager, D. L., Ximenez-Fyvie, L. A., Haffajee, A. D. & Socransky, S. S. Distribution of selected bacterial species on intraoral surfaces. *J Clin Periodontol* **30**, 644–654 (2003).
239. Welch, J. L. M., Dewhirst, F. E. & Borisy, G. G. Biogeography of the Oral Microbiome: The Site-Specialist Hypothesis. *Annu Rev Microbiol* **73**, 1–24 (2019).
240. Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I. & Dewhirst, F. E. Defining the Normal Bacterial Flora of the Oral Cavity. *J Clin Microbiol* **43**, 5721–5732 (2005).
241. Segata, N. *et al.* Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol* **13**, R42 (2012).
242. Diaz, P. I. *et al.* Using high throughput sequencing to explore the biodiversity in oral bacterial communities. *Mol Oral Microbiol* **27**, 182–201 (2012).
243. McLean, A. R., Torres-Morales, J., Dewhirst, F. E., Borisy, G. G. & Welch, J. L. M. Site-tropism of streptococci in the oral microbiome. *Mol Oral Microbiol* (2022) doi:10.1111/omi.12387.
244. Liu, C. R. *et al.* Gemella morbillorum mitral valve endocarditis in a patient with a history of mitral valve annuloplasty. *Bmj Case Reports* **15**, e247033 (2022).
245. Kwong, T. N. Y. *et al.* Association Between Bacteremia From Specific Microbes and Subsequent Diagnosis of Colorectal Cancer. *Gastroenterology* **155**, 383-390.e8 (2018).
246. Chen, Y. *et al.* Bacteremia caused by Bergeyella zoohelcum in an infective endocarditis patient: case report and review of literature. *Bmc Infect Dis* **17**, 271 (2017).
247. Peters, B. A. *et al.* Oral Microbiome Composition Reflects Prospective Risk for Esophageal Cancers. *Cancer Res* **77**, 6777–6787 (2017).
248. Tarca, A. L., Carey, V. J., Chen, X., Romero, R. & Drăghici, S. Machine Learning and Its Applications to Biology. *PLoS Comput. Biol.* **3**, e116 (2007).

249. Friedman, J. *et al.* Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **1**, 1–22 (2010).
250. Marcos-Zambrano, L. J. *et al.* Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment. *Front. Microbiol.* **12**, 634511 (2021).
251. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
252. Fukui, H. *et al.* Usefulness of Machine Learning-Based Gut Microbiome Analysis for Identifying Patients with Irritable Bowels Syndrome. *J. Clin. Med.* **9**, 2403 (2020).
253. Ellis, D. E. *et al.* Comparing LASSO and random forest models for predicting neurological dysfunction among fluoroquinolone users. *Pharmacoepidemiol. Drug Saf.* **31**, 393–403 (2022).
254. Shin, J. M. *et al.* Microbial Communities Associated with Primary and Metastatic Head and Neck Squamous Cell Carcinoma – A High Fusobacterial and Low Streptococcal Signature. *Sci Rep-uk* **7**, 9934 (2017).
255. Eun, Y.-G. *et al.* Oral microbiome associated with lymph node metastasis in oral squamous cell carcinoma. *Sci. Reports* **11**, 23176 (2021).
256. Belibasakis, G. N. *et al.* Periodontal microbiology and microbial etiology of periodontal diseases: Historical concepts and contemporary perspectives. *Periodontol. 2000* (2023) doi:10.1111/prd.12473.
257. Abdulkareem, A. A. *et al.* Current concepts in the pathogenesis of periodontitis: from symbiosis to dysbiosis. *J. Oral Microbiol.* **15**, 2197779 (2023).
258. Uchibori, S., Tsuzukibashi, O., Kobayashi, T. & Aida, M. Localization of the Genus *Rothia* in the Oral Cavity. *Int. J. Oral-Méd. Sci.* **11**, 207–210 (2013).
259. Deo, P. N. & Deshmukh, R. Oral microbiome: Unveiling the fundamentals. *J Oral Maxillofac Pathology Jomfp* **23**, 122–128 (2019).
260. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* **35**, 833–844 (2017).
261. Cour, C. D. de la *et al.* Prevalence of human papillomavirus in oral epithelial dysplasia: Systematic review and meta-analysis. *Head Neck* **42**, 2975–2984 (2020).
262. Clarke, T. *et al.* Sampling from four geographically divergent young female populations demonstrates forensic geolocation potential in microbiomes. *Sci. Rep.* **12**, 18547 (2022).
263. Cho, H.-W. & Eom, Y.-B. Forensic Analysis of Human Microbiome in Skin and Body Fluids Based on Geographic Location. *Front. Cell. Infect. Microbiol.* **11**, 695191 (2021).

## 12 Appendix A Full results from Table 36

**Table 89.** Results of MaAsLin2 analysis showing taxa with significant changes in abundance ( $p_{adj} < 0.1$ ) associated with site. (Buc/Alv/Gin [buccal mucosa/alveolus/gingiva] used as reference) (negative values denote reduced abundance). DT/LBT=dorsum/lateral border tongue, FOM/VT=floor of mouth/ventral tongue, PAL=palate

| Taxa                                                                                                                                                                             | Site   | Coefficient | p value  | q value |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-------------|----------|---------|
| <i>Streptococcus mitis pneumoniae</i>                                                                                                                                            | DT/LBT | -0.71       | 0.0002   | 0.003   |
| <i>Gemella haemolysans</i>                                                                                                                                                       | DT/LBT | -0.63       | 0.0001   | 0.003   |
| <i>Streptococcus</i> sp. HMT423 <i>mitis</i>                                                                                                                                     | DT/LBT | -0.51       | 8.09E-05 | 0.002   |
| <i>Streptococcus</i> sp. HMT064 <i>mitis</i>                                                                                                                                     | DT/LBT | -0.51       | 0.0002   | 0.003   |
| <i>Streptococcus sanguinis</i>                                                                                                                                                   | DT/LBT | -0.37       | 0.001    | 0.02    |
| <i>Parvimonas micra</i>                                                                                                                                                          | DT/LBT | -0.37       | 0.004    | 0.04    |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 070                                                                                                                    | DT/LBT | -0.34       | 0.003    | 0.04    |
| <i>Rothia dentocariosa</i>                                                                                                                                                       | DT/LBT | -0.32       | 0.007    | 0.06    |
| <i>Peptostreptococcaceae</i> XI G9 <i>brachy</i>                                                                                                                                 | DT/LBT | -0.30       | 0.008    | 0.06    |
| <i>Prevotella denticola</i>                                                                                                                                                      | DT/LBT | -0.30       | 0.01     | 0.08    |
| <i>Actinomyces</i> sp. HMT169                                                                                                                                                    | DT/LBT | -0.30       | 0.008    | 0.06    |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 071                                                                                                                    | DT/LBT | -0.28       | 0.01     | 0.09    |
| <i>Campylobacter gracilis</i>                                                                                                                                                    | DT/LBT | -0.28       | 0.009    | 0.07    |
| <i>Actinomyces oris</i>                                                                                                                                                          | DT/LBT | -0.27       | 0.0001   | 0.003   |
| <i>Corynebacterium matruchotii</i>                                                                                                                                               | DT/LBT | -0.27       | 0.005    | 0.04    |
| <i>Mogibacterium timidum</i>                                                                                                                                                     | DT/LBT | -0.25       | 0.004    | 0.04    |
| <i>Streptococcus</i> sp. HMT064 <i>australis oralis</i> subsp. <i>dentisani</i> clade 398 sp. HMT423 <i>infantis</i> clade 431 <i>infantis</i> clade 638 <i>mitis pneumoniae</i> | DT/LBT | -0.22       | 0.006    | 0.05    |
| <i>Capnocytophaga granulosa</i>                                                                                                                                                  | DT/LBT | -0.22       | 0.01     | 0.08    |
| <i>Leptotrichia buccalis</i>                                                                                                                                                     | DT/LBT | -0.22       | 0.01     | 0.08    |
| <i>Peptostreptococcaceae</i> XI G4 <i>bacterium</i> HMT369                                                                                                                       | DT/LBT | -0.22       | 0.008    | 0.06    |
| <i>Prevotella</i> sp. HMT300                                                                                                                                                     | DT/LBT | -0.19       | 0.003    | 0.04    |
| <i>Fusobacterium nucleatum</i> subsp. <i>animalis</i>                                                                                                                            | DT/LBT | -0.19       | 0.02     | 0.09    |
| <i>Fretibacterium</i> sp. HMT359                                                                                                                                                 | DT/LBT | -0.18       | 0.01     | 0.09    |
| <i>Bulleidia extructa</i>                                                                                                                                                        | DT/LBT | -0.14       | 0.01     | 0.08    |
| <i>Haemophilus pittmaniae</i>                                                                                                                                                    | DT/LBT | 0.16        | 0.003    | 0.03    |
| <i>Streptococcus infantis</i> clade 431 <i>infantis</i> clade 638                                                                                                                | DT/LBT | 0.16        | 0.01     | 0.09    |
| <i>Prevotella</i> sp. HMT396                                                                                                                                                     | DT/LBT | 0.20        | 0.006    | 0.05    |
| <i>Streptococcus thermophilus salivarius</i>                                                                                                                                     | DT/LBT | 0.21        | 0.001    | 0.01    |
| <i>Actinomyces lingnae</i> Not Validly Published                                                                                                                                 | DT/LBT | 0.38        | 0.0006   | 0.01    |
| <i>Streptococcus australis</i>                                                                                                                                                   | DT/LBT | 0.38        | 0.002    | 0.02    |
| <i>Veillonella rogosae dispar parvula atypica</i>                                                                                                                                | DT/LBT | 0.38        | 3.42E-06 | 0.0003  |
| <i>Actinomyces odontolyticus</i>                                                                                                                                                 | DT/LBT | 0.38        | 0.002    | 0.02    |
| <i>Granulicatella adiacens</i>                                                                                                                                                   | DT/LBT | 0.39        | 1.71E-06 | 0.00016 |
| <i>Streptococcus</i> sp. HMT066 sp. HMT074                                                                                                                                       | DT/LBT | 0.39        | 0.006    | 0.05    |
| <i>Gemella sanguinis</i>                                                                                                                                                         | DT/LBT | 0.43        | 0.0010   | 0.01    |
| <i>Streptococcus infantis</i> clade 638                                                                                                                                          | DT/LBT | 0.48        | 0.0003   | 0.006   |

|                                                                             |        |       |          |          |
|-----------------------------------------------------------------------------|--------|-------|----------|----------|
| <i>Streptococcus</i> sp. HMT061                                             | DT/LBT | 0.48  | 1.82E-05 | 0.0008   |
| <i>Veillonella denticariosi</i>                                             | DT/LBT | 0.51  | 1.17E-05 | 0.0006   |
| <i>Rothia mucilaginosa</i>                                                  | DT/LBT | 0.51  | 4.56E-09 | 6.40E-07 |
| <i>Actinomyces</i> sp. HMT180 <i>odontolyticus</i>                          | DT/LBT | 0.51  | 2.89E-05 | 0.001    |
| <i>Streptococcus parasanguinis</i> clade 411                                | DT/LBT | 0.54  | 0.0002   | 0.004    |
| <i>Streptococcus parasanguinis</i> clade 411 <i>parasanguinis</i> clade 721 | DT/LBT | 0.58  | 7.01E-05 | 0.002    |
| <i>Rothia aeria dentocariosa</i>                                            | DT/LBT | 0.84  | 6.52E-12 | 3.66E-09 |
| <i>Veillonella rogosae dispar</i>                                           | DT/LBT | 0.85  | 2.28E-09 | 4.27E-07 |
| <i>Gemella haemolysans</i>                                                  | FOM/VT | -0.57 | 0.01     | 0.07     |
| <i>Streptococcus constellatus intermedius</i>                               | FOM/VT | -0.54 | 0.004    | 0.04     |
| <i>Parvimonas micra</i>                                                     | FOM/VT | -0.45 | 0.01     | 0.08     |
| <i>Mycoplasma faucium</i>                                                   | FOM/VT | -0.30 | 0.01     | 0.08     |
| <i>Parvimonas micra</i>                                                     | PAL    | -0.62 | 0.001    | 0.02     |
| <i>Rothia dentocariosa</i>                                                  | PAL    | -0.62 | 0.0005   | 0.009    |
| <i>Fusobacterium nucleatum subsp vincentii</i>                              | PAL    | -0.56 | 0.002    | 0.02     |
| <i>Streptococcus constellatus intermedius</i>                               | PAL    | -0.50 | 0.01     | 0.08     |
| <i>Porphyromonas endodontalis</i>                                           | PAL    | -0.50 | 0.007    | 0.06     |
| <i>Prevotella denticola</i>                                                 | PAL    | -0.48 | 0.006    | 0.05     |
| <i>Haemophilus haemolyticus</i>                                             | PAL    | -0.45 | 0.001    | 0.01     |
| <i>Peptostreptococcaceae</i> XI G 9 <i>brachy</i>                           | PAL    | -0.41 | 0.01     | 0.09     |
| <i>Campylobacter gracilis</i>                                               | PAL    | -0.40 | 0.01     | 0.09     |
| <i>Leptotrichia hongkongensis</i>                                           | PAL    | -0.37 | 0.003    | 0.04     |
| <i>Eikenella corrodens</i>                                                  | PAL    | -0.34 | 0.009    | 0.06     |
| <i>Tannerella</i> sp. HMT286                                                | PAL    | -0.34 | 0.01     | 0.09     |
| <i>Tannerella forsythia</i>                                                 | PAL    | -0.33 | 0.009    | 0.06     |
| <i>Mogibacterium timidum</i>                                                | PAL    | -0.31 | 0.02     | 0.09     |
| <i>Abiotrophia defectiva</i>                                                | PAL    | -0.27 | 0.007    | 0.06     |
| <i>Veillonella</i> sp. HMT780                                               | PAL    | -0.26 | 0.008    | 0.06     |
| <i>Prevotella</i> sp. HMT300                                                | PAL    | -0.26 | 0.006    | 0.05     |
| <i>Campylobacter rectus</i>                                                 | PAL    | -0.18 | 0.02     | 0.09     |
| <i>Stomatobaculum longum</i>                                                | PAL    | 0.31  | 0.01     | 0.09     |
| <i>Rothia mucilaginosa</i>                                                  | PAL    | 0.32  | 0.01     | 0.08     |
| <i>Lachnoanaerobaculum orale</i>                                            | PAL    | 0.34  | 0.0002   | 0.005    |
| <i>Selenomonas</i> sp. HMT136                                               | PAL    | 0.40  | 2.84E-05 | 0.001    |
| <i>Peptostreptococcaceae</i> XI G 1 <i>sulci</i>                            | PAL    | 0.41  | 2.08E-07 | 2.34E-05 |
| <i>Stomatobaculum</i> sp. HMT097                                            | PAL    | 0.42  | 0.0005   | 0.009    |
| <i>Streptococcus salivarius</i>                                             | PAL    | 0.42  | 6.38E-05 | 0.002    |
| <i>Oribacterium sinus</i>                                                   | PAL    | 0.45  | 0.008    | 0.06     |
| <i>Leptotrichia</i> sp. HMT215                                              | PAL    | 0.45  | 0.003    | 0.04     |
| <i>Campylobacter concisus</i>                                               | PAL    | 0.46  | 0.004    | 0.04     |
| <i>Streptococcus vestibularis thermophilus salivarius</i>                   | PAL    | 0.47  | 0.0001   | 0.003    |
| <i>Actinomyces lingnae</i> Not Validly Published                            | PAL    | 0.49  | 0.002    | 0.03     |
| <i>Saccharibacteria</i> TM7 G1 <i>bacterium</i> HMT352                      | PAL    | 0.55  | 6.49E-06 | 0.0004   |
| <i>Prevotella melaninogenica</i>                                            | PAL    | 0.55  | 0.005    | 0.05     |
| <i>Streptococcus parasanguinis</i> clade 411 <i>parasanguinis</i> clade 721 | PAL    | 0.55  | 0.01     | 0.07     |



|                                              |     |      |          |          |
|----------------------------------------------|-----|------|----------|----------|
| <i>Prevotella</i> sp. HMT305                 | PAL | 0.56 | 3.74E-10 | 1.05E-07 |
| <i>Streptococcus</i> sp. HMT057              | PAL | 0.56 | 0.004    | 0.04     |
| <i>Actinomyces</i> sp. HMT172                | PAL | 0.60 | 0.0006   | 0.01     |
| <i>Leptotrichia</i> sp. HMT417               | PAL | 0.62 | 0.006    | 0.05     |
| <i>Actinomyces odontolyticus</i>             | PAL | 0.62 | 0.0007   | 0.01     |
| <i>Solobacterium moorei</i>                  | PAL | 0.62 | 5.48E-05 | 0.002    |
| <i>Prevotella</i> sp. HMT313                 | PAL | 0.65 | 0.0006   | 0.01     |
| <i>Mogibacterium diversum</i>                | PAL | 0.65 | 4.43E-06 | 0.0003   |
| <i>Alloprevotella</i> sp. HMT308             | PAL | 0.66 | 5.78E-05 | 0.002    |
| <i>Streptococcus parasanguinis</i> clade 721 | PAL | 0.68 | 0.0006   | 0.01     |
| <i>Actinomyces graevenitzii</i>              | PAL | 0.69 | 0.001    | 0.01     |
| <i>Prevotella pallens</i>                    | PAL | 0.72 | 9.99E-06 | 0.0006   |
| <i>Atopobium parvulum</i>                    | PAL | 0.82 | 5.94E-05 | 0.002    |
| <i>Prevotella salivae</i>                    | PAL | 0.96 | 1.38E-05 | 0.0006   |
| <i>Prevotella histicola</i>                  | PAL | 1.05 | 2.00E-05 | 0.0008   |

## 13 Appendix B Full results from Table 39

**Table 90.** Results of MaAsLin2 analysis of a site and smoking model showing taxa with significant changes in abundance (padj<0.1) (Negative coefficient values denote reduced abundance)

| Taxa                                              | Variable | Coefficient | p value  | Adjusted p value |
|---------------------------------------------------|----------|-------------|----------|------------------|
| <i>Neisseria perflava flavescens</i>              | Current  | -1.21       | 9.52E-11 | 2.97E-08         |
| <i>Lautropia mirabilis</i>                        | Current  | -0.75       | 1.49E-11 | 6.95E-09         |
| <i>Haemophilus parainfluenzae</i>                 | Current  | -0.71       | 1.12E-08 | 1.31E-06         |
| <i>Streptococcus</i> sp. HMT066 sp. HMT074        | Current  | -0.70       | 5.05E-07 | 3.93E-05         |
| <i>Rothia aeria</i>                               | Current  | -0.69       | 3.85E-09 | 6.00E-07         |
| <i>Streptococcus australis</i>                    | Current  | -0.66       | 5.64E-08 | 5.86E-06         |
| <i>Kingella oralis</i>                            | Current  | -0.66       | 6.39E-07 | 4.60E-05         |
| <i>Streptococcus sanguinis</i>                    | Current  | -0.53       | 3.71E-06 | 0.0002           |
| <i>Neisseria macacae flava mucosa sicca</i>       | Current  | -0.51       | 0.0003   | 0.004            |
| <i>Actinomyces</i> sp. HMT169                     | Current  | -0.49       | 9.30E-06 | 0.0003           |
| <i>Streptococcus</i> sp. HMT066                   | Current  | -0.49       | 0.0005   | 0.007            |
| <i>Fusobacterium periodonticum hwasookii</i>      | Current  | -0.46       | 0.0002   | 0.003            |
| <i>Stomatobaculum</i> sp. HMT097                  | Current  | -0.41       | 7.63E-07 | 5.10E-05         |
| <i>Capnocytophaga sputigena</i>                   | Current  | -0.39       | 4.35E-05 | 0.001            |
| <i>Bergeyella</i> sp. HMT322                      | Current  | -0.39       | 3.14E-06 | 0.0002           |
| <i>Abiotrophia defectiva</i>                      | Current  | -0.32       | 3.82E-06 | 0.0002           |
| <i>Neisseria oralis</i>                           | Current  | -0.31       | 2.23E-05 | 0.0006           |
| <i>Gemella morbillorum</i>                        | Current  | -0.31       | 0.004    | 0.04             |
| <i>Streptococcus intermedius</i>                  | Current  | -0.31       | 0.0005   | 0.007            |
| <i>Campylobacter concisus</i>                     | Current  | -0.30       | 0.007    | 0.05             |
| <i>Cardiobacterium hominis</i>                    | Current  | -0.29       | 6.54E-06 | 0.0003           |
| <i>Selenomonas artemidis</i>                      | Current  | -0.25       | 0.003    | 0.03             |
| <i>Neisseria elongata</i>                         | Current  | -0.22       | 0.007    | 0.05             |
| <i>Aggregatibacter</i> sp. HMT512                 | Current  | -0.21       | 0.002    | 0.02             |
| <i>Capnocytophaga gingivalis</i>                  | Current  | -0.21       | 0.005    | 0.04             |
| <i>Aggregatibacter</i> sp. HMT458                 | Current  | -0.20       | 0.006    | 0.04             |
| <i>Lachnoanaerobaculum umeaense</i>               | Current  | -0.19       | 0.006    | 0.04             |
| <i>Neisseria bacilliformis</i>                    | Current  | -0.16       | 0.02     | 0.08             |
| <i>Porphyromonas catoniae</i>                     | Current  | -0.12       | 0.004    | 0.04             |
| <i>Bulleidia extructa</i>                         | Current  | 0.13        | 0.02     | 0.08             |
| <i>Treponema</i> sp. HMT237                       | Current  | 0.20        | 0.005    | 0.04             |
| <i>Prevotella</i> sp. HMT526                      | Current  | 0.20        | 0.003    | 0.03             |
| <i>Fusobacterium</i> sp. HMT203                   | Current  | 0.21        | 0.004    | 0.03             |
| <i>Veillonella rogosae dispar parvula atypica</i> | Current  | 0.23        | 0.006    | 0.04             |
| <i>Streptococcus</i> sp. HMT064 sp. HMT423 mitis  | Current  | 0.23        | 0.006    | 0.04             |
| <i>Tannerella forsythia</i>                       | Current  | 0.23        | 0.007    | 0.04             |
| <i>Peptostreptococcaceae XI G6 nodatum</i>        | Current  | 0.23        | 0.004    | 0.03             |
| <i>Clostridiales F1 G1 bacterium</i> HMT093       | Current  | 0.23        | 1.70E-05 | 0.0005           |
| <i>Streptococcus thermophilus salivarius</i>      | Current  | 0.24        | 0.0002   | 0.004            |

|                                                                                                                                                                                        |         |       |          |          |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|-------|----------|----------|
| <i>Bacteroidales G2 bacterium</i> HMT274                                                                                                                                               | Current | 0.26  | 0.004    | 0.04     |
| <i>Peptostreptococcaceae XI G4 bacterium</i> HMT369                                                                                                                                    | Current | 0.28  | 0.0006   | 0.008    |
| <i>Peptostreptococcaceae XI G9 brachy</i>                                                                                                                                              | Current | 0.28  | 0.01     | 0.07     |
| <i>Streptococcus vestibularis salivarius</i>                                                                                                                                           | Current | 0.29  | 0.01     | 0.07     |
| <i>Stomatobaculum longum</i>                                                                                                                                                           | Current | 0.32  | 0.0002   | 0.004    |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 070                                                                                                                          | Current | 0.33  | 0.005    | 0.04     |
| <i>Streptococcus parasanguinis</i> clade 721                                                                                                                                           | Current | 0.34  | 0.01     | 0.07     |
| <i>Mogibacterium timidum</i>                                                                                                                                                           | Current | 0.36  | 5.42E-05 | 0.001    |
| <i>Fusobacterium nucleatum</i> subsp. <i>vincentii</i>                                                                                                                                 | Current | 0.36  | 0.004    | 0.03     |
| <i>Veillonella denticariosi</i>                                                                                                                                                        | Current | 0.36  | 0.002    | 0.02     |
| <i>Mycoplasma faucium</i>                                                                                                                                                              | Current | 0.36  | 3.45E-05 | 0.0009   |
| <i>Filifactor alocis</i>                                                                                                                                                               | Current | 0.36  | 0.004    | 0.03     |
| <i>Streptococcus parasanguinis</i> clade 411                                                                                                                                           | Current | 0.36  | 0.01     | 0.08     |
| <i>Alloprevotella tannerae</i>                                                                                                                                                         | Current | 0.37  | 0.008    | 0.05     |
| <i>Actinomyces lingnae</i> Not Validly Published                                                                                                                                       | Current | 0.39  | 0.0004   | 0.006    |
| <i>Parvimonas micra</i>                                                                                                                                                                | Current | 0.40  | 0.002    | 0.02     |
| <i>Prevotella</i> sp. HMT313                                                                                                                                                           | Current | 0.42  | 0.001    | 0.01     |
| <i>Streptococcus</i> sp. HMT057                                                                                                                                                        | Current | 0.43  | 0.001    | 0.01     |
| <i>Porphyromonas endodontalis</i>                                                                                                                                                      | Current | 0.45  | 0.0003   | 0.005    |
| <i>Actinomyces graevenitzi</i>                                                                                                                                                         | Current | 0.50  | 0.0005   | 0.007    |
| <i>Peptostreptococcaceae XI G5 saphenum</i>                                                                                                                                            | Current | 0.50  | 1.99E-06 | 0.0001   |
| <i>Prevotella pleuritidis</i>                                                                                                                                                          | Current | 0.52  | 9.62E-07 | 6.00E-05 |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 071                                                                                                                          | Current | 0.53  | 3.49E-06 | 0.0002   |
| <i>Prevotella histicola</i>                                                                                                                                                            | Current | 0.59  | 0.0004   | 0.007    |
| <i>Atopobium parvulum</i>                                                                                                                                                              | Current | 0.59  | 1.51E-05 | 0.0005   |
| <i>Streptococcus parasanguinis</i> clade 411<br><i>parasanguinis</i> clade 721                                                                                                         | Current | 0.76  | 1.24E-07 | 1.16E-05 |
| <i>Actinomyces</i> sp. HMT169                                                                                                                                                          | Former  | -0.50 | 7.46E-05 | 0.001    |
| <i>Streptococcus intermedius</i>                                                                                                                                                       | Former  | -0.41 | 3.76E-05 | 0.0009   |
| <i>Streptococcus sanguinis</i>                                                                                                                                                         | Former  | -0.36 | 0.005    | 0.04     |
| <i>Actinomyces</i> sp. HMT175                                                                                                                                                          | Former  | -0.33 | 0.004    | 0.04     |
| <i>Selenomonas artemidis</i>                                                                                                                                                           | Former  | -0.32 | 0.0009   | 0.01     |
| <i>Rothia aerea</i>                                                                                                                                                                    | Former  | -0.31 | 0.02     | 0.08     |
| <i>Lautropia mirabilis</i>                                                                                                                                                             | Former  | -0.29 | 0.02     | 0.08     |
| <i>Actinomyces</i> sp. HMT448                                                                                                                                                          | Former  | -0.27 | 0.005    | 0.04     |
| <i>Cardiobacterium hominis</i>                                                                                                                                                         | Former  | -0.23 | 0.001    | 0.02     |
| <i>Streptococcus</i> sp. HMT064 <i>oralis</i> subsp. <i>dentisani</i><br>clade 398                                                                                                     | Former  | -0.23 | 0.01     | 0.08     |
| <i>Corynebacterium durum</i>                                                                                                                                                           | Former  | -0.19 | 0.01     | 0.07     |
| <i>Ruminococcaceae G2 bacterium</i> HMT085                                                                                                                                             | Former  | 0.22  | 0.01     | 0.07     |
| <i>Streptococcus</i> sp. HMT064 <i>australis oralis</i> subsp.<br><i>dentisani</i> clade 398 sp. HMT423 <i>infantis</i> clade 431<br><i>infantis</i> clade 638 <i>mitis pneumoniae</i> | Former  | 0.25  | 0.006    | 0.04     |
| <i>Streptococcus mitis pneumoniae</i>                                                                                                                                                  | DT/LBT  | -0.70 | 0.0002   | 0.004    |
| <i>Gemella haemolysans</i>                                                                                                                                                             | DT/LBT  | -0.59 | 0.0002   | 0.004    |
| <i>Streptococcus</i> sp. HMT064 <i>mitis</i>                                                                                                                                           | DT/LBT  | -0.51 | 0.0002   | 0.003    |
| <i>Streptococcus</i> sp. HMT423 <i>mitis</i>                                                                                                                                           | DT/LBT  | -0.50 | 0.0001   | 0.002    |
| <i>Parvimonas micra</i>                                                                                                                                                                | DT/LBT  | -0.39 | 0.002    | 0.02     |

|                                                                                                                                                                                  |        |       |          |          |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-------|----------|----------|
| <i>Streptococcus sanguinis</i>                                                                                                                                                   | DT/LBT | -0.38 | 0.0007   | 0.009    |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 070                                                                                                                    | DT/LBT | -0.36 | 0.002    | 0.02     |
| <i>Rothia dentocariosa</i>                                                                                                                                                       | DT/LBT | -0.34 | 0.005    | 0.04     |
| <i>Actinomyces</i> sp. HMT169                                                                                                                                                    | DT/LBT | -0.31 | 0.004    | 0.03     |
| <i>Peptostreptococcaceae</i> XI G9 <i>brachy</i>                                                                                                                                 | DT/LBT | -0.31 | 0.005    | 0.04     |
| <i>Streptococcus oralis</i> subsp. <i>tigurinus</i> clade 071                                                                                                                    | DT/LBT | -0.30 | 0.007    | 0.04     |
| <i>Prevotella denticola</i>                                                                                                                                                      | DT/LBT | -0.30 | 0.01     | 0.07     |
| <i>Campylobacter gracilis</i>                                                                                                                                                    | DT/LBT | -0.29 | 0.007    | 0.04     |
| <i>Actinomyces oris</i>                                                                                                                                                          | DT/LBT | -0.28 | 7.44E-05 | 0.001    |
| <i>Corynebacterium matruchotii</i>                                                                                                                                               | DT/LBT | -0.27 | 0.003    | 0.03     |
| <i>Mogibacterium timidum</i>                                                                                                                                                     | DT/LBT | -0.26 | 0.003    | 0.03     |
| <i>Actinomyces</i> sp. HMT175                                                                                                                                                    | DT/LBT | -0.24 | 0.02     | 0.08     |
| <i>Peptostreptococcaceae</i> XI G4 <i>bacterium</i> HMT369                                                                                                                       | DT/LBT | -0.22 | 0.005    | 0.04     |
| <i>Leptotrichia buccalis</i>                                                                                                                                                     | DT/LBT | -0.22 | 0.01     | 0.07     |
| <i>Capnocytophaga granulosa</i>                                                                                                                                                  | DT/LBT | -0.22 | 0.02     | 0.08     |
| <i>Streptococcus</i> sp. HMT064 <i>australis oralis</i> subsp. <i>dentisani</i> clade 398 sp. HMT423 <i>infantis</i> clade 431 <i>infantis</i> clade 638 <i>mitis pneumoniae</i> | DT/LBT | -0.21 | 0.010    | 0.06     |
| <i>Tannerella forsythia</i>                                                                                                                                                      | DT/LBT | -0.20 | 0.02     | 0.09     |
| <i>Fusobacterium nucleatum</i> subsp. <i>animalis</i>                                                                                                                            | DT/LBT | -0.20 | 0.01     | 0.06     |
| <i>Prevotella</i> sp. HMT300                                                                                                                                                     | DT/LBT | -0.19 | 0.003    | 0.03     |
| <i>Fretibacterium</i> sp. HMT359                                                                                                                                                 | DT/LBT | -0.18 | 0.01     | 0.07     |
| <i>Abiotrophia defectiva</i>                                                                                                                                                     | DT/LBT | -0.16 | 0.02     | 0.08     |
| <i>Bulleidia extructa</i>                                                                                                                                                        | DT/LBT | -0.14 | 0.010    | 0.06     |
| <i>Haemophilus pittmaniae</i>                                                                                                                                                    | DT/LBT | 0.17  | 0.002    | 0.02     |
| <i>Streptococcus infantis</i> clade 431 <i>infantis</i> clade 638                                                                                                                | DT/LBT | 0.17  | 0.01     | 0.06     |
| <i>Prevotella</i> sp. HMT396                                                                                                                                                     | DT/LBT | 0.19  | 0.006    | 0.04     |
| <i>Streptococcus thermophilus salivarius</i>                                                                                                                                     | DT/LBT | 0.21  | 0.0008   | 0.01     |
| <i>Streptococcus australis infantis</i> clade 431                                                                                                                                | DT/LBT | 0.29  | 0.01     | 0.08     |
| <i>Haemophilus parainfluenzae</i>                                                                                                                                                | DT/LBT | 0.31  | 0.009    | 0.06     |
| <i>Streptococcus</i> sp. HMT066                                                                                                                                                  | DT/LBT | 0.34  | 0.01     | 0.08     |
| <i>Actinomyces odontolyticus</i>                                                                                                                                                 | DT/LBT | 0.36  | 0.003    | 0.03     |
| <i>Actinomyces lingnae</i> Not Validly Published                                                                                                                                 | DT/LBT | 0.37  | 0.0007   | 0.009    |
| <i>Veillonella rogosae</i> disp.ar <i>parvula atypica</i>                                                                                                                        | DT/LBT | 0.38  | 2.94E-06 | 0.0002   |
| <i>Granulicatella adiacens</i>                                                                                                                                                   | DT/LBT | 0.39  | 2.13E-06 | 0.0001   |
| <i>Streptococcus australis</i>                                                                                                                                                   | DT/LBT | 0.39  | 0.001    | 0.01     |
| <i>Streptococcus</i> sp. HMT066 sp. HMT074                                                                                                                                       | DT/LBT | 0.40  | 0.003    | 0.03     |
| <i>Gemella sanguinis</i>                                                                                                                                                         | DT/LBT | 0.43  | 0.0009   | 0.01     |
| <i>Neisseria perflava flavescens</i>                                                                                                                                             | DT/LBT | 0.43  | 0.02     | 0.08     |
| <i>Streptococcus</i> sp. HMT061                                                                                                                                                  | DT/LBT | 0.48  | 1.60E-05 | 0.0005   |
| <i>Streptococcus infantis</i> clade 638                                                                                                                                          | DT/LBT | 0.50  | 0.0002   | 0.003    |
| <i>Veillonella denticariosi</i>                                                                                                                                                  | DT/LBT | 0.50  | 1.30E-05 | 0.0004   |
| <i>Rothia mucilaginsa</i>                                                                                                                                                        | DT/LBT | 0.51  | 5.85E-09 | 7.81E-07 |
| <i>Actinomyces</i> sp. HMT180 <i>odontolyticus</i>                                                                                                                               | DT/LBT | 0.52  | 2.91E-05 | 0.0008   |
| <i>Streptococcus parasanguinis</i> clade 411                                                                                                                                     | DT/LBT | 0.53  | 0.0003   | 0.004    |
| <i>Streptococcus parasanguinis</i> clade 411 <i>parasanguinis</i> clade 721                                                                                                      | DT/LBT | 0.56  | 5.99E-05 | 0.001    |
| <i>Veillonella rogosae</i> dispar                                                                                                                                                | DT/LBT | 0.84  | 3.51E-09 | 6.00E-07 |

|                                                                  |        |       |          |          |
|------------------------------------------------------------------|--------|-------|----------|----------|
| <i>Rothia aeria dentocariosa</i>                                 | DT/LBT | 0.85  | 5.41E-12 | 5.06E-09 |
| <i>Gemella haemolysans</i>                                       | FOM/VT | -0.56 | 0.01     | 0.06     |
| <i>Streptococcus constellatus intermedius</i>                    | FOM/VT | -0.56 | 0.003    | 0.03     |
| <i>Parvimonas micra</i>                                          | FOM/VT | -0.49 | 0.005    | 0.04     |
| <i>Porphyromonas endodontalis</i>                                | FOM/VT | -0.40 | 0.02     | 0.08     |
| <i>Fusobacterium nucleatum subsp. vincentii</i>                  | FOM/VT | -0.39 | 0.02     | 0.09     |
| <i>Streptococcus infantis</i> clade 431                          | FOM/VT | -0.37 | 0.02     | 0.08     |
| <i>Peptostreptococcaceae</i> XI G5 <i>saphenum</i>               | FOM/VT | -0.35 | 0.01     | 0.08     |
| <i>Mycoplasma faucium</i>                                        | FOM/VT | -0.34 | 0.004    | 0.03     |
| <i>Mogibacterium timidum</i>                                     | FOM/VT | -0.32 | 0.007    | 0.04     |
| <i>Treponema socranskii</i>                                      | FOM/VT | -0.21 | 0.02     | 0.09     |
| <i>Bulleidia extructa</i>                                        | FOM/VT | -0.17 | 0.02     | 0.09     |
| <i>Actinomyces</i> sp. HMT448                                    | FOM/VT | 0.26  | 0.02     | 0.09     |
| <i>Streptococcus gordonii</i>                                    | Former | -0.62 | 0.0006   | 0.008    |
| <i>Parvimonas micra</i>                                          | PAL    | -0.62 | 0.0010   | 0.01     |
| <i>Rothia dentocariosa</i>                                       | PAL    | -0.61 | 0.0006   | 0.008    |
| <i>Fusobacterium nucleatum</i> subsp. <i>vincentii</i>           | PAL    | -0.57 | 0.001    | 0.02     |
| <i>Streptococcus constellatus intermedius</i>                    | PAL    | -0.50 | 0.01     | 0.07     |
| <i>Porphyromonas endodontalis</i>                                | PAL    | -0.50 | 0.006    | 0.04     |
| <i>Prevotella denticola</i>                                      | PAL    | -0.48 | 0.006    | 0.04     |
| <i>Fusobacterium nucleatum</i> subsp. <i>vincentii</i> naviforme | PAL    | -0.46 | 0.02     | 0.09     |
| <i>Haemophilus haemolyticus</i>                                  | PAL    | -0.46 | 0.0008   | 0.009    |
| <i>Peptostreptococcaceae</i> XI G9 <i>brachy</i>                 | PAL    | -0.41 | 0.01     | 0.07     |
| <i>Campylobacter gracilis</i>                                    | PAL    | -0.39 | 0.01     | 0.08     |
| <i>Leptotrichia hongkongensis</i>                                | PAL    | -0.37 | 0.004    | 0.03     |
| <i>Eikenella corrodens</i>                                       | PAL    | -0.34 | 0.009    | 0.06     |
| <i>Tannerella</i> sp. HMT286                                     | PAL    | -0.34 | 0.01     | 0.08     |
| <i>Tannerella forsythia</i>                                      | PAL    | -0.33 | 0.008    | 0.05     |
| <i>Mogibacterium timidum</i>                                     | PAL    | -0.31 | 0.01     | 0.07     |
| <i>Peptostreptococcaceae</i> XI G4 <i>bacterium</i> HMT369       | PAL    | -0.28 | 0.02     | 0.09     |
| <i>Abiotrophia defectiva</i>                                     | PAL    | -0.27 | 0.006    | 0.04     |
| <i>Veillonella</i> sp. HMT780                                    | PAL    | -0.27 | 0.007    | 0.05     |
| <i>Prevotella</i> sp. HMT300                                     | PAL    | -0.26 | 0.007    | 0.04     |
| <i>Campylobacter rectus</i>                                      | PAL    | -0.18 | 0.02     | 0.08     |
| <i>Porphyromonas catoniae</i>                                    | PAL    | -0.14 | 0.02     | 0.09     |
| <i>Actinomyces</i> sp. HMT180                                    | PAL    | 0.22  | 0.02     | 0.09     |
| <i>Megasp. haera</i> <i>micronuciformis</i>                      | PAL    | 0.25  | 0.02     | 0.09     |
| <i>Stomatobaculum longum</i>                                     | PAL    | 0.30  | 0.01     | 0.08     |
| <i>Rothia mucilaginosa</i>                                       | PAL    | 0.32  | 0.01     | 0.06     |
| <i>Lachnoanaerobaculum orale</i>                                 | PAL    | 0.34  | 0.0003   | 0.004    |
| <i>Actinomyces</i> sp. HMT172 sp. HMT180 <i>odontolyticus</i>    | PAL    | 0.39  | 0.02     | 0.08     |
| <i>Selenomonas</i> sp. HMT136                                    | PAL    | 0.40  | 2.76E-05 | 0.0007   |
| <i>Peptostreptococcaceae</i> XI G1 <i>sulci</i>                  | PAL    | 0.41  | 2.09E-07 | 1.78E-05 |
| <i>Streptococcus salivarius</i>                                  | PAL    | 0.42  | 7.02E-05 | 0.001    |
| <i>Stomatobaculum</i> sp. HMT097                                 | PAL    | 0.43  | 0.0003   | 0.005    |
| <i>Leptotrichia</i> sp. HMT215                                   | PAL    | 0.45  | 0.003    | 0.03     |

|                                                                                |     |      |          |          |
|--------------------------------------------------------------------------------|-----|------|----------|----------|
| <i>Oribacterium sinus</i>                                                      | PAL | 0.46 | 0.007    | 0.04     |
| <i>Campylobacter concisus</i>                                                  | PAL | 0.46 | 0.004    | 0.03     |
| <i>Streptococcus vestibularis thermophilus salivarius</i>                      | PAL | 0.47 | 0.0001   | 0.003    |
| <i>Actinomyces lingnae</i> Not Validly Published                               | PAL | 0.49 | 0.002    | 0.02     |
| <i>Streptococcus parasanguinis</i> clade 411                                   | PAL | 0.50 | 0.02     | 0.09     |
| <i>Saccharibacteria</i> TM7 G1 <i>bacterium</i> HMT352                         | PAL | 0.55 | 6.95E-06 | 0.0003   |
| <i>Prevotella melaninogenica</i>                                               | PAL | 0.55 | 0.004    | 0.04     |
| <i>Streptococcus parasanguinis</i> clade 411<br><i>parasanguinis</i> clade 721 | PAL | 0.55 | 0.007    | 0.04     |
| <i>Streptococcus</i> sp. HMT057                                                | PAL | 0.56 | 0.004    | 0.03     |
| <i>Prevotella</i> sp. HMT305                                                   | PAL | 0.56 | 4.27E-10 | 9.97E-08 |
| <i>Actinomyces</i> sp. HMT172                                                  | PAL | 0.60 | 0.0006   | 0.008    |
| <i>Leptotrichia</i> sp. HMT417                                                 | PAL | 0.61 | 0.007    | 0.04     |
| <i>Solobacterium moorei</i>                                                    | PAL | 0.62 | 5.78E-05 | 0.001    |
| <i>Actinomyces odontolyticus</i>                                               | PAL | 0.63 | 0.0005   | 0.007    |
| <i>Prevotella</i> sp. HMT313                                                   | PAL | 0.65 | 0.0006   | 0.008    |
| <i>Mogibacterium diversum</i>                                                  | PAL | 0.65 | 4.69E-06 | 0.0002   |
| <i>Alloprevotella</i> sp. HMT308                                               | PAL | 0.66 | 6.25E-05 | 0.001    |
| <i>Streptococcus parasanguinis</i> clade 721                                   | PAL | 0.67 | 0.0006   | 0.008    |
| <i>Actinomyces graevenitzi</i>                                                 | PAL | 0.68 | 0.0010   | 0.01     |
| <i>Prevotella pallens</i>                                                      | PAL | 0.72 | 1.08E-05 | 0.0004   |
| <i>Atopobium parvulum</i>                                                      | PAL | 0.81 | 3.93E-05 | 0.0009   |
| <i>Prevotella salivae</i>                                                      | PAL | 0.96 | 1.27E-05 | 0.0004   |
| <i>Prevotella histicola</i>                                                    | PAL | 1.05 | 1.66E-05 | 0.0005   |

## 14 Appendix C Patient information leaflet

### Patient Information Leaflet

#### **A study to understand the microbiology of oral leukoplakia (white patches), erythroplakia (red patches), erythroleukoplakia (mixed white and red patches) and squamous cell carcinoma (mouth cancer)**

**1. Introduction:** Oral leukoplakia, erythroplakia and erythroleukoplakia are poorly understood conditions that appear as white and/or red patches inside the mouth. They are associated with an increased risk of developing oral cancer. While it is known that these patches develop much more commonly in people who smoke and drink alcohol, they also develop in non-smokers and non-drinkers. It is possible that they could arise as a result of changes in the types of bacteria normally present in the mouth. Scientists refer to these bacteria as the “microbiome” and research is showing that changes to the normal, healthy microbiome can affect health.

Researchers here in the Dublin Dental University Hospital are investigating the possibility that changes in the oral microbiome might be responsible for oral leukoplakia, erythroplakia, erythroleukoplakia or cancer. We are using a new technique that allows us to identify all of the bacteria in a patient’s mouth and it is hoped that this new technique will identify changes in the microbiome associated with these lesions.

**2. Procedures:** This study requires the participation of patients presenting with oral leukoplakia, erythroplakia, erythroleukoplakia and cancer and patients with no signs of oral disease.

If you agree to take part in this study, you will be asked to give a sample from an oral leukoplakia/erythroplakia/erythroleukoplakia/cancer and/or from some normal oral tissue. The samples will be taken by gently swabbing the surface of the mouth to remove some of the bacteria at the site. This is a painless procedure which does not damage the lining of the mouth or result in bleeding. The plaque and calculus on your teeth, will also be recorded as this is likely to influence the bacteria present on the lining of your mouth.

The samples will be anonymised and will be held in a locked freezer in an access-controlled laboratory in the Dublin Dental University Hospital until analysis. All data will be retained on a password-protected computer at the DDUH. These data will be retained for up to 10 years after which time it will be destroyed.

**3. Benefits:** You will receive no direct benefit from this study. We hope that the study will benefit patients suffering from oral leukoplakia, erythroplakia, erythroleukoplakia or cancer in the future by allowing clinicians to detect it earlier and to develop new treatments.

**4. Risks:** The sampling procedure does not present any risk to the patient.

**5. Exclusion from participation:** Patients will be excluded if they are less than 18 years of age, suffer from diabetes, Crohn's disease, or ulcerative colitis, have/had bowel or stomach cancer, have a dry mouth, have any colds or flu, have taken antibiotics in the last 2 months or have taken systemic, inhaled or local oral immunosuppressive agents in the last 2 months.

**6. Alternative treatment:** You do not have to be a part of this study to be treated.

**7. Confidentiality:** Your identity will remain confidential. Your name will not be published and will not be disclosed to anyone outside the research team.

**8. Processing of personal data:** Your Dental Hospital record may be accessed by the consultants involved in your care, in relation to this study. This may be to clarify details such as age, gender and alcohol and cigarette consumption. Any personal data used in the study will be anonymised.

**9. Voluntary Participation:** Participation in this study is completely voluntary. You may quit at any time. If you decide not to participate, or if you quit, you will not be penalised and will not give up any benefits you had before entering the study.

**10. Stopping the study:** Your doctor may stop your participation in the study at any time without your consent.

**11. Permission:** This study has been approved by the Joint St James'/Tallaght Research Ethics Committee.

**12. Further information:** *You can get more information or answers to your questions about the study, your participation in the study, and your rights, from Dr Sheila Galvin or Prof Claire Healy who can be telephoned at (01) 6127314. If your doctor learns of important new information that might affect your desire to remain in the study, he or she will tell you.*



## 15 Appendix D Patient consent form

**Title of research study:**

**An analysis of the microbiology of oral leukoplakia (white patches), erythroplakia (red patches), erythroleukoplakia (mixed white and red patches) and squamous cell carcinoma (mouth cancer)**

This study and this consent form have been explained to me. My doctor has answered all my questions to my satisfaction. I believe I understand what will happen if I agree to be part of this study.

I have read, or had read to me, this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction. I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights. I have received a copy of this agreement.

I consent to the taking of swabs from my mouth.

☐

I consent to the accessing of my Dental Hospital records by the consultants involved in my care in relation to this study.

☐

I consent to the recording of dental plaque and calculus levels.

☐

**PARTICIPANT'S NAME:**

**PARTICIPANT'S SIGNATURE:**

**Date:**

**Date on which the participant was first furnished with this form:**

**Statement of investigator's responsibility:** I have explained the nature, purpose, procedures, benefits, risks of, or alternatives to, this research study. I have offered to answer any questions and fully answered such questions. I believe that the participant understands my explanation and has freely given informed consent.

**Physician's signature:**

**Date:**

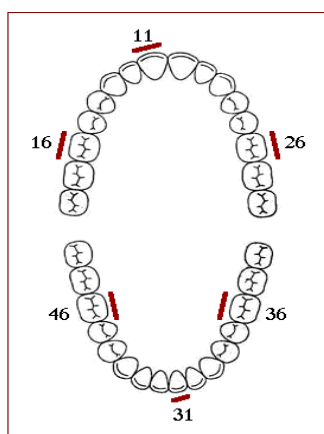
(Keep the original of this form in the participant's medical record, give one copy to the participant, keep one copy in the investigator's records)

## 16 Appendix E Data recording form

|                                                                                     |  |
|-------------------------------------------------------------------------------------|--|
| Patient name                                                                        |  |
| DDUH number                                                                         |  |
| Gender                                                                              |  |
| Age                                                                                 |  |
| Clinical features                                                                   |  |
| Biopsy results                                                                      |  |
| Smoker (Never, previous, current)                                                   |  |
| Alcohol (units/week)                                                                |  |
| Tooth brushing frequency                                                            |  |
| Mouthwash use – yes/no<br>Frequency and brand                                       |  |
| Denture wearer – yes/no<br>Removed at night – yes/no<br>Denture cleaning agent used |  |
| <b>Exclusion criteria</b>                                                           |  |
| Diabetes                                                                            |  |
| Crohn's disease/ulcerative colitis                                                  |  |
| History gastric or bowel cancer                                                     |  |
| Current cold or flu                                                                 |  |
| Antibiotics in last 2 months                                                        |  |
| Systemic, inhaled, local steroids in last 2 months                                  |  |
| Dry mouth                                                                           |  |
| SWABS TAKEN – STATE SITE and LESION TYPE                                            |  |
| SWAB TAKEN NORMAL – STATE SITE                                                      |  |
| OHI-S                                                                               |  |
| Number of teeth present                                                             |  |

## 17 Appendix F OHI-S

| Score | Criteria for classifying debris                                                                                                                                                                                              |
|-------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 0     | No debris or stain present                                                                                                                                                                                                   |
| 1     | Soft debris covering not more than one third of the tooth surface, or presence of extrinsic stains without other debris regardless of surface area covered                                                                   |
| 2     | Soft debris covering more than one third, but not more than two thirds, of the exposed tooth surface.                                                                                                                        |
| 3     | Soft debris covering more than two thirds of the exposed tooth surface.                                                                                                                                                      |
| Score | Criteria for classifying calculus                                                                                                                                                                                            |
| 0     | No calculus present                                                                                                                                                                                                          |
| 1     | Supragingival calculus covering not more than third of the exposed tooth surface.                                                                                                                                            |
| 2     | Supragingival calculus covering more than one third but not more than two thirds of the exposed tooth surface or the presence of individual flecks of subgingival calculus around the cervical portion of the tooth or both. |
| 3     | Supragingival calculus covering more than two third of the exposed tooth surface or a continuous heavy band of subgingival calculus around the cervical portion of the tooth or both.                                        |



Note:  
 If 11/31 absent, use 21/41  
 If 6s absent use 7s or 8s

### Calculation:

Score for debris ÷ no of surfaces scored

Score for calculus ÷ no of surfaces scored

OHI – S = Debris score + calculus score (max 6)

## 18 Appendix G Published Review



## OPEN ACCESS

## EDITED BY

Arun Singh,  
Tata Memorial Hospital, India

## REVIEWED BY

Ashu Sharma,  
University at Buffalo, United States

## \*CORRESPONDENCE

Sheila Galvin  
✉ sheila.galvin@dental.tcd.ie

## SPECIALTY SECTION

This article was submitted to Oral Cancers, a section of the journal Frontiers in Oral Health

RECEIVED 14 February 2023

ACCEPTED 01 March 2023

PUBLISHED 23 March 2023

## CITATION

Galvin S, Moran GP and Healy CM (2023)  
Influence of site and smoking on malignant  
transformation in the oral cavity: Is the  
microbiome the missing link?  
Front. Oral. Health 4:1166037.  
doi: 10.3389/froh.2023.1166037

## COPYRIGHT

© 2023 Galvin, Moran and Healy. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Influence of site and smoking on malignant transformation in the oral cavity: Is the microbiome the missing link?

Sheila Galvin<sup>1\*</sup>, Gary P. Moran<sup>2</sup> and Claire M. Healy<sup>1</sup>

<sup>1</sup>Division of Oral and Maxillofacial Surgery, Oral Medicine and Oral Pathology, School of Dental Science, Trinity College Dublin, Dublin Dental University Hospital, Dublin, Ireland, <sup>2</sup>Division of Oral Biosciences, School of Dental Science, Trinity College Dublin, Dublin Dental University Hospital, Dublin, Ireland

The tongue and floor of the mouth are high-risk sites for oral squamous cell carcinoma (OSCC), while smoking is its most significant risk factor. Recently, questions have been raised as to the role of the oral microbiome in OSCC because of a wealth of evidence demonstrating that the microbiome of OSCC differs from that of healthy mucosa. However, oral site and smoking also have a significant impact on oral microbial communities, and to date, the role these factors play in influencing the dysbiotic microbial communities of OSCC and precursor lesions has not been considered. This review aims to examine the influence of site and smoking on the oral microbiome and, in turn, whether these microbiome changes could be involved in oral carcinogenesis.

## KEYWORDS

oral cancer, oral microbiome, OSCC, smoking, oral leukoplakia, oral potentially malignant disorders

## 1. Introduction: micro-organisms and cancer

Micro-organisms are estimated to cause up to 20% of fatal human malignancies, with approximately 2.2 million infection-attributed cancers diagnosed worldwide in 2018 (1). To date, 11 microbes have been designated as human carcinogens by the International Agency for Cancer Research (IARC): *Helicobacter pylori*, Hepatitis B and C viruses, *Opisthorchis viverrini*, *Clonorchis sinensis*, Human papillomavirus, Epstein-Barr virus, Human herpes virus 8, Human T-cell Lymphotropic Virus 1, *Schistosoma haematobium*, and Human Immunodeficiency Virus-1 (2). Recent advances in microbiome research, however, have demonstrated that many other microbial species are likely to play a role in carcinogenesis, including the oral bacterial pathogen *Fusobacterium nucleatum*, which is increasingly being linked with colorectal carcinoma (3).

There is general agreement that the microbiome of oral squamous cell carcinoma (OSCC) differs from that of healthy tissue, and there is emerging evidence that pre-cancerous lesions also harbour different microbial communities to healthy mucosa (4–10). There is also a consensus that the mucosal site and tobacco use have a significant impact on the composition of the oral microbiome (11–17). However, what has not been considered to date is the role these factors play in influencing the dysbiotic microbial communities of OSCC and precursor lesions and whether these microbiome changes are involved in malignant progression. This mini review will, therefore, examine the influence of site and smoking on the oral microbiome and, in turn, the evidence for any related dysbiosis in oral carcinogenesis.

## 2. Risk factors of OSCC

Established risk factors for OSCC are tobacco and alcohol consumption, betel nut use, age and presence of an oral potentially malignant disorder (OPMD), a group of mucosal conditions associated with a statistically increased risk of transformation to cancer. Tobacco is the most significant of these risk factors, with up to 25% of oral cancers directly attributable to smoking, the risk of OSCC increasing with the number of cigarettes smoked, and up to 80% of OSCCs occurring in smokers (18, 19). Tobacco and alcohol are known to act synergistically, increasing the risk of oral cancer up to 15-fold, particularly in the floor of the mouth (18, 20). However, up to 15% of OSCCs in older adults and up to 26% of OSCCs in younger adults have little to no exposure to tobacco or alcohol (21, 22), with Deneuve et al. also finding that their non-smoking tongue cancer cohort had a significantly higher prevalence of oral leukoplakia (OLK) than their smoking cohort (21). OLK is the most common OPMD with a global estimated incidence of 4.1% and malignant transformation rate of 9.8% (23). Risk factors for the development of OLK include the same risk factors as oral cancer, namely tobacco smoking, consumption of alcohol, and use of betel nut. However, OLKs in non-smokers are generally considered to be at a higher risk of malignant transformation than OLKs in smokers (24, 25).

Finally, epidemiological studies and meta-analyses have associated poor oral hygiene, tooth loss, and periodontal disease with OSCC (26–30). However, it is still unclear as to whether these associations are due to chronic periodontal disease sharing risk factors with OSCC, namely smoking and alcohol use, or to the microbial dysbiosis and/or inflammation associated with these conditions.

## 3. Microbiome of OSCC and OPMD

There is a wealth of literature confirming that the microbiome of OSCC is different from that of healthy mucosa. However, the literature differs as to what microbial changes are responsible for these differences, the conflicting evidence being due at least in part to the various methods by which the oral microbiome is sampled, including mucosal swab, tissue biopsy, and saliva collection. Mucosal surface swabs are non-invasive and enable direct and targeted sampling of specific oral mucosal surfaces. Their principal advantage, however, is that they allow for separate sampling of lesional and matched normal sites within the same individual and therefore control of confounding factors such as smoking, alcohol consumption, and oral hygiene. Direct tissue sampling enables the investigation of invasive species within oral lesions. This method of sampling is, however, invasive, requiring mucosal biopsy, and so carries risks for patients such as post-operative pain, bleeding, and infection. Therefore, in recent studies, ethical considerations have prevented the biopsy sampling of normal tissue as a control. Nonetheless, tissue microbiome studies of oral cancers have

shown interesting results, particularly in relation to *Fusobacterium nucleatum*. Finally, the ease of collection of saliva makes it a popular choice for microbiome studies. However, saliva collection does not allow for sampling of specific sites/lesions and has also been shown to be strongly biased towards tongue and palate communities (12, 13, 31). The saliva sampled can also be contaminated with food and other debris, and although there are various protocols described for pre-sampling rinsing to minimise contamination, they also provide for different sampling methods among studies.

Studies have also used different methods, which have evolved significantly over the years, to characterise the oral microbiome, from initial culture-based methods to DNA-DNA hybridisation and more recently to next-generation sequencing (NGS)-based methods such as marker gene sequencing and metagenomics. Marker gene sequencing usually involves sequencing specific variable (V) regions of the highly conserved bacterial 16S rRNA gene to identify bacteria present in a sample, whereas metagenomic methods sequence the entire genome and therefore can provide not only taxonomy but also the metabolic pathways and functional profiles of the microbial community sampled (32, 33). Metagenomic methods are currently limited by cost and accessibility, but due to the kind of information they can provide, it is likely that they will supersede marker gene sequencing in time. It should be noted that the choice of the variable region (V1–V9) to sequence also needs consideration. Regions V1–V5 are typically used for the oral microbiota, with recent reports suggesting that the V1–V3 region is preferable to the more commonly used V3–V4 region as it is more divergent and therefore provides more phylogenetic resolution and a more accurate assessment of population diversity (34, 35). Several platforms are available for sequencing, of which Illumina remains the most commonly used due to its wide availability, high output, and high level of accuracy (33).

The majority of the microbiome studies on OSCC and OLK to date have sequenced various parts of the V1–V5 region of the 16S rRNA gene using NGS. While the choice of region can impact phylogenetic resolution and diversity as mentioned previously, the choice of the sampling method has an even greater impact on results. Studies using swab and biopsy sampling methods have found an increased abundance of *Fusobacterium* and reduced abundance of *Streptococcus* and *Rothia* at OSCC sites (4, 5, 36–38), while some of those sampling saliva have identified Firmicutes, *Streptococci*, and *Rothia* as more abundant in OSCC patients (7, 39, 40). However, no OSCC microbiome studies to date have controlled for significant influences such as smoking, oral site, or oral hygiene, all factors known to influence the oral microbiome and that need to be considered before any definite conclusions can be drawn on the influence of the microbiome on the development of OSCC.

To date, only two authors have looked at the microbiome of OPMDs sampled by swabbing. Schmidt et al. (4) demonstrated that abundances of Firmicutes, Actinobacteria, and *Streptococcus* spp. were significantly reduced at OPMD sites relative to contralateral normal sites, while Amer et al. (10) found a significant enrichment of OLKs with *Fusobacterium*, *Leptotrichia*,

*Rothia mucilaginosa*, and *Campylobacter* spp. relative to contralateral normal sites. However, Amer et al. also demonstrated that site and smoking status had a more significant influence on the microbiome than the presence of an OLK. Changes in the abundance of *Fusobacterium nucleatum* subsp. *vincentii* were particularly site-dependent with increased abundance found on buccal OLKs and on lateral tongue contralateral normal sites. Lateral tongue normal sites also showed an enrichment of *Rothia mucilaginosa*. These data highlight the necessity for control samples to be matched by site, which is possible only by mucosal swab sampling.

## 4. Influence of site on the microbiome

The tongue is the most common intra-oral site for cancer, accounting for approximately 46.7% of oral cavity (excluding salivary gland) cancers in the United States in 2016 (41). Despite an overall decrease in tobacco and alcohol-associated OSCC in the US between 1985 and 2009, the proportion of tongue cancers in adults under 50 years of age increased over the same 25-year period (42, 43).

Unsurprisingly, the tongue is also the most common, or second most common, site of OLK and is considered a high-risk site for malignant transformation (23, 44, 45), with the most recent systematic review on the topic finding that 56.3% of OLKs that transformed to cancer were on the tongue (23). A slightly older systematic review identified the buccal mucosa (18.4%) and the tongue (16.4%) as the most common sites for OLK (45), with the authors commenting that the site involved was related to the lifestyle factors of populations studied; the buccal mucosa is more likely to be involved in regions where betel nut is used, whereas the tongue and floor of mouth are more commonly involved in Western populations where smoking and alcohol consumption are more prevalent. As stated previously, malignant transformation was most common in OLKs affecting the tongue (24.2%), followed by the tongue/floor of the mouth combined (14.9%), while the buccal mucosa had the lowest rate of malignant transformation (3.5%).

The oral microbiome also shows site specificity (11–13). Prior to the advent of next-generation sequencing (NGS), Mager et al. (12) used checkerboard DNA-DNA hybridisation to investigate the microbiomes of saliva, dorsum, ventral and lateral tongue, maxillary gingiva, floor of mouth, buccal mucosa, labial mucosa, and hard palate in 225 individuals. Saliva and lateral and dorsal tongue samples clustered together, with the other sites forming a second cluster. There were significantly higher proportions of Gram-negative, and mostly anaerobic, organisms such as *Veillonella parvula*, *Prevotella melaninogenica*, *Eikenella corrodens*, *Neisseria mucosa*, *Actinomyces odontolyticus*, *Fusobacterium periodonticum*, *Fusobacterium nucleatum* subsp. *vincentii*, and *Porphyromonas gingivalis* in the saliva/dorsal/lateral tongue cluster, while the second cluster was dominated by Gram-positive aerobes *Streptococcus mitis* and *Streptococcus oralis* and a single Gram-negative anaerobe, *Selenomonas noxia*. Aas et al.'s (11) findings, pre-NGS using 16S rRNA cloning, were broadly

similar, although the dominant colonisers of the tongue differed, likely because of the different methodologies used for bacterial identification.

Finally, Segata et al. (13) sampled 200 subjects at seven oral sites: buccal mucosa, attached gingiva, hard palate, saliva, tongue, and two tooth surfaces along with tonsils, throat, and stool. Clustering similar to that in Mager et al.'s study was identified with buccal mucosa, gingiva, and hard palate clustered together (Group 1), while saliva and tongue grouped with throat and tonsils (Group 2). The Group 1 cluster consisted mostly of Firmicutes, followed by Proteobacteria, Bacteroidetes, Actinobacteria, and Fusobacteria, while the Group 2 cluster had a reduced relative abundance of Firmicutes and increased abundance of Bacteroidetes, Fusobacteria, Actinobacteria, and TM7. At the genera level, Group 1 was dominated by *Streptococcus*, while Group 2 had a more even distribution of *Streptococcus*, *Veillonella*, *Prevotella*, *Neisseria*, *Fusobacterium*, *Actinomyces*, and *Leptotrichia*.

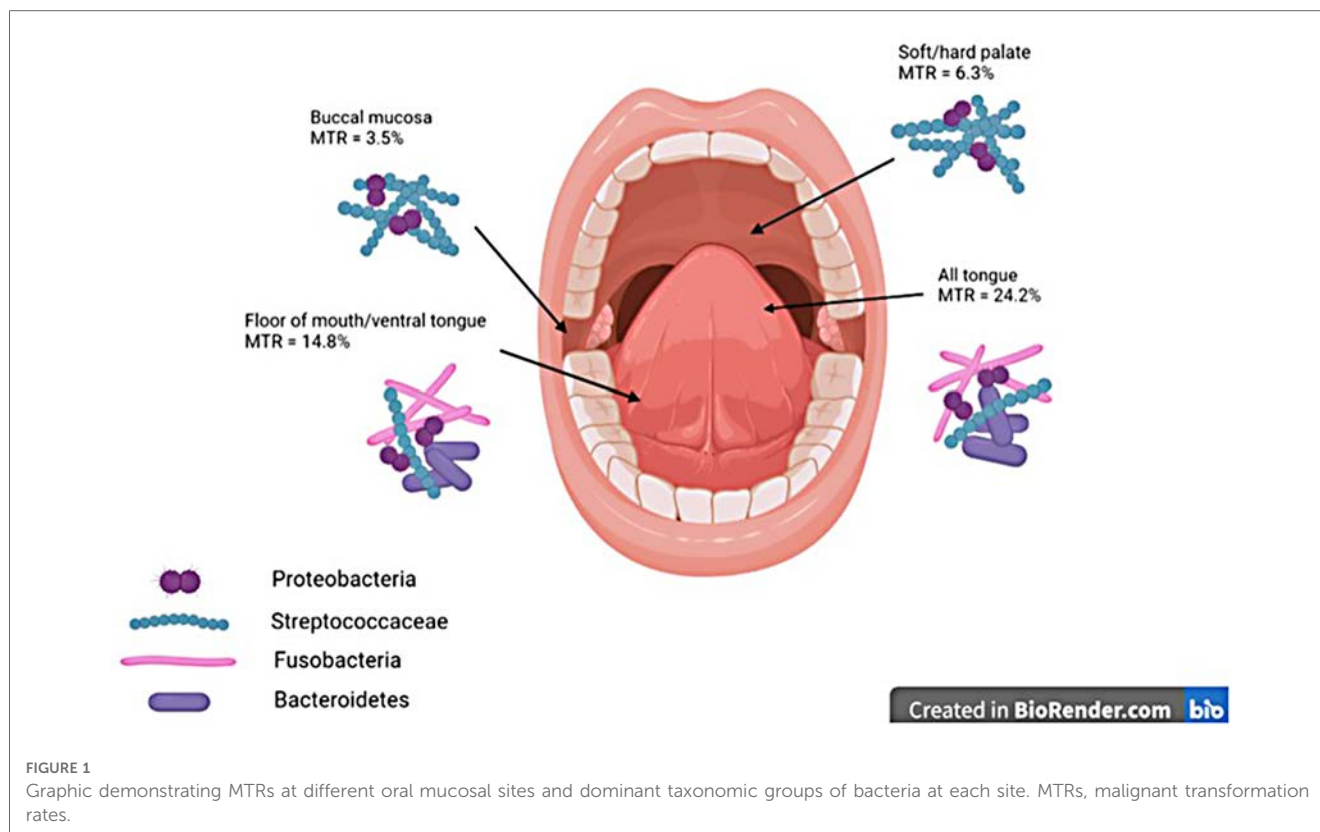
It is thus clear that the tongue, a site known to have high rates of malignant transformation, also has high levels of colonisation with Gram-negative organisms of phyla Fusobacteria and Bacteroidetes, whereas the buccal mucosa and palate, which have much lower rates of malignant transformation, are predominantly colonised by Gram-positive *Streptococci* (Figure 1).

## 5. Influence of smoking, betel nut, and electronic cigarettes on the oral microbiome

Along with site, smoking is also known to have a profound effect on the oral microbiome, although as for OSCC microbiome studies, results differ because of variations in sampling methodology and failure to control for factors such as periodontal disease for which smoking is an established risk factor. Depletion of *Neisseria* spp. in smokers is widely reported and many of these studies also report an increase in Fusobacteria (Table 1) (10, 14–17, 46–48). A switch to a community rich in Fusobacteria may increase the inflammatory stimulation on mucosal surfaces; *Fusobacterium nucleatum*, in particular, has been shown to induce IL-6-mediated inflammation (49). However, smoking is also a risk factor for periodontal disease, which, in turn, can increase the burden of Gram-negative bacteria in the oral cavity and may explain why some studies identify increases in *Fusobacterium nucleatum* in smokers. What influence, if any, this has on OSCC risk or malignant transformation of OLK has yet to be elucidated.

To date, few studies have investigated the influence of betel nut or electronic cigarette (e-cigarettes) use on the oral microbiome. Hernandez et al. (50) found significantly increased abundances of *Streptococcus infantis* and reduced abundances of *Parascardovia* and other *Streptococci* in current betel nut chewers, while Hsiao et al. (16) associated betel nut use with an increased abundance of *Prevotella intermedia*. E-cigarette use has been associated with increased abundances of *Veillonella* and *Porphyromonas* (51–54).





## 6. Microbial methods of carcinogenesis

To date, only one study has shown that bacteria may induce malignant transformation in the oral mucosa, where *Porphyromonas gingivalis* and *Fusobacterium nucleatum* were shown to promote carcinogenesis in a chemically induced murine model of tongue OSCC (55). The exact mechanisms of microbial carcinogenesis are still unclear, although research suggests that bacterial-induced epithelial change, promotion of cellular invasion and proliferation, acetaldehyde (ACH) production, and inflammation are likely involved.

The epithelial adhesion molecule E-cadherin, a marker of the epithelial state, is lost early in oral dysplasia (56) and is considered an early event in epithelial-mesenchymal transition (EMT), a biological process whereby an epithelial cell changes to a mesenchymal cell phenotype and is therefore able to migrate, invade, and resist apoptosis. EMT has, therefore, been associated with the initiation of carcinogenesis and facilitation of cancer spread. Bacteria can induce EMT, with *Helicobacter pylori* being a well-known example, inducing EMT *via* its cytotoxin-associated gene A (Cag-A) (57). Long-term infection with *Porphyromonas gingivalis* has also been associated with initial EMT changes in oral epithelial cells, including decreased E-cadherin expression and upregulation of Vimentin (58). Zhang et al. (49) recently showed that the *Fusobacterium nucleatum* adhesin, FadA, can facilitate EMT *via* downregulation and transfer of E-cadherin to the cytoplasm and upregulation of N-cadherin, Vimentin, and SNAI1 and miRNA MIR4435-2HG, thus providing a possible

mechanism for the role of *Fusobacterium nucleatum* in carcinogenesis. *Fusobacterium nucleatum* is also known to promote cell proliferation *via* upregulation of multiple kinases (59), while both *Fusobacterium nucleatum* and *Porphyromonas gingivalis* promote cellular invasion (59, 60).

Microbial production of ACH is also likely to play a role in oral carcinogenesis. ACH, produced by the reduction of ethanol by host or microbial acetaldehyde dehydrogenase (61), is designated as a Class 1 carcinogen by the IARC (62, 63) and is a known risk factor for OSCC. ACH is a genotoxic carcinogen capable of inducing DNA damage and secondary hyperproliferation of epithelium (61, 64–67). Mutagenic levels (>100  $\mu$ M) of ACH can be detected in saliva following the ingestion of 0.5 g alcohol per kg body weight. Smoking also significantly increases salivary ACH production, with active smoking resulting in an additional 200–400  $\mu$ M peak ACH concentration in saliva for the duration of the smoking (64, 68–70). Interestingly, the amount of ACH produced is significantly reduced after 3 days of using a chlorhexidine mouthrinse (68), supporting the role of bacteria in ACH production.

All species of *Neisseria*, along with *Rothia mucilaginosa*, *Streptococcus mitis*, and *Prevotella histicola*, have been shown to produce ACH from ethanol (71–73). As *Neisseria* are known to be decreased in smokers (72), it is likely that other bacteria are responsible for ACH production in this cohort. *Rothia mucilaginosa* strains isolated from dysplastic OLKs have been shown to be capable of producing mutagenic levels of ACH in the presence of alcohol (74), while other studies have shown that certain *Streptococci*, particularly viridans group *Streptococci*, have



TABLE 1 Significant differences between the oral microbiomes of smokers and non-smokers in the literature.

| Author                 | Sampling method and cohort                                                                               | Periodontal health                                                | Increased in smokers                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Decreased in smokers                                                                                                                                                                                                                                                                                                                        |
|------------------------|----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Thomas et al. (2014)   | Mucosal swabs<br>13 smokers<br>9 never smokers                                                           | Not measured                                                      | <i>Fusobacteria</i><br><i>Actinobacteria</i><br><i>Capnocytophaga</i><br><i>Prevotella</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Firmicutes<br><i>Cyanobacteria</i><br>TM7<br>BD1-5<br><i>Neisseria</i><br><i>Peptostreptococcus</i><br><i>Gemella</i><br><i>Granulicatella</i><br><i>Staphylococcus</i>                                                                                                                                                                     |
| Moon et al. (2015)     | Subgingival plaque<br>134 smokers<br>134 non-smokers                                                     | Moderate periodontitis                                            | <i>Fusobacteria</i><br><i>Fusobacterium</i> <i>Streptococcus</i> <i>Veillonella</i><br><i>Corynebacterium</i> TM7<br><i>Filifactor</i><br><i>Parvimonas</i><br><i>Tanarella</i><br><i>Rothia</i>                                                                                                                                                                                                                                                                                                                                                    | Bacteroidetes<br><i>Neisseria</i><br><i>Prevotella</i><br><i>Aggregatibacter</i><br><i>Selenomonas</i><br><i>Veillonellaceae</i><br>GQ422718                                                                                                                                                                                                |
| Mason et al. (2015)    | Subgingival plaque<br>200 subjects                                                                       | Periodontally healthy                                             | <i>Fusobacterium nucleatum</i><br><i>Fusobacterium naviforme</i><br><i>Filifactor alocis</i><br><i>Dialister microaerophile</i> <i>Megasphaera</i><br><i>sueciensis</i><br><i>Megasphaera geminatus</i><br><i>Megasphaera elsdenii</i><br><i>Megasphaera micronuciformis</i><br><i>Acinetobacter johnsonii</i><br><i>Acinetobacter guillouiae</i><br><i>Acinetobacter schindleri</i><br><i>Acinetobacter baumannii</i><br><i>Acinetobacter haemolyticus</i> <i>Pseudomonas</i><br><i>pseudoalcaligenes</i><br><i>Pseudoramibacter alactolyticus</i> | <i>Streptococcus sanguinis</i><br><i>Streptococcus</i><br><i>parasanguinis</i><br><i>Streptococcus oralis</i><br><i>Gemella elegans</i><br><i>Gemella adiacens</i><br><i>Actinomyces viscosus</i><br><i>Actinomyces israelii</i><br><i>Actinomyces dentalis</i><br><i>Neisseria subflava</i><br><i>Haemophilus</i><br><i>parainfluenzae</i> |
| Wu et al. (2016)       | Saliva<br>112 smokers<br>571 former smokers<br>521 never smokers                                         | Not measured, no oral examinations                                | <i>Actinobacteria</i><br>Firmicutes<br><i>Atopobium</i> <i>Streptococcus</i><br><i>Bifidobacterium</i> <i>Lactobacillus</i><br><i>Veillonella</i>                                                                                                                                                                                                                                                                                                                                                                                                   | Proteobacteria<br><i>Neisseria</i><br><i>Haemophilus</i><br><i>Aggregatibacter</i><br><i>Capnocytophaga</i><br><i>Corynebacterium</i><br><i>Porphyromonas</i><br><i>Prevotella</i><br><i>Peptostreptococcus</i><br><i>Leptotrichia</i><br><i>Abiotrophia</i><br><i>Selenomonas</i>                                                          |
| Amer et al. (2017)     | Mucosal swabs (OLK study, subset analysis)<br>19 current smokers<br>10 former smokers<br>7 never smokers | Not measured                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | <i>Fusobacterium nucleatum</i><br><i>Neisseria</i><br><i>Leptotrichia</i>                                                                                                                                                                                                                                                                   |
| Hsiao et al. (2018)    | Saliva<br>104 male smokers<br>47 male never smokers                                                      | Not measured                                                      | <i>Fusobacterium nucleatum</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                             |
| Wirth et al. (2020)    | Saliva<br>11 smokers<br>11 non-smokers (8 never, 3 former—>1 year cessation)                             | Dental examinations, moderation and severe periodontitis excluded | <i>Prevotella</i><br><i>Megasphaera</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <i>Neisseria</i><br><i>Capnocytophaga</i><br><i>Porphyromonas</i><br><i>Oribacterium</i>                                                                                                                                                                                                                                                    |
| Gopinath et al. (2021) | Buccal mucosa swabs<br>17 male smokers<br>13 male non-smokers                                            | Not measured, no oral examinations                                | <i>Fusobacterium</i> spp. including<br><i>Fusobacterium nucleatum</i><br><i>Saccharibacterium</i> spp.<br><i>Shuttleworthia</i> spp.                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                             |

OLK, oral leukoplakia.

significant ACH-producing ability. However, *Streptococci* are typically associated with a healthy microbiome and reduced in OPMDs and OSCC (16, 73, 75). Therefore, it is clear that more studies are needed to investigate the role of bacteria in ACH

production and the various factors that influence its production; however, it is apparent that many bacteria within the normal oral microbiome are potent producers of ACH, raising questions about their role in oral carcinogenesis.

Finally, inflammation, which has been recognised as a fundamental component of malignant processes (76), is another potential link between the oral microbiome and OSCC. Chronic infection causes inflammation, and as demonstrated by the associations between hepatitis and hepatocellular carcinoma (HCC) and *Helicobacter pylori*-induced gastritis and gastric cancer, this infection-induced inflammation can precede cancer development (76). The role of infection-associated inflammation in OSCC is more complex as the oral cavity harbours hundreds of bacteria and fungi as commensals without being in a constant state of inflammation. However, polymicrobial dysbiosis in the oral cavity does lead to inflammation, as demonstrated by periodontal disease, and therefore may also play a role in carcinogenesis. Bacterial pathogens can stimulate host immune cells, such as macrophages and neutrophils, and oral epithelial cells, *via* their surface receptors to activate the STAT3 and NF- $\kappa$ B signalling pathways, which, through the promotion of proliferation, angiogenesis, invasion, and enhanced survival, can play a role in tumorigenesis (77, 78). *Fusobacterium nucleatum*, in particular, has been shown to bind and enter host epithelial and endothelial cells and induce NK- $\kappa$ B- and IL-6-mediated inflammation *via* its FadA adhesin molecule (79, 80).

Although the definitive pathways of bacterial carcinogenesis have yet to be elucidated, it is likely that, rather than a single bacteria or pathway, there is a polymicrobial dysbiosis and therefore multiple mechanisms are involved. This dysbiosis is likely to be influenced by smoking habit and it is also possible that due to baseline differences in mucosal site colonisation, some mucosal sites are more susceptible to colonisation by a potentially carcinogenic microbiome.

## 7. Discussion

OLK, the most common OPMD, is most commonly found on the buccal mucosa and tongue. However, OLKs at these sites display very different malignant transformation rates. In health, both sites have distinctly different microbial communities, which raises the question of the role of the microbiome in the malignant transformation of OLKs at these sites. While the microbiome may not initiate dysplastic change, sites with a significantly higher burden of Gram-negative bacteria (e.g., tongue) may be subject to greater pro-inflammatory stimuli, which could play a role in driving abnormal cellular phenotypes. OLKs on the palate and buccal mucosa where *Streptococci*

predominate may not be exposed to the same level of inflammatory stimulation.

Smoking, a major risk factor for oral cancer, also significantly influences the oral microbiome, depleting common oral taxa such as *Neisseria* spp., which may allow pathobionts to proliferate. Smoking may also contribute to the Gram-negative burden through its role in promoting periodontal disease. It is not yet known whether this effect of smoking on the microbiome contributes to oral carcinogenesis. However, it is clear that smoking is a potential confounding factor in establishing the carcinogenic potential of the oral microbiome. In future, site-specific studies, with anatomically matched control sites, in non-smokers and smokers may provide answers to these questions.

## Author contributions

SG drafted and prepared the manuscript. GPM and CMH reviewed and edited the final version. All authors contributed to the article and approved the submitted version.

## Funding

GPM and CMH were supported by an award from the Irish Health Research Board (grant no. ILP-POR-2019-030), which is funding several studies on the oral microbiome.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob Health*. (2019) 8:e180–90. doi: 10.1016/S2214-109X(19)30488-7
2. Carbone A, Blum HE, Cesarman E, Chen CJ, Fang Y, Fontham ETH. Biological agents. Volume 100 B. A review of human carcinogens. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. *IARC Monogr Eval Carcinog Risks Hum*. (2012) 100:1–441.
3. Castellarin M, Warren RL, Douglas Freeman J, Dreolini L, Krzywinski M, Strauss J, et al. *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res*. (2012) 22:299–306. doi: 10.1101/gr.126516.111
4. Schmidt BL, Kuczynski J, Bhattacharya A, Huey B, Corby PM, et al. Changes in abundance of oral microbiota associated with oral cancer. *PLoS One*. (2014) 9:e98741. doi: 10.1371/journal.pone.0098741

5. Zhao H, Chiu M, Huang Z, Yang X, Ran S, et al. Variations in oral microbiota associated with oral cancer. *Sci Rep.* (2017) 7:11773. doi: 10.1038/s41598-017-11779-9
6. Takahashi Y, Park J, Hosomi K, Yamada T, Kobayashi A, et al. Analysis of oral microbiota in Japanese oral cancer patients using 16S rRNA sequencing. *J Oral Biosci.* (2019) 127:512–9. doi: 10.1016/j.job.2019.03.003
7. Guerrero-Preston R, Godoy-Vitorino F, Jedlicka A, Rodriguez-Hilario A, Gonzalez H, et al. 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment. *Oncotarget.* (2016) 7:51320–34. doi: 10.18632/oncotarget.9710
8. Yang C-Y, Yeh Y-M, Yu H-Y, Chin C-Y, Hsu C-W, et al. Oral microbiota community dynamics associated with oral squamous cell carcinoma staging. *Front Microbiol.* (2018) 9:862. doi: 10.3389/fmicb.2018.00862
9. Pushalkar S, Ji X, Li Y, Estilo C, Yegnanarayana R. Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. *BMC Microbiol.* (2012) 12:144. doi: 10.1186/1471-2180-12-144
10. Amer A, Galvin S, Healy CM, Moran GP. The microbiome of potentially malignant oral leukoplakia exhibits enrichment for *Fusobacterium*, *Leptotrichia*, *Campylobacter*, and *Rothia* species. *Front Microbiol.* (2017) 8:2391. doi: 10.3389/fmicb.2017.02391
11. Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol.* (2005) 43:5721–32. doi: 10.1128/JCM.43.11.5721-5732.2005
12. Mager DL, Ximenez-Fyvie LA, Haffajee AD, Socransky SS. Distribution of selected bacterial species on intraoral surfaces. *J Clin Periodontol.* (2003) 30:644–54. doi: 10.1034/j.1600-051X.2003.00376.x
13. Segata N, Haake SK, Mannon P, Lerman KP, Waldron L, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol.* (2012) 13:R42. doi: 10.1186/gb-2012-13-6-r42
14. Moon J-H, Lee J-H, Lee J-Y. Subgingival microbiome in smokers and non-smokers in Korean chronic periodontitis patients. *Mol Oral Microbiol.* (2015) 30:227–41. doi: 10.1111/omi.12086
15. Mason MR, Preshaw PM, Nagaraja HN, Dabdoub SM, Rahman A, et al. The subgingival microbiome of clinically healthy current and never smokers. *ISME J.* (2015) 9:268–72. doi: 10.1038/ismej.2014.114
16. Hsiao J-R, Chang C-C, Lee W-T, Huang C-C, Ou C-Y, et al. The interplay between oral microbiome, lifestyle factors and genetic polymorphisms in the risk of oral squamous cell carcinoma. *Carcinogenesis.* (2018) 39:778–87. doi: 10.1093/carcin/bgy053
17. Thomas AM, Gleber-Netto FO, Fernandes GR, Amorim M, Barbosa LF, et al. Alcohol and tobacco consumption affects bacterial richness in oral cavity mucosa biofilms. *BMC Microbiol.* (2014) 14:250. doi: 10.1186/s12866-014-0250-2
18. Hashibe M, Brennan P, Chuang S-C, Boccia S, Castellsague X, Chen C, et al. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiol Prev Biomarkers.* (2009) 18:541–50. doi: 10.1158/1055-9965.EPI-08-0347
19. Scully C, Bagan J. Oral squamous cell carcinoma: overview of current understanding of aetiopathogenesis and clinical implications. *Oral Dis.* (2009) 15:388–99. doi: 10.1111/j.1601-0825.2009.01563.x
20. Radoi L, Paget-Bailly S, Cyr D, Papadopoulos A, Guida F. Tobacco smoking, alcohol drinking and oral cavity cancer by subsite. *Eur J Cancer Prev.* (2013) 22:268–76. doi: 10.1097/CEJ.0b013e3283592c2c
21. Deneuve S, Guerlain J, Dupres-Bories A, Majoufre C, Philouze P, Ceruse P, et al. Oral tongue squamous cell carcinomas in young patients according to their smoking status: a GETTEC study. *Eur Arch Otorhinolaryngol.* (2022) 279:415–24. doi: 10.1007/s00405-021-06793-7
22. Llewellyn CD, Linklater K, Bell J, Johnson NW, Warnakulasuriya KAAS. Squamous cell carcinoma of the oral cavity in patients aged 45 years and under: a descriptive analysis of 116 cases diagnosed in the South East of England from 1990 to 1997. *Oral Oncol.* (2003) 39:106–14. doi: 10.1016/S1368-8375(02)00026-X
23. Aguirre-Urizar JM, Lafuente-Ibáñez de Mendoza I, Warnakulasuriya S. Malignant transformation of oral leukoplakia: systematic review and meta-analysis of the last 5 years. *Oral Dis.* (2021) 27:1881–95. doi: 10.1111/odi.13810
24. Ho MW, Risk JM, Woolgar JA, Field EA, Field JK, Steele JC, et al. The clinical determinants of malignant transformation in oral epithelial dysplasia. *Oral Oncol.* (2012) 48:969–76. doi: 10.1016/j.oraloncology.2012.04.002
25. Rock LD, Rosin MP, Zhang L, Chan B, Shariati B, Laronde DM. Characterization of epithelial oral dysplasia in non-smokers: first steps towards precision medicine. *Oral Oncol.* (2018) 78:119–25. doi: 10.1016/j.oraloncology.2018.01.028
26. Yao Q-W, Zhou D-S, Peng H-J, Ji P, Liu D-S. Association of periodontal disease with oral cancer: a meta-analysis. *Tumor Biol.* (2014) 35:7073–7. doi: 10.1007/s13277-014-1951-8
27. Zheng T, Boyle P, Hu H, Duan J, Jiang P, Ma D, et al. Dentition, oral hygiene, and risk of oral cancer: a case-control study in Beijing, People's Republic of China. *Cancer Cause Control.* (1990) 1:235–41. doi: 10.1007/BF00117475
28. Ahrens W, Pohlmann H, Foraita R, Nelis M, Lagiou P, Lagiou A, et al. Oral health, dental care and mouthwash associated with upper aerodigestive tract cancer risk in Europe: the ARCA study. *Oral Oncol.* (2014) 50:616–25. doi: 10.1016/j.oraloncology.2014.03.001
29. Javed F, Warnakulasuriya S. Is there a relationship between periodontal disease and oral cancer? A systematic review of currently available evidence. *Crit Rev Oncol Hemat.* (2016) 97:197–205. doi: 10.1016/j.critrevonc.2015.08.018
30. Hashim D, Sartori S, Brennan P, Curado MP, Wunsch-Filho V, Divaris K, et al. The role of oral hygiene in head and neck cancer: results from International Head and Neck Cancer Epidemiology (INHANCE) consortium. *Ann Oncol.* (2016) 27:1619–25. doi: 10.1093/annonc/mdw224
31. Wade WG, Prosdocimi EM. Profiling of oral bacterial communities. *J Dent Res.* (2020) 99:621–9. doi: 10.1177/0022034520914594
32. Deo PN, Deshmukh R. Oral microbiome: unveiling the fundamentals. *J Oral Maxillofac Pathology.* (2019) 23:122–8. doi: 10.4103/jomfp.JOMFP\_77\_18
33. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol.* (2017) 35:833–44. doi: 10.1038/nbt.3935
34. Kim M, Morrison M, Yu Z. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J Microbiol Meth.* (2011) 84:81–7. doi: 10.1016/j.mimet.2010.10.020
35. Zheng W, Tsompana M, Ruscitto A, Sharma A, Genco R. An accurate and efficient experimental approach for characterization of the complex oral microbiota. *Microbiome.* (2015) 3(48). doi: 10.1186/s40168-015-0110-9
36. Al-hebshi NN, Nasher AT, Maryoud MY, Homeida HE, Chen T, et al. Inflammatory bacteriome featuring *Fusobacterium nucleatum* and *Pseudomonas aeruginosa* identified in association with oral squamous cell carcinoma. *Sci Rep.* (2017) 7:1834. doi: 10.1038/s41598-017-02079-3
37. Perera M, Al-hebshi NN, Perera I, Ipe D, Ulett GC, et al. Inflammatory bacteriome and oral squamous cell carcinoma. *J Dent Res.* (2018) 97:725–32. doi: 10.1177/0022034518767118
38. Zhang L, Liu Y, Zheng HJ, Zhang CP. The oral microbiota may have influence on oral cancer. *Front Cell Infect Microbiol.* (2019) 9:476. doi: 10.3389/fcimb.2019.00476
39. Mager DM, Ximenez-Fyvie LA, Haffajee AD, Socransky SS, et al. The salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. *J Transl Med.* (2005) 3:27. doi: 10.1186/1479-5876-3-27
40. Pushalkar S, Mane SP, Ji X, Li Y, Evans C, et al. Microbial diversity in saliva of oral squamous cell carcinoma. *FEMS Immunol Med Microbiol.* (2011) 61:269–77. doi: 10.1111/j.1574-695X.2010.00773.x
41. Ellington TD, Henley SJ, Senkomago V, O'Neil ME, Wilson RJ, et al. Trends in incidence of cancers of the oral cavity and pharynx—United States 2007–2016. *MMWR Morb Mortal Wkly Rep.* (2020) 69:433–8. doi: 10.15585/mmwr.mm6915a1
42. Montero PH, Patel PD, Palmer FL, Patel SG, Shah JP, et al. Changing trends in smoking and alcohol consumption in patients with oral cancer treated at memorial Sloan-Kettering Cancer Center from 1985 to 2009. *Arch Otolaryngol Head Neck Surg.* (2012) 138:817–22. doi: 10.1001/archoto.2012.1792
43. Tota JE, Anderson WF, Coffey C, Califano J, Cozen W, et al. Rising incidence of oral tongue cancer among white men and women in the United States, 1973–2012. *Oral Oncol.* (2017) 67:146–52. doi: 10.1016/j.oraloncology.2017.02.019
44. Pinto AC, Carames J, Francisco H, Chen A, Azul AM, Marques D. Malignant transformation rate of oral leukoplakia—systematic review. *Oral Surg Oral Med Oral Pathol Oral Radiol.* (2020) 129:600–11. doi: 10.1016/j.oooo.2020.02.017
45. Warnakulasuriya S, Ariyawardana A. Malignant transformation of oral leukoplakia: a systematic review of observational studies. *J Oral Pathol Med.* (2016) 45:155–66. doi: 10.1111/jop.12339
46. Wu J, Peters BA, Dominianni C, Zhang Y, Pei Z. Cigarette smoking and the oral microbiome in a large study of American adults. *ISME J.* (2016) 10:ismej201637.
47. Wirth R, Maróti G, Mihók R, Simon-Fiala D, Antal M, Pap B, et al. A case study of salivary microbiome in smokers and non-smokers in Hungary: analysis by shotgun metagenome sequencing. *J Oral Microbiol.* (2020) 12:1773067. doi: 10.1080/20002297.2020.1773067
48. Gopinath D, Wie CC, Banerjee M, Thangavelu L, Pradeep K, Nallaswamy D, et al. Compositional profile of mucosal bacteriome of smokers and smokeless tobacco users. *Clin Oral Invest.* (2021) 26:1–10. doi: 10.1007/s00784-021-04137-7
49. Zhang S, Li C, Liu J, Geng F, Shi X, Li Q, et al. *Fusobacterium nucleatum* promotes epithelial-mesenchymal transition through regulation of the lncRNA MIR4435-2HG/miR-296-5p/Akt2/SNAI1 signaling pathway. *FEBS J.* (2020) 287:4032–47. doi: 10.1111/febs.15233
50. Hernandez BY, Zhu X, Goodman MT, Gatewood R, Mendiola P. Betel nut chewing, oral premalignant lesions, and the oral microbiome. *PLoS ONE.* (2017) 12:e0172196. doi: 10.1371/journal.pone.0172196
51. Yang I, Rodriguez J, Wright CY, Hu Y. The oral microbiome of electronic cigarette users: a cross-sectional exploration. *Oral Dis.* (2022). doi: 10.1111/odi.14186

52. Chopyk J, Bojanowski CM, Shin J, Moshensky A, Fuentes AL, Bonde SS, et al. Compositional differences in the oral microbiome of e-cigarette users. *Front Microbiol.* (2021) 12:599664. doi: 10.3389/fmicb.2021.599664
53. Ganesan SM, Dabdoub SM, Nagaraja HN, Scott ML, Pamulapati S, Kumar PS, et al. Adverse effects of electronic cigarettes on the disease-naïve oral microbiome. *Sci Adv.* (2020) 6:eaz0108. doi: 10.1126/sciadv.aaz0108
54. Pushalkar S, Paul B, Li Q, Yang J, Vasconcelos R, Makwana S, et al. Electronic cigarette aerosol modulates the oral microbiome and increases risk of infection. *Isience.* (2020) 23:100884. doi: 10.1016/j.isci.2020.100884
55. Gallimidi AB, Fischman S, Revach B, Bulvik R, Maliutina A, Rubinstein AM, et al. Periodontal pathogens *Porphyromonas gingivalis* and *Fusobacterium nucleatum* promote tumor progression in an oral-specific chemical carcinogenesis model. *Oncotarget.* (2015) 6:22613–23. doi: 10.18632/oncotarget.4209
56. Abdalla Z, Walsh T, Thakker N, Ward CM. Loss of epithelial markers is an early event in oral dysplasia and is observed within the safety margin of dysplastic and T1 OSCC biopsies. *PLoS ONE.* (2017) 12:e0187449. doi: 10.1371/journal.pone.0187449
57. Yu H, Zeng J, Liang X, Wang W, Zhou Y, Sun Y, et al. *Helicobacter pylori* promotes epithelial-mesenchymal transition in gastric cancer by downregulating programmed cell death protein 4 (PDCD4). *PLoS ONE.* (2014) 9:e105306. doi: 10.1371/journal.pone.0105306
58. Lee J, Roberts JS, Atanasova KR, Chowdhury N, Han K, Yilmaz Ö. Human primary epithelial cells acquire an epithelial-mesenchymal-transition phenotype during long-term infection by the oral opportunistic pathogen, *Porphyromonas gingivalis*. *Front Cell Infect Microbiol.* (2017) 7:493. doi: 10.3389/fcimb.2017.00493
59. Uitto V-J, Baillie D, Wu Q, Gendron R, Grenier D, Putnins EE, et al. *Fusobacterium nucleatum* increases collagenase 3 production and migration of epithelial cells. *Infect Immun.* (2005) 73:1171–9. doi: 10.1128/IAI.73.2.1171-1179.2005
60. Inaba H, Amano A, Lamont RJ, Murakami Y. Involvement of protease-activated receptor 4 in over-expression of matrix metalloproteinase 9 induced by *Porphyromonas gingivalis*. *Med Microbiol Immun.* (2015) 204:605–12. doi: 10.1007/s00430-015-0389-y
61. Nosova T, Jokelainen K, Kaihovaara P, Jousimies-Somer H, Siitonen A, Heine R, et al. Aldehyde dehydrogenase activity and acetate production by aerobic bacteria representing the normal flora of human large intestine. *Alcohol Alcohol.* (1996) 31:555–64. doi: 10.1093/oxfordjournals.alcalca.a008191
62. Baan R, Straif K, Grosse Y, Secretan B, Ghissassi FE, Bouvard V, et al. Carcinogenicity of alcoholic beverages. *Lancet Oncol.* (2007) 8:292–3. doi: 10.1016/S1470-2045(07)70099-2
63. Acetaldehyde. IARC monographs on the evaluations of carcinogenic risks to humans. *IARC Monogr Eval Carcinog Risks Hum.* (1999) 71:319–35.
64. Homann N, Tillonen J, Rintamaki H, Salaspuro M, Lindqvist C, Meurman JH. Poor dental status increases acetaldehyde production from ethanol in saliva: a possible link to increased oral cancer risk among heavy drinkers. *Oral Oncol.* (2001) 37:153–8. doi: 10.1016/S1368-8375(00)00076-2
65. Brooks PJ, Theruvathu JA. DNA adducts from acetaldehyde: implications for alcohol-related carcinogenesis. *Alcohol.* (2005) 35:187–93. doi: 10.1016/j.alcohol.2005.03.009
66. Seitz HK, Stickel F. Molecular mechanisms of alcohol-mediated carcinogenesis. *Nat Rev Cancer.* (2007) 7:599–612. doi: 10.1038/nrc2191
67. Meurman JH. Infectious and dietary risk factors of oral cancer. *Oral Oncol.* (2010) 46:411–3. doi: 10.1016/j.oraloncology.2010.03.003
68. Homann N, Jousimies-Somer H, Jokelainen K, Heine R, Salaspuro M. High acetaldehyde levels in saliva after ethanol consumption: methodological aspects and pathogenetic implications. *Carcinogenesis.* (1997) 18:1739–43. doi: 10.1093/carcin/18.9.1739
69. Homann N, Tillonen J, Meurman JH, Rintamaki H, Lindqvist C, Rautio M, et al. Increased salivary acetaldehyde levels in heavy drinkers and smokers: a microbiological approach to oral cavity cancer. *Carcinogenesis.* (2000) 21:663–8. doi: 10.1093/carcin/21.4.663
70. Marttila E, Uittamo J, Rusanen P, Lindqvist C, Salaspuro M, Rautemaa R. Acetaldehyde production and microbial colonization in oral squamous cell carcinoma and oral lichenoid disease. *Oral Surg Oral Med Oral Pathol Oral Radiol.* (2013) 116:61–8. doi: 10.1016/j.oooo.2013.02.009
71. Alnuaimi A, Ramadan AN, Wiesenfeld D, O'Brien-Simpson NM, Kolev SD, Reynolds EC, et al. *Candida* virulence and ethanol-derived acetaldehyde production in oral cancer and non-cancer subjects. *Oral Dis.* (2016) 22:805–14. doi: 10.1111/odi.12565
72. Muto M, Hitomi Y, Ohtsu A, Shimada H, Kashiwase Y, Sasaki H, et al. Acetaldehyde production by non-pathogenic *Neisseria* in human oral microflora: implications for carcinogenesis in upper aerodigestive tract. *Int J Cancer.* (2000) 88:342–50. doi: 10.1002/1097-0215(20001101)88:3<342::AID-IJC4>3.0.CO;2-I
73. Moritani K, Takeshita T, Shibata Y, Ninomiya T, Kiyohara Y, Yamashita Y. Acetaldehyde production by major oral microbes. *Oral Dis.* (2015) 21:748–54. doi: 10.1111/odi.12341
74. Amer A, Whelan A, Al-Hebshi NN, Healy CM, Moran GP. Acetaldehyde production by *Rothia mucilaginosa* isolates from patients with oral leukoplakia. *J Oral Microbiol.* (2020) 12:1743066. doi: 10.1080/20002297.2020.1743066
75. Kurkivuori J, Salaspuro V, Kaihovaara P, Kari K, Rautemaa R, Grönroos L, et al. Acetaldehyde production from ethanol by oral streptococci. *Oral Oncol.* (2007) 43:181–6. doi: 10.1016/j.oraloncology.2006.02.005
76. Francescone R, Hou V, Grivennikov SI. Microbiome, inflammation, and cancer. *Cancer J.* (2014) 20:181–9. doi: 10.1097/PPO.0000000000000048
77. O'Grady I, Anderson A, O'Sullivan J. The interplay of the oral microbiome and alcohol consumption in oral squamous cell carcinomas. *Oral Oncol.* (2020) 110:105011. doi: 10.1016/j.oraloncology.2020.105011
78. Landskron G, De la Fuente M, Thuwajit P, Thuwajit C, Hermoso MA. Chronic inflammation and cytokines in the tumor microenvironment. *J Immunol Res.* (2014) 2014:149185. doi: 10.1155/2014/149185
79. Han Yw, Shi W, Huang GT-J, Haake SJ, Park N-H, Kuramitsu H, et al. Interactions between periodontal bacteria and human oral epithelial cells: *Fusobacterium nucleatum* adheres to and invades epithelial cells. *Infect Immun.* (2000) 68:3140–6. doi: 10.1128/IAI.68.6.3140-3146.2000
80. Xu M, Yamada M, Li M, Liu H, Chen SG, Han YW. Fada from *Fusobacterium nucleatum* utilizes both secreted and nonsecreted forms for functional oligomerization for attachment and invasion of host cells. *J Biol Chem.* (2007) 282:25000–9. doi: 10.1074/jbc.M611567200

## 19 Appendix H Published paper based on the findings detailed in Chapter 6



## Smoking, tooth loss and oral hygiene practices have significant and site-specific impacts on the microbiome of oral mucosal surfaces: a cross-sectional study

Sheila Galvin, Sviatlana Anishchuk, Claire M. Healy & Gary P. Moran

**To cite this article:** Sheila Galvin, Sviatlana Anishchuk, Claire M. Healy & Gary P. Moran (2023) Smoking, tooth loss and oral hygiene practices have significant and site-specific impacts on the microbiome of oral mucosal surfaces: a cross-sectional study, *Journal of Oral Microbiology*, 15:1, 2263971, DOI: [10.1080/20002297.2023.2263971](https://doi.org/10.1080/20002297.2023.2263971)

**To link to this article:** <https://doi.org/10.1080/20002297.2023.2263971>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 02 Oct 2023.



[Submit your article to this journal](#)



Article views: 141



[View related articles](#)



[View Crossmark data](#)



RESEARCH ARTICLE



# Smoking, tooth loss and oral hygiene practices have significant and site-specific impacts on the microbiome of oral mucosal surfaces: a cross-sectional study

Sheila Galvin<sup>a</sup>, Sviatlana Anishchuk<sup>a</sup>, Claire M. Healy<sup>a</sup> and Gary P. Moran <sup>b</sup>

<sup>a</sup>Division of Oral and Maxillofacial Surgery, Oral Medicine and Oral Pathology, School of Dental Science, Trinity College Dublin, Dublin Dental University Hospital, Dublin, Ireland; <sup>b</sup>Division of Oral Biosciences, School of Dental Science, Trinity College Dublin, Dublin Dental University Hospital, Dublin, Ireland

## ABSTRACT

We investigated bacterial colonisation patterns of healthy mucosa (buccal, tongue, palate and floor of mouth) in a cohort of adults in order to determine how smoking, tooth loss, plaque levels and oral hygiene practices impacted on mucosal colonisation. A total of 322 swabs were recovered from 256 participants, of whom 46% were current smokers. We analysed colonization by sequencing the V1-V3 regions of the 16S rRNA gene. Palate and tongue microbiomes generally exhibited greater biodiversity than buccal and floor of mouth. Although *Neisseria*, *Lautropia* and *Haemophilus* spp. showed reduced abundance in smokers, buccal mucosa specifically showed a significant increase in *Prevotella* spp., whereas tongue and floor of mouth tended towards increased abundance of *Streptococcus* spp. Unexpectedly, tooth brushing frequency had a greater impact on mucosal community structure than plaque levels. Tooth loss was associated with significant reductions in mucosal biodiversity and had site-specific impacts, with buccal communities showing increased abundance of periodontitis-associated species and *Rothia mucilaginosa*, whereas tongue communities exhibited increased abundance of several streptococcal OTUs and reduced abundance of *Haemophilus* spp. This study highlights the complex relationship between mucosal colonisation and host factors, highlighting the need for careful consideration of these factors in mucosal microbiome studies.

## ARTICLE HISTORY

Received 18 April 2023  
Revised 9 September 2023  
Accepted 13 September 2023

## KEYWORDS

Oral mucosa; microbiome; 16S rRNA; smoking; tooth loss; oral hygiene

## Introduction

The oral microbiome has undergone intense investigation to understand its core composition and role in common human diseases, such as dental decay (caries) and gum disease (periodontitis) [1–4]. Early culture-independent, molecular techniques using 16S rRNA cloning demonstrated that this microbiome was highly diverse and site specific, with a high proportion of uncultured species [5–7]. With the advent of next-generation sequencing, the Human Microbiome Consortium sampled the oral cavity of North American adults and characterised a rich bacterial community that differed between intraoral niches, with a conserved core microbiome that was shared between individuals [8–10]. Amplification and sequencing of the variable regions of the 16S rRNA gene is still the method of choice in studies, and the development of the aerodigestive habitat-specific expanded Human Oral Microbiome Database (eHOMD) has allowed species level or supra-species level classification of 16S rRNA reads [11,12]. Further, fine-scale sub-species typing can be achieved by oligotyping, which was used by Eren et al. to show that different oligotypes of *Streptococcus*, *Neisseria* and *Veillonella* (which could not

be resolved using traditional analysis) exhibited very different distributions among oral sites [13].

Although our understanding of oral ecology has greatly improved, the factors that result in interindividual variability are still poorly understood. We know that events early in life such as infant feeding methods and antibiotic usage can have long-term impacts on oral microbiome structure and that events such as tooth eruption play an important role in maturation of the oral microbiome [2,14–16]. Variation in the adult microbiome may be influenced by smoking, oral hygiene, diet (notably sugar consumption) and periodontal health [3,16–21]. Many of these studies analyse the salivary microbiome as a representation of the oral microbiome [22], or in the case of studies examining oral hygiene and periodontal health, focus on plaque composition [17,18,23]. The influence of host and environmental factors on the microbiome of mucosal surfaces is less well understood and given the site specificity of the oral microbiome, the site-specific impacts of these factors are also poorly understood. As an example, the largest study to date of smoking and the oral

**CONTACT** Gary P. Moran  [gpmoran@dental.tcd.ie](mailto:gpmoran@dental.tcd.ie)  Lincoln Place, Dublin Dental University Hospital, Dublin D02 F859, Ireland

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/20002297.2023.2263971>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

microbiome focuses solely on saliva [17]. The impact of smoking on mucosal surfaces has been reported, but the studies are relatively small and often have low proportions of current smokers [24,25].

The current study was designed to examine the microbiome of four distinct oral niches, namely buccal, tongue, palate and floor of mouth (FOM). Healthy mucosal sites were sampled from patients attending for routine dental care and from patients attending an oral dysplasia clinic, with the latter group containing a high proportion of smokers and alcohol users with relatively high levels of tooth-loss. Multivariate analyses were used to investigate the impacts of variables including smoking, tooth loss and oral hygiene practices on the oral microbiome and to determine if any site-specific impacts could be identified.

## Materials and methods

### Sample collection

Ethical Approval for this study was granted by the Joint Hospitals' Research Ethics Committee, Tallaght Hospital, Dublin (Ref: 2017-11-Chairman's actions [7]). Following written informed patient consent, mucosal swabs were collected at a single site, the Dublin Dental University Hospital. Recruitment was carried out among patients attending general dental clinics for routine check-ups ( $n=61$ ) and from patients attending the oral medicine clinic ( $n=195$ ). Only healthy mucosal tissue was sampled. Swabs were taken using sterile polyurethane foam swabs (CultureSwab EZ, BD) by a dental hygienist by gently rubbing the surface of the mucosa 10 times with rotation of the swab. Samples were recovered at least 1 h after eating, oral hygiene or mouthwash use. Participants were recorded as current smokers, former smokers or never smokers and alcohol consumption was recorded as units of alcohol consumed per week. Oral health of the participants was recorded including number of teeth, plaque levels (determined using the simplified oral hygiene index, OHI-S [26]), tooth brushing frequency, mouthwash use and denture wearing.

Patients having taken antibiotics or used topical steroids intra-orally in the past 3 months were excluded along with patients with diabetes mellitus, Crohn's disease, ulcerative colitis, current viral infection (cold/flu), or history of gastrointestinal malignancy. A total of 322 oral mucosal swabs were collected as detailed in Table 1. For this analysis, a minimum sample size of 235 was determined based on previous data [27]. Using the chi-square test to determine if there is an association between mucosal sites (four categories) and a variable with three levels (e.g. never, current and former smoker)

**Table 1.** Mucosal sites sampled (percentages in brackets).

| Anatomical areas     | Sub-sites                      | Number (%)  |
|----------------------|--------------------------------|-------------|
| Floor of mouth (Fom) | Floor of mouth, ventral tongue | 43 (13.3%)  |
| Buccal (Buc)         | Buccal mucosa, Gingiva         | 128 (39.8%) |
| Tongue (Ton)         | Dorsum, lateral border tongue  | 115 (35.7%) |
| Palate (Pal)         | Palate                         | 36 (11.2%)  |
| Total                |                                | 322 (100%)  |

an effect size of 0.25 can be identified assuming an estimated power of 0.8, level of significance 0.05.

### DNA extraction and sequencing

Swabs were immediately stored at  $-80^{\circ}\text{C}$  for a maximum of 6 weeks prior to DNA extraction. DNA was extracted from mucosal swabs using the MasterPure Complete DNA/RNA Purification Kit (Epicentre Biotechnologies). The manufacturer's extraction protocol was supplemented with an additional incubation step with Ready-Lyse lysozyme (Epicentre Biotechnologies) and a bead disruption step, as described by Amer et al. [27]. The DNA was resuspended in a Tris-EDTA buffer (pH 7.5) and stored at  $-80^{\circ}\text{C}$ . Amplification of the V1-V3 region of the 16S rRNA gene was carried out using the primers 27F-YM and 519 R (27F-YM: 5'-AGAGTTTGATYMTGGCTCAG; 519 R: 5'-GWATTACCGCGGCKGCTG) [28] and was performed in duplicate using separate template dilutions (1:1 and 1:10). Amplification and sequencing was performed by Integrated Microbiome Resources (IMR; Dalhousie, Canada) according to their standardised, in-house protocols described by Comeau et al. [29] and detailed here: <https://www.protocols.io/workspaces/integrated-microbiome-resource-imr/publications>. Paired end sequencing was performed using the Illumina 600 cycle MiSeq reagent kit. All sequence data has been submitted to the NCBI sequence read archive (SRA), submission SUB13676038, BioProject PRJNA994083.

### Data analysis

Bacterial 16S rRNA sequences were processed and filtered using Dada2 with the following parameters: maxN = 0, maxEE = c(2,5) and truncQ = 2 [30]. The 10 terminal bases were trimmed from reverse reads due to the drop in sequence quality. Following error estimation and correction using the Dada2 algorithm, paired reads were merged and chimeras were removed following their identification with the removeBimeraDenovo command. Taxonomy was assigned using the Human Oral Microbiome Database (HOMD) classifications (eHOMD 16S rRNA RefSeq version 15.1) [7]. The packages Phyloseq [31] and MicrobiotaProcess [32] were used to calculate the alpha diversity indices for each



sample and for comparison of beta diversity (community structure) [31]. Further visualisations (biplots) and analysis of community structure, including PERMANOVA, were conducted using the package MicroViz in R studio [33,34]. Multivariate analysis was conducted with MaAsLin 2 using relative abundance data and the default linear models (LM) setting with log transformation of the data [34]. MaAsLin 2 was used to control for the influence of patient variables such as smoking, oral hygiene and tooth loss and to exclude any possible associations between them. We also included the clinic attended (either oral medicine clinic or general dental clinic) as a random effect to test if this had a significant impact on the results. The indicated patient metadata categories were used as fixed-effects with 'healthy' variables set as the reference (e.g. never smokers, 0 alcohol units/week, good oral hygiene, <5 missing teeth, brushing >1 per day) [34]. MaAsLin2 generates tabular outputs (see supplementary tables) detailing a coefficient of variance from the reference and adjusted *p* values (q-values; Benjamini-Hochberg FDR). Heatmaps show the 50 most significant values (default *Padj* < 0.25) with the effect size calculated as  $(-\log(q\text{-value}) * (\text{coefficient}))$ .

## Results

Swabs from healthy oral mucosal surfaces were analysed to investigate the influence of mucosal site, age, gender, smoking, alcohol consumption, plaque levels, tooth loss and denture wear on the microbiome. A total of 322 sites were sampled from 256 patients. Of these patients, 195 were attending the oral medicine clinic and 61 were attending general dental clinics at the Dublin Dental Hospital. Oral sites swabbed included four main anatomical areas: floor of mouth (FOM), buccal, tongue and palate (Table 1). Buccal and tongue samples were recovered from almost all participants, where healthy mucosa was present. A proportion of patients were also sampled at the palate (*n* = 36) and floor of mouth (FOM, *n* = 43). There was no statistical difference in the proportions of sites swabbed in the oral medicine and general dental patients (Table S1). However, those attending the oral medicine clinic were more likely to smoke, drink alcohol, had poorer oral hygiene and had fewer teeth compared to those attending the general dental clinics (Table S1).

Analysis of alpha diversity showed that tongue and palate communities exhibited similar Chao1 richness values and Shannon diversity index values (Figure 1a). Communities from the FOM and buccal regions had significantly reduced Chao1 values compared to tongue communities (Figure 1a). Biodiversity was lowest in FOM communities, which was significantly reduced compared to both palate

(Shannon and Simpson) and tongue (Simpson) (Figure 1a and Fig. S1a).

Although different mucosal sites had similar core microbiomes, mucosal site was found to have a significant influence on microbial community structure. Buccal and FOM sites had the greatest proportions of *Streptococcus* species, tongue sites showed increased abundance of *Rothia* and *Neisseria* species, while palatal sites showed increased abundances of *Prevotella* species (Figure 1b). PCA (principal component analysis) of community membership by site also highlighted these genera as being responsible for differences in community structure (Fig. S1b). Inter-site comparisons of Bray-Curtis dissimilarity values using pairwise PERMANOVA confirmed this, showing significant differences in community structure between all sites with the exception of buccal and floor of mouth communities (Table 2).

MaAsLin2 analysis was used to carry out a multivariate analysis to investigate which species showed significant differences in abundance at each site. Using buccal mucosa as reference, 59 species showed significant changes (*Padj* < 0.05) in abundance, with 28 on the tongue, 30 on the palate, but only one on FOM (Figure 1c and Table S2). The most significant changes at tongue sites were increased abundances of *Rothia* spp. including *R. mucilaginosa* and *R. aeria:dentocariosa*, *Veillonella* species and *Actinomyces* species. Abundances of *Granulicatella adiacens* and *S. parasanguinis* were also significantly increased (Figure 1c and Table S1). Eleven species showed reduced abundance at tongue sites including OTUs related to *S. mitis* (Figure 1c and Table S1).

Significant changes at palate sites included increased abundance of six *Prevotella* species, four *Actinomyces* species and three *Streptococcus* species. Relative to both tongue and palate, buccal mucosa had significantly higher levels of *S. mitis*, *R. dentocariosa*, *S. sanguinis* and *P. micra* (Figure 1c and Table S2). Multivariate analysis showed that these major site-specific changes were independent of smoking, oral hygiene and tooth loss and whether the patient attended oral medicine or general dental clinics (Table S3).

## Smoking and alcohol consumption and the mucosal microbiome

Smoking was found to have a significant influence on the microbial community of mucosal surfaces. Although no significant changes in alpha diversity metrics were observed (Fig. S2), visualisation of community structure using NMDS plots of Bray-Curtis dissimilarity values showed separation of communities from current and never smokers, particularly along axis 3 (Figure 2a). Pairwise PERMANOVA

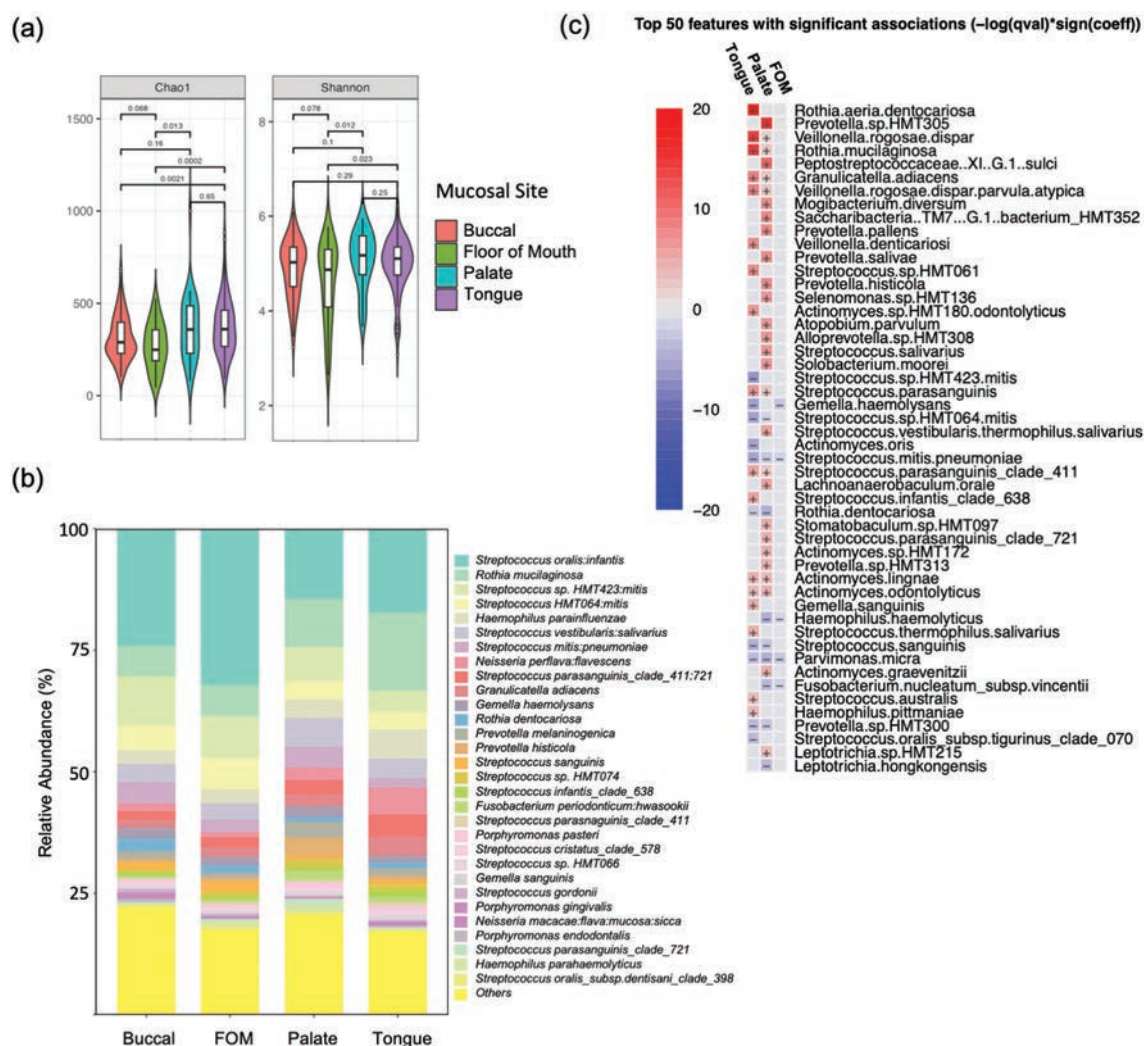


**Figure 1.** Analysis of healthy mucosal microbiomes by site. (a) Alpha diversity values for buccal, floor of mouth (FOM), palate and tongue communities. (b) Bar plot showing relative abundance (%) of the 30 most abundant species. (c) Results of multivariate analysis carried out using MaAsLin2 comparing FOM, tongue and palate communities with buccal mucosa.

**Table 2.** Pairwise PERMANOVA results for mucosal site comparisons.

| Site comparisons | p value | Adjusted p value |
|------------------|---------|------------------|
| Tongue vs Palate | 0.004   | 0.004            |
| Tongue vs Buccal | 0.0001  | 0.0002           |
| Tongue vs FOM    | 0.0001  | 0.0002           |
| Palate vs Buccal | 0.0001  | 0.0002           |
| Palate vs FOM    | 0.0001  | 0.0002           |
| Buccal vs FOM    | 0.13    | 0.13             |

comparisons showed that there were significant differences between the microbial communities of all three categories of smokers ( $\text{Padj} < 0.003$ ; Table S4). MaAslin2 was used for multivariate analysis to identify which species were responsible for these differences. Twenty-eight species showed significantly decreased, and 29 species significantly increased abundances (all  $\text{Padj} < 0.05$ ) associated with smoking (Figure 2b, Table S5). Smoking was associated with loss of species typically associated with healthy mucosa, including several species of *Neisseria* and *Streptococcus*, *L. mirabilis*, *R. aeria* and *H. parainfluenza*. Several Gram-negative species

from the genera *Aggregatibacter*, *Kingella*, *Bergeyella*, *Fusobacterium* and *Capnocytophaga* also showed significant decreases in abundance in association with smoking (Figure 2b, Table S5). The species showing increased abundance associated with smoking were predominantly periodontitis-associated species from the genera *Prevotella*, *Porphyromonas*, *Fusobacterium*, *Tannerella*, *Parvimonas*, *Filifactor* and *Peptostreptococcaceae* (all  $\text{Padj} < 0.05$ ; Figure 2b, Table S5).

Examination of the impact of smoking at different mucosal sites revealed site-specific effects (Fig. S2b). *Neisseria*, *Lautropia* and *Kingella* spp. were all found to be reduced on tongue, buccal and floor of mouth in smokers. No significant changes were identified on the palate. Uniquely, buccal communities exhibited significant increased abundance of several genera associated with periodontal disease, including *Tannerella*, *Filifactor* and *Porphyromonas* (Fig. S2b). Analysis of buccal community composition showed that smoking was associated with a shift from *Neisseria* and *Haemophilus* towards communities

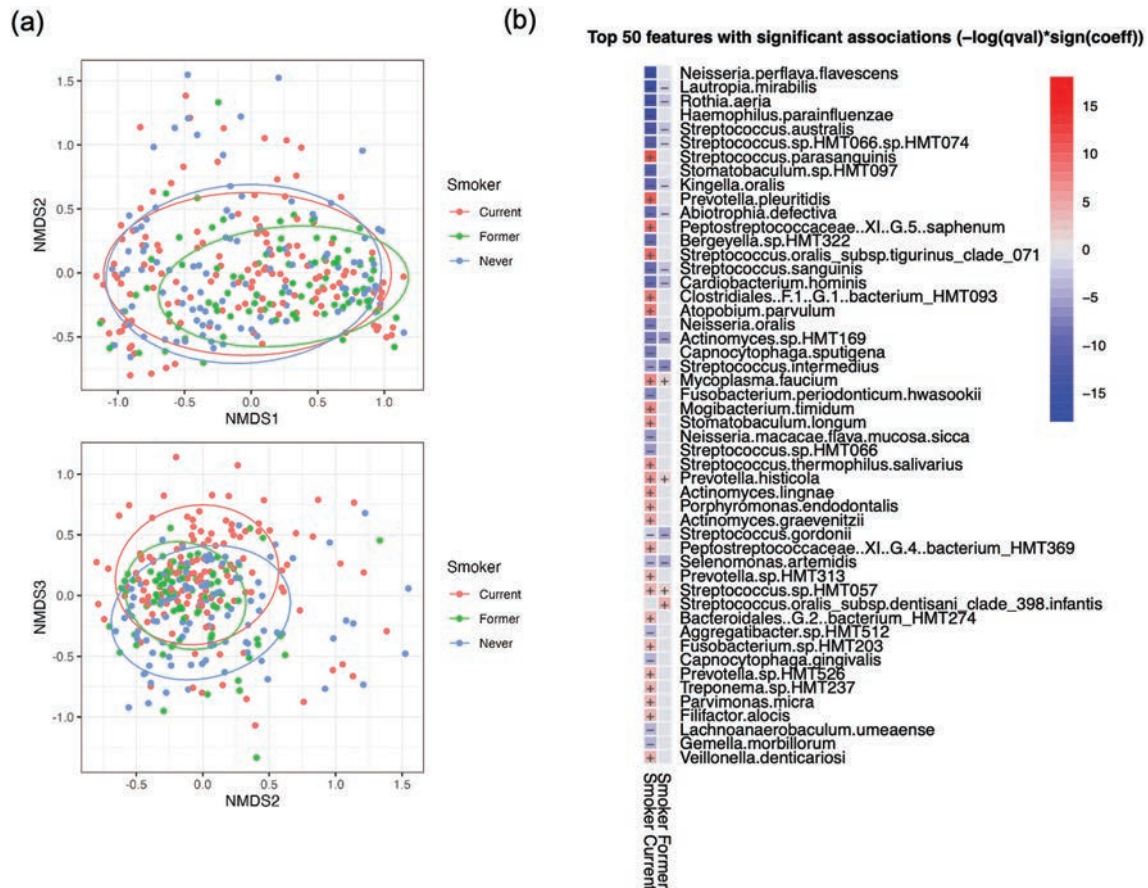


**Figure 2.** Analysis of the impact of smoking on the mucosal microbiome. (a) NMDS plots comparing community structure of samples based on Bray-Curtis dissimilarity index values. (b) Results of multivariate analysis carried out using MaAsLin2 with comparing the mucosal microbiomes of current and former smokers with never smokers.

rich in *Prevotella* (MaAsLin 2  $P_{adj} = 0.013$ ; Figure 3a, b). PCA plots examining tongue communities also showed that smoking was associated with a shift from *Neisseria* rich communities towards communities with increased abundance of *Rothia* or *Streptococcus* (Figure 3a,b), although the increases in *Rothia* and *Streptococcus* were not highly significant (Figure 2d).

Alcohol consumption was not associated with any significant shifts in alpha or beta-diversity ( $p = 0.38$ ) (Fig. S3). A multivariate analysis of smoking and alcohol consumption, while confirming the previous associations with smoking, identified that consumption of >20 units of alcohol per week resulted in increased levels of *C. concisus*, *P. melaninogenica* and reduced levels of *F. alocis* (Fig. S3).

### Oral hygiene and the mucosal microbiome

We categorised our participants based on oral hygiene practices including tooth-brushing frequency (brushing <1, 1 or >1 per day), mouthwash use and on the basis of observed levels of plaque accumulation using the simplified oral hygiene index (OHI-S). The number of individuals who brushed frequently (>1 per day,  $n = 186$ ) outnumbered those that brushed infrequently

(≤1 Per day,  $n = 67$ ). Brushing less than once per day was associated with reduced biodiversity as indicated by significantly reduced Shannon index values (Figure 4a). Analysis of community structure using NMDS plots comparing Bray-Curtis dissimilarity values demonstrated separation of samples where brushing occurred more than once per day (PERMANOVA  $p = 0.032$ ; Figure 4b). Bar plots of relative species abundance indicated that infrequent brushing (≤1 Per day) was associated with higher levels of streptococci (Figure 4c). A multivariate analysis of brushing frequency with smoking showed that those who brushed infrequently (≤1 Per day) had significantly increased levels of the highly abundant OTU *S. oralis:infantis*, as well as *H. parahaemolyticus* and *G. elegans* (Table S6 and Fig. S4).

Measurement of plaque levels in dentate participants was assessed using the OHI-S index, which categorised the participants as having good ( $n = 77$ ), fair ( $n = 77$ ) and poor ( $n = 47$ ) oral hygiene. Oral hygiene category had less of an impact on mucosal alpha diversity and population structure compared to brushing frequency (Figure 4a,b). It appeared that the shifts in population structure related to oral hygiene were not highly significant (Figure 4b; PERMANOVA  $P_{adj} > 0.05$ ). However, in a multivariate analysis (Table S6, Fig. S4), poor oral hygiene was



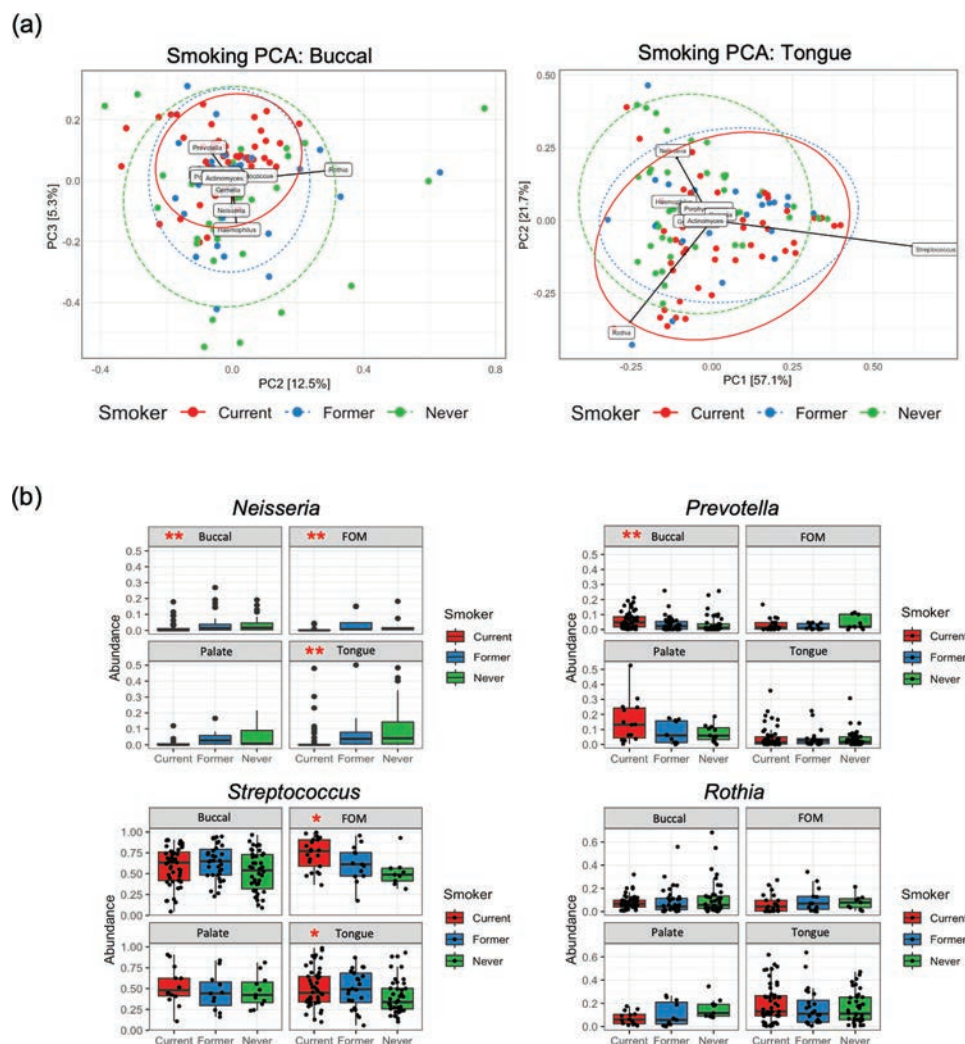


**Figure 3.** Site-specific effects of smoking on the mucosal microbiome. (a) PCA biplots comparing the genus level composition of buccal and tongue communities in smokers, former smokers and never smokers. Labelled arrows indicate the influence of specific genera in separating samples along the respective axis. (b) Box-plots showing relative abundance of *Neisseria*, *Prevotella*, *Streptococcus* and *Rothia* species in current, former and never smokers at each mucosal site. Significance values (MaAslin 2) are indicated by \*\* (Padj < 0.05) and \* (Padj < 0.2).

associated with a significantly increased abundance of several plaque and periodontitis-associated species including *P. denticola* (Padj 0.09), *Leptotrichia* sp. HMT498 (Padj = 0.09), *Saccharibacteria* sp. HMT346 (Padj = 0.087), *Lachnospiraceae* sp. HMT100 (Padj = 0.023), *Tannerella* sp. HMT286 (Padj = 0.058) and *F. nucleatum* sp. HMT204 (Padj = 0.077) (Table S6, Fig. S4a).

Mouthwash users had significantly different population structures to those who did not use mouthwash (Fig. S5, PERMANOVA  $p = 0.03$ ). MaAsLin2 analysis showed that this difference was associated with reduced levels of the high abundance streptococcal OTU *S. mitis:pneumoniae* (Padj = 0.01) in mouthwash users. *Gr. adiacens* and *R. mucilaginosa* appeared to show increased abundance with mouthwash use, although the adjusted  $p$  values were not as significant (Fig. S4, Table S6).

### Tooth loss and denture wear and the mucosal microbiome

The alpha diversities of communities in individuals missing more than 15 teeth were significantly different on all six alpha diversity indices analysed when compared to those with less than five missing teeth (Figure 5a and Fig. S6a). Tooth number had a significant influence on the mucosal microbial population (Figure 5b), with NMDS plots showing clustering of communities from those who had lost more than 15 teeth (PERMANOVA Padj < 0.05; Figure 5c). Missing 5 to 15 and >15 teeth was associated with altered abundance (Padj < 0.1) of 14 and 17 species, respectively (Table S7; Fig. S6b). Missing 5–15 teeth was associated with a reduced abundance of *Haemophilus* spp. and increased abundance of *S. parasanguinis* in addition to many taxa associated with periodontal diseases such as *E. corrodens*, *P. micra*, *Saccharibacteria* sp. HMT349, *Fretibacterium* sp. HMT359 and *Treponema*

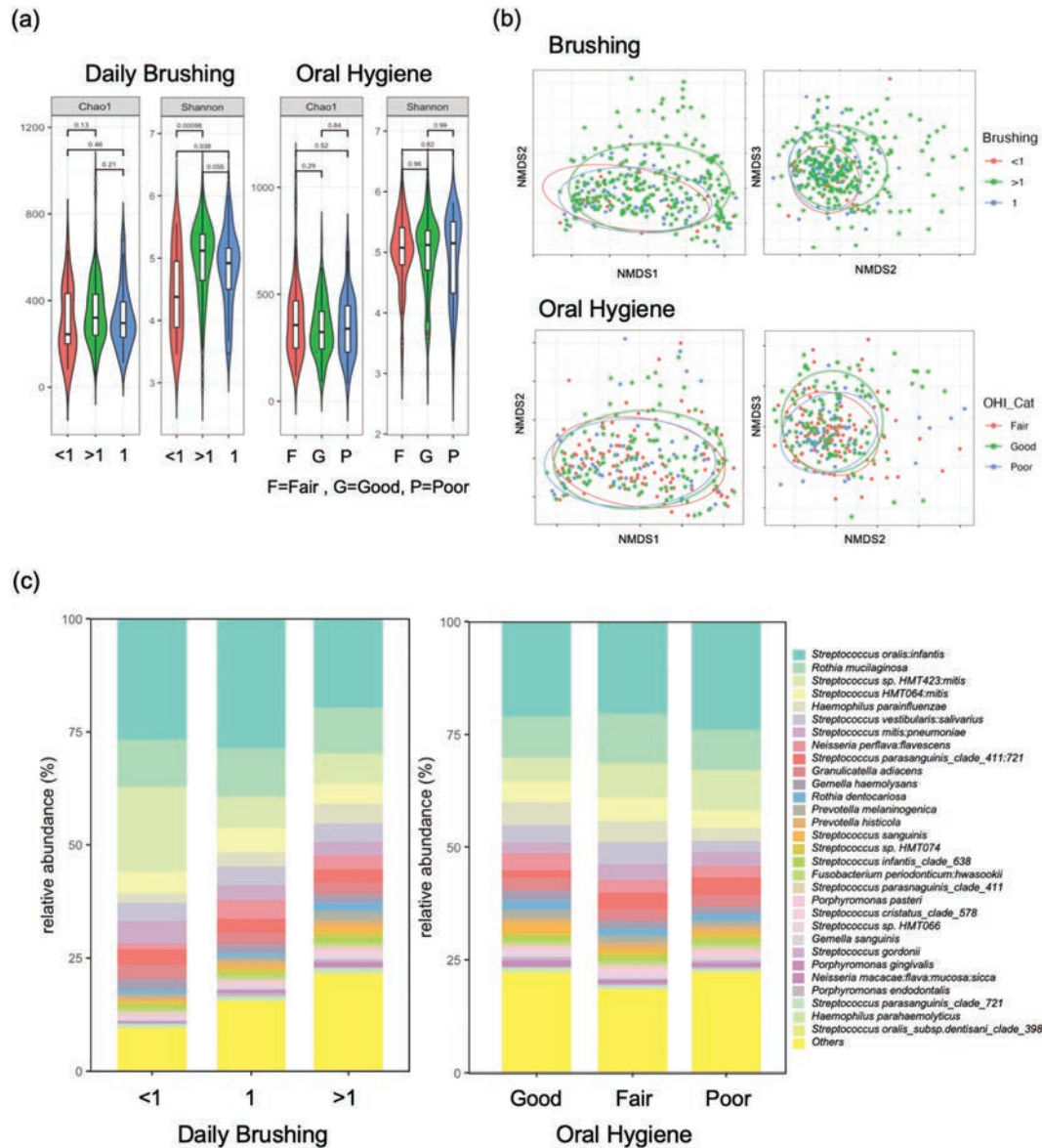


**Figure 4.** Impact of tooth brushing frequency (<1, 1 or >1 per day) and oral hygiene (OHI-5 index), on the mucosal microbiome. (a) Alpha diversity values according to daily brushing frequency and oral hygiene. (b) Community structure analysed by NMDS of Bray-Curtis dissimilarity values according to daily brushing frequency and oral hygiene. (c) Bar plot showing relative abundance (%) of the 30 most abundant species according to daily brushing frequency and oral hygiene.

sp. HMT237 (all  $P_{adj} < 0.1$ ; Table S7). Missing more than 15 teeth was associated with the loss of many taxa reflecting the reduction in richness and biodiversity including *H. parainfluenzae*, *L. mirabilis*, *S. sanguinis* and *R. aeria* (all  $P_{adj} < 0.1$ ; Table S7, Fig. S6).

Due to the highly significant changes associated with tooth loss, we examined if these changes were similar across different mucosal sites. Analysis of buccal and tongue sites identified significant changes at those locations (Fig. S8). Using a heatmap to compare these changes at the buccal and tongue sites, we observed that loss of >15 teeth was associated with decreased abundance of *S. sanguinis* and *H. sputorum* and increased abundance of *S. parasanguinis* at both sites (Figure 5d). However, buccal sites uniquely exhibited increased *R. mucilaginosa* and decreased

*R. dentocariosa*, whereas tongue sites exhibited greatly decreased abundance of *H. parainfluenzae* and *H. parahaemolyticus* and increased *S. infantis* (Figure 5d). Analysis of tooth loss in FOM and palate sites did not identify any highly significant results (perhaps influenced by the smaller sample sizes).

Denture wearing also impacted on alpha diversity with reduced biodiversity (Shannon and Simpson) in those with partial or complete dentures (Fig. S8a). Denture wear was associated with significant shifts in population structure ( $p = 0.007$ ; Fig. S8b). We included denture wearing in a multivariate model including smoking and tooth loss to determine if any denture-specific changes could be identified (Fig. S8c). This identified a significant reduction in *S. oralis* subsp. *tigurinus*:

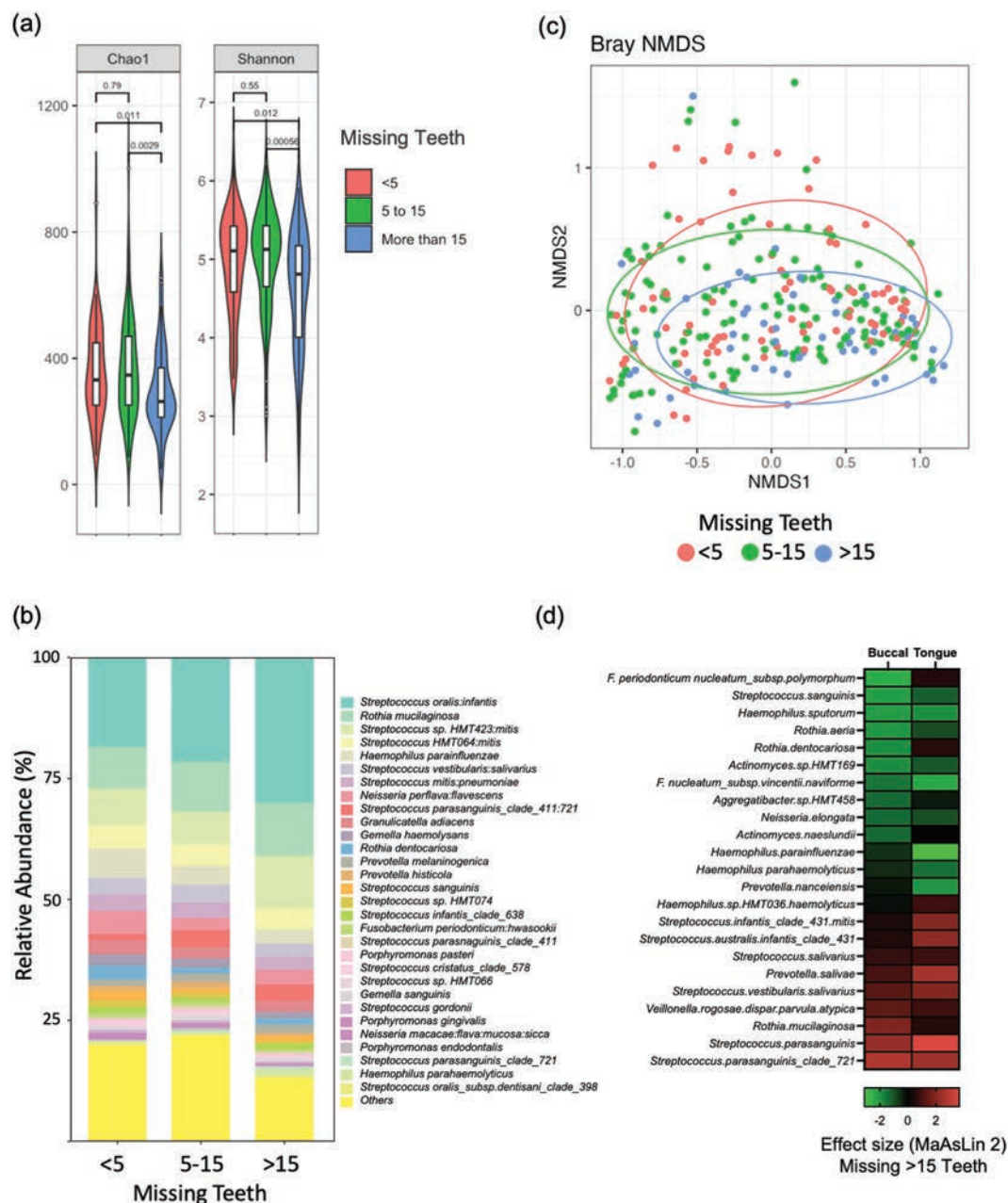


**Figure 5.** The influence of tooth loss on the mucosal microbiome (a) comparison of Chao1 and Shannon index values in individuals categorised by tooth loss (missing teeth < 5, 5–15 and > 15). (b) Bar plot showing relative abundance (%) of the 30 most abundant species according to number of missing teeth. (c) Community structure analysed by NMDS of Bray-Curtis dissimilarity values according to number of missing teeth. (d) Heatmap comparing the significant changes in species abundance associated with loss of > 15 teeth on buccal and tongue mucosa identified using MaAsLin 2 (Fig. S7). Colour intensity corresponds to effect size where reduced abundance (negative values) is indicated in green and increased abundance (positive values) is indicated in red.

*sanguinis* independent of tooth loss in both partial and complete denture wearers ( $P_{adj} < 0.04$ ; Figure S8c).

### Impact of age and gender on the mucosal microbiome

Differences in mucosal surface colonisation patterns associated with age and gender were investigated. A multivariate analysis was carried out on age category (<40, 40–60 and >60) and gender including smoking, oral hygiene, brushing frequency and

missing teeth (Fig. S9; Table S8). Age over 60 years was associated with a reduced abundance of *Actinomyces* spp., *S. oralis* subsp. *dentisani* and increased abundance of *V. rogosae:dispar*, *Solobacterium moorei* and *Megasphaera micronuciformis* (Fig. S9, Table S8).

Surprisingly, male gender was significantly associated ( $P_{adj} < 0.05$ ) with increased abundance of several taxa including *Selenomonas sputigena* and *S. parasanguinis* clade 721 (Fig. S9, Table S8). Increased levels of the periodontal disease-associated organisms *T. sockraskii* and *Peptostreptococcaceae*



[XI] [G5] *saphenum* were also detected in males, although the *p* values were less significant (Padj 0.11; Fig. S9, Table S8). Although current smoking was significantly associated with male gender (*p* = 0.047), these changes appeared to be independent of smoking and oral hygiene (Fig. S9).

## Discussion

The oral cavity is one of the most diverse microbiomes in the human body, and this biodiversity is facilitated by the wide range of niches available for microbes to colonise. However, few studies have investigated how these different sites are impacted by different habits, such as smoking and alcohol consumption, oral hygiene practices and the presence/absence of teeth. These studies are important as investigations of the microbiome and its role in mucosal diseases may be confounded by differences associated with these factors when comparing cases and controls. Although our study presents a large group of participants with disparate oral health, smoking and alcohol consumption patterns, it must be noted that most participants (195/256) were sampled while attending a clinic to assess mucosal dysplasia. Sampling at this clinic permitted us to recruit large numbers of current smokers and drinkers which would not have been achievable at general clinics. Although only healthy mucosa were sampled for this study, we cannot exclude that the presence of dysplasia at one site could impact colonisation at other oral sites. To control this, we included 'clinic attended' as a variable in our multivariate analysis (Table S3) which demonstrated that this factor did not significantly impact the major findings associated with site, smoking and the other oral health parameters analysed.

There is a general agreement in the literature that the oral microbiome displays site-specific differences and that certain sites can be clustered according to their colonisation patterns, with most agreeing that the dorsum and lateral tongue have similar microbiota while buccal mucosa, palate, FOM and gingival sites also tend to cluster [5,35–39]. In the current study, although oral sites had broadly similar compositions, tongue and palate bacterial communities were generally richer than buccal and FOM samples and this was generally due to increased abundance of the genera *Veillonella*, *Rothia*, *Neisseria*, *Granulicatella* and *Actinomyces* in tongue samples and increased *Prevotella*, *Veillonella* and *Actinomyces* in palate communities. Previously described streptococcal tropisms could also be identified in the current study, with higher abundance of *S. sanguinis* and *S. mitis* OTUs on buccal mucosa, increased *S. infantis* and *S. parasanguinis* on the tongue and greater abundance of *S. salivarius* and

*S. parasanguinis* in palate communities [40]. These species distributions are broadly similar to those described in the literature using a variety of technologies [5,9,40].

However, the high abundance of *Rothia* spp. we noted in this study on the tongue was not seen in some other NGS studies [5,35,37]. However, culture-based studies have shown that the tongue is the main habitat of *R. mucilaginosa* and have estimated relative abundances (~20%) in line with those described here [41]. Tobacco may influence carriage rates of *R. mucilaginosa* on the tongue [42] and the current study contains a high proportion of current smokers (46%) which may have influenced the carriage rates detected here.

Smoking is arguably the most significant influencer of the oral microbiome. The largest study of smoking published to date by Wu et al. sampled saliva and demonstrated that smoking resulted in decreased *Neisseria* and *Haemophilus* and increased *Veillonella* and *Streptococcus* species [17]. Most published studies that have sampled mucosal surfaces have been relatively small and have significantly more non-smokers than smokers in their sample populations, which may potentially influence reported results [24,25]. Our study contained significant numbers of smokers (46%) and our sample size allowed us to analyse the impact in individual mucosal sites. Our findings agreed with many previous studies, including Wu et al. [17], with current smoking being associated with reduced abundance of several species of *Neisseria*, *H. parainfluenza*, *L. mirabilis* and *R. aeria*. We also observed reduced *S. australis* and *S. sanguinis* and increased abundance of *S. parasanguinis*. Similar to Wu et al. [17], several Gram-negative bacteria associated with plaque from the genera *Aggregatibacter*, *Bergeyella*, *Capnocytophaga*, *Selenomonas* and *Kingella* also showed significant decreases in abundance in association with smoking (Table S5, **Error! Reference source not found.**). The species showing increased abundance associated with smoking were predominantly subgingival plaque and periodontal disease-associated pathogens, including species from the *Prevotella*, *Porphyromonas*, *Tannerella*, *Parvimonas*, *Filifactor*, *Bacteroidales* [G2] and *Peptostreptococcaceae* genera. Wu et al. [17] did not observe any increase in the levels of these periodontal pathogens in their large study of saliva recovered from smokers. However, studies of subgingival plaque have revealed that smoking is associated with increased levels of periodontal pathogens, which agrees with our findings [23]. Either the sampling method of Wu et al. (mouthwash) may not have detected these organisms or the levels of tooth loss and periodontal disease (which were not recorded) may have been very low in their population.

Unexpectedly, some of these smoking-related changes in abundance were found to be highly site-specific. While almost all surfaces examined (buccal, tongue, palate and FOM) exhibited reduced abundance of *Neisseria*, *Lautropia* and *Haemophilus*, depletion of these taxa appears to result in other site-specific shifts in abundance. Buccal sites in smokers showed a highly significant increase in *Prevotella* spp. and generally exhibited the most significant increases in the abundance of putative periodontal pathogens compared to other sites (Fig S2). Increased *Prevotella* was also observed on the palate in some smokers and previous smokers, although this was not statistically significant. This *Prevotella* shift was not evident in FOM samples where smokers exhibited a shift towards *Streptococcus* species. In tongue samples, PCA plots indicated that smokers exhibited decreased levels of *Neisseria*; however, smokers exhibited a divergence in community structure with increased abundance of either *Rothia* or *Streptococcus* species. The heterogeneity of community structure in tongue samples from smokers made it difficult to determine whether these shifts in abundance were statistically significant. It seems on the tongue at least, the response to smoking may be specific to the individual and could be influenced by factors such as the baseline microbiome community structure prior to smoking or other factors not examined here such as diet. For example, individual levels of *Rothia* may be influenced by a diet rich in green leafy vegetables [43].

Very few studies have examined the impact of oral hygiene and periodontal disease on the mucosal microbiome [38]. As subgingival plaque develops and the bacterial population increases, there is a shift to a more anaerobic population dominated by Gram-negative rods and spirochetes [44]. Although we did not measure periodontal disease *per se*, we recorded tooth number and denture wear and assessed plaque levels using the Simplified Greene and Vermillion Index (OHI-S) [26], along with brushing habits and mouthwash use. We found that all of these factors influenced oral mucosal colonisation to some degree. Plaque levels only had a moderate impact on gross community structure as assessed with the Bray-Curtis index but were associated with significantly increased abundance of several periodontitis-associated taxa, including several *Prevotella* species, *Tannerella* sp. HMT286, along with plaque-associated species of genera *Leptotrichia*, *Saccharibacteria* [G1] and *Lachnospiraceae* [G3]. It can take up to 4 days of no oral hygiene for dental plaque to significantly change in structure [44] and as only 5.1% (13 of 256) of our sample population brushed less than once per day, it is perhaps not unexpected that plaque accumulation did not significantly influence the microbiome structure. The impact of moderate plaque accumulation on mucosal surfaces observed here is likely the result of dispersal of plaque associated species onto mucosal

sites. In the current study, it is difficult to determine whether this is transient or long-term colonisation and follow-up studies with oral hygiene interventions would be required to determine this.

Surprisingly, brushing frequency appeared to have a more significant impact on mucosal community structure compared to oral hygiene. We found that brushing more than once per day was associated with a reduced abundance of streptococci and an overall increase in biodiversity. We did not ask participants who brushed their teeth regularly whether they also brushed mucosal surfaces. However, the increased biodiversity associated with brushing >1 per day may be associated with frequent rinsing of the oral cavity which may have the effect of reducing the overall abundance of streptococci. Similarly, mouthwash use was associated with reduced levels of some streptococci (most significantly the *S. mitis: pneumonia* OTU) with increased abundance of *G. adiacens* and *R. mucilaginosa*.

Tooth loss was associated with significant changes in alpha and beta diversity. We compared participants who had lost fewer than five teeth with those missing 5–15 teeth or more than 15 teeth. Although the absence of a full periodontal exam is a weakness of the current analysis, we used levels of tooth loss as proxy for periodontal disease status. Both categories (missing 5–15 teeth or more than 15 teeth) exhibited a significant shift in community structure, but only the >15 missing teeth category showed significant changes in richness and biodiversity. This loss in species richness was characterised by reduced abundance of many plaque-associated species including *C. durum*, *L. mirabilis*, *Haemophilus* spp., *Actinomyces* spp., *Rothia* spp. and *Fusobacterium* spp., presumably related to the loss of tooth surfaces available for colonisation. Those in the 5–15 tooth loss category exhibited microbial signatures indicative of active periodontal disease including increased abundance of *E. corrodens*, *P. micra*, *Treponema* sp. HMT237, *Fretibacterium* spp. and *Fusobacterium* spp., presumably due to the greater number of disease active sites shedding pathogens on mucosal surfaces. Socranksy and Huffajee [38] using checkerboard DNA – DNA hybridization also concluded that the soft tissue microbiota differs between healthy subjects and those who have periodontal disease or are edentulous. Their study also showed that mucosal colonisation (tongue and buccal) by *F. nucleatum* and *E. corrodens* was significantly increased in those with periodontitis. Although the current analysis would benefit from periodontal health data, the nature of the microbiome shifts associated with loss of >5 teeth indicate that tooth loss is a reasonable indicator of the presence of periodontal disease.



Interestingly, we also noted site-specific changes in the buccal and tongue communities associated with loss of >15 teeth. On tongue communities, we specifically observed reduced abundance of *H. parainfluenza* and *H. parahaemolyticus* with a concomitant increase in *H. haemolyticus*. Conversely on buccal mucosa, tooth loss was associated with loss of *R. dentocariosa* and increased *R. mucilaginosa*. These site-specific changes may be linked to the baseline microbiomes at these sites. For example, *R. dentocariosa* normally has a higher abundance on buccal mucosa. Tooth loss may result in a decrease in shedding of plaque associated *R. dentocariosa* on buccal mucosa, which may provide space for *R. mucilaginosa* to colonise.

Finally, we also identified some age and gender associated differences in a multivariate analysis with smoking, tooth-loss and other oral hygiene metrics. Loss of *Actinomyces* sp. HMT175 and *S. oralis* subsp. *dentisani* was deemed highly significant in those over 60 with an increase in *V. rogosae:dispar*. As these changes are independent of tooth loss, other age-related factors such as immunosenescence or dry mouth may be responsible for this. Interestingly, men tended to harbour increased levels of periodontal pathogens *T. sockranskii* and *Peptostreptococcaceae* [XI][G5] *saphenum*. Our multivariate analysis indicated that this was independent of smoking and tooth-loss. The incidence of periodontal disease is known to be higher in males than females and could be related to different abundances of periodontal pathogens. Liu et al. [45] carried out an extensive analysis of gender differences in the oral microbiome and also found several putative periodontal pathogens to have greater abundance in men, including a species of *Treponema*. These changes were linked to hormonal and metabolic differences between the sexes.

We have demonstrated that smoking, tooth loss and oral site are the most significant influencers of the oral mucosal microbiome. The influence of site suggests that niches such as palate, tongue and buccal mucosa should be considered as independent variables in studies of oral mucosal diseases. Our study highlights the need for carefully matched controls in studies examining mucosal colonisation pattern associated with specific pathologies. Due to the difficulties in matching cases and control subjects for these myriad factors, it may be useful to take samples from contralateral healthy mucosal sites (if present) so each patient can act as their own matched control.

## Disclosure statement

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no financial support for this work that could have influenced its outcome.

## Funding

GPM and CMH are funded by a grant from the Irish Health Research Board (Grant no. ILP-POR-2019-030).

## Author contributions

SG: DNA extraction, data analysis, manuscript preparation; SA: Sample collection, oral health data collection, manuscript preparation; CMH: hypothesis development, funding, manuscript preparation; GPM: hypothesis development, funding, data analysis, manuscript preparation.

## ORCID

Gary P. Moran  <http://orcid.org/0000-0003-0469-1788>

## References

- [1] de Jesus VC, Shikder R, Oryniak D, et al. Sex-based diverse plaque microbiota in children with severe caries. *J Dent Res*. 2020;99(6):703–712. doi: [10.1177/0022034520908595](https://doi.org/10.1177/0022034520908595)
- [2] Dzidic M, Collado MC, Abrahamsson T, et al. Oral microbiome development during childhood: an ecological succession influenced by postnatal factors and associated with tooth decay. *ISME J*. 2018;12(9):2292–2306. doi: [10.1038/s41396-018-0204-z](https://doi.org/10.1038/s41396-018-0204-z)
- [3] Califf KJ, Schwarzberg-Lipson K, Garg N, et al. Multi-omics analysis of periodontal pocket microbial communities pre- and posttreatment. *mSystems*. 2017;2(3):e00016–17. doi: [10.1128/mSystems.00016-17](https://doi.org/10.1128/mSystems.00016-17)
- [4] Iniesta M, Chamorro C, Ambrosio N, et al. Subgingival microbiome in periodontal health, gingivitis and different stages of periodontitis. *J Clin Periodontol*. 2023;1(7):905–920. doi: [10.1111/jcpe.13793](https://doi.org/10.1111/jcpe.13793)
- [5] Aas JA, Paster BJ, Stokes LN, et al. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol*. 2005;43(11):5721–5732. doi: [10.1128/JCM.43.11.5721-5732.2005](https://doi.org/10.1128/JCM.43.11.5721-5732.2005)
- [6] Bik EM, Long CD, Armitage GC, et al. Bacterial diversity in the oral cavity of 10 healthy individuals. *ISME J*. 2010;4(8):962–974. doi: [10.1038/ismej.2010.30](https://doi.org/10.1038/ismej.2010.30)
- [7] Dewhirst FE, Chen T, Izard J, et al. The human oral microbiome. *J Bacteriol*. 2010;192(19):5002–5017. doi: [10.1128/JB.00542-10](https://doi.org/10.1128/JB.00542-10)
- [8] Gevers D, Knight R, Petrosino JF, et al. The human microbiome project: a community resource for the healthy human microbiome. *PLoS Biol*. 2014;10(8):e1001377. doi: [10.1371/journal.pbio.1001377](https://doi.org/10.1371/journal.pbio.1001377)
- [9] Segata N, Haake SK, Mannon P, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol*. 2012;13(6):R42–R42. doi: [10.1186/gb-2012-13-6-r42](https://doi.org/10.1186/gb-2012-13-6-r42)
- [10] Huttenhower C, Gevers D, Knight R, et al. Structure, function and diversity of the healthy human microbiome. *Nature*. 2014;486:207.
- [11] Wade WG, Prosdociimi EM. Profiling of oral bacterial communities. *J Dent Res*. 2020;99(6):621–629. doi: [10.1177/0022034520914594](https://doi.org/10.1177/0022034520914594)

- [12] Escapa IF, Huang Y, Chen T, et al. Construction of habitat-specific training sets to achieve species-level assignment in 16S rRNA gene datasets. *Microbiome*. 2020;8(1):65. doi: [10.1186/s40168-020-00841-w](https://doi.org/10.1186/s40168-020-00841-w)
- [13] Eren AM, Borisy GG, Huse SM, et al. Oligotyping analysis of the human oral microbiome. *Proc Natl Acad Sci*. 2014;111(28):E2875–E2884. doi: [10.1073/pnas.1409644111](https://doi.org/10.1073/pnas.1409644111)
- [14] Crielaard W, Zaura E, Schuller AA, et al. Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health. *BMC Med Genomics*. 2011;4(1):22. doi: [10.1186/1755-8794-4-22](https://doi.org/10.1186/1755-8794-4-22)
- [15] Mason MR, Chambers S, Dabdoub SM, et al. Characterizing oral microbial communities across dentition states and colonization niches. *Microbiome*. 2018;6(1):67. doi: [10.1186/s40168-018-0443-2](https://doi.org/10.1186/s40168-018-0443-2)
- [16] Gomez A, Espinoza JL, Harkins DM, et al. Host Genetic control of the oral microbiome in health and disease. *Cell Host Microbe*. 2017;22(3):269–278.e3. doi: [10.1016/j.chom.2017.08.013](https://doi.org/10.1016/j.chom.2017.08.013)
- [17] Wu J, Peters BA, Dominianni C, et al. Cigarette smoking and the oral microbiome in a large study of American adults. *ISME J*. 2016;10(10):2435–2446. doi: [10.1038/ismej.2016.37](https://doi.org/10.1038/ismej.2016.37)
- [18] Nearing JT, DeClercq V, Limbergen JV, et al. Assessing the variation within the oral microbiome of healthy adults. *mSphere*. 2020;5(5): doi: [10.1128/mSphere.00451-20](https://doi.org/10.1128/mSphere.00451-20)
- [19] Marsh PD. Microbial ecology of dental plaque and its significance in health and disease. *Adv Dent Res*. 1994;8(2):263–271. doi: [10.1177/08959374940080022001](https://doi.org/10.1177/08959374940080022001)
- [20] Griffen AL, Beall CJ, Campbell JH, et al. Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *ISME J*. 2011;6(6):1176–1185. doi: [10.1038/ismej.2011.191](https://doi.org/10.1038/ismej.2011.191)
- [21] Socransky SS, Haffajee AD, Cugini MA, et al. Microbial complexes in subgingival plaque. *J Clin Periodontol*. 2005;25(2):134–144. doi: [10.1111/j.1600-051X.1998.tb02419.x](https://doi.org/10.1111/j.1600-051X.1998.tb02419.x)
- [22] Hsiao J-R, Chang C-C, Lee W-T, et al. The interplay between oral microbiome, lifestyle factors and genetic polymorphisms in the risk of oral squamous cell carcinoma. *Carcinogenesis*. 2018;39(6):778–787. doi: [10.1093/carcin/bgy053](https://doi.org/10.1093/carcin/bgy053)
- [23] Moon JH, Lee JH, Lee JY. Subgingival microbiome in smokers and non-smokers in Korean chronic periodontitis patients. *Mol Oral Microbiol*. 2015;30:227–241. doi: [10.1111/omi.12086](https://doi.org/10.1111/omi.12086)
- [24] Charlson ES, Chen J, Custers-Allen R, et al. Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS One*. 2010;5(12):e15216. doi: [10.1371/journal.pone.0015216](https://doi.org/10.1371/journal.pone.0015216)
- [25] Thomas AM, Gleber-Netto FO, Fernandes GR, et al. Alcohol and tobacco consumption affects bacterial richness in oral cavity mucosa biofilms. *BMC Microbiol*. 2014;14(1):250. doi: [10.1186/s12866-014-0250-2](https://doi.org/10.1186/s12866-014-0250-2)
- [26] Greene JG, Vermillion JR. The simplified oral hygiene index. *J Am Dent Assoc*. 1964;68(1):7–13. doi: [10.14219/jada.archive.1964.0034](https://doi.org/10.14219/jada.archive.1964.0034)
- [27] Amer A, Galvin S, Healy CM, et al. The microbiome of potentially malignant oral leukoplakia exhibits enrichment for *Fusobacterium*, *Leptotrichia*, *Campylobacter*, and *Rothia* species. *Front Microbiol*. 2017;8:2391. doi: [10.3389/fmicb.2017.02391](https://doi.org/10.3389/fmicb.2017.02391)
- [28] Frank JA, Reich CI, Sharma S, et al. Critical evaluation of two primers commonly used for Amplification of bacterial 16S rRNA genes. *Appl Environ Microb*. 2008;74(8):2461–2470. doi: [10.1128/AEM.02272-07](https://doi.org/10.1128/AEM.02272-07)
- [29] Comeau AM, Douglas GM, Langille MGI, et al. Microbiome helper: a custom and streamlined workflow for microbiome Research. *mSystems*. 2017;2(1):e00127–16. doi: [10.1128/mSystems.00127-16](https://doi.org/10.1128/mSystems.00127-16)
- [30] Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581–583. doi: [10.1038/nmeth.3869](https://doi.org/10.1038/nmeth.3869)
- [31] McMurdie PJ, Holmes S, Watson M. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8(4):e61217. doi: [10.1371/journal.pone.0061217](https://doi.org/10.1371/journal.pone.0061217)
- [32] Xu S, Li Z, Tang W, et al. MicrobiotaProcess: a comprehensive R package for managing and analyzing microbiome and other ecological data within the tidy framework. *Innovation (Camb)*. 2023;4:100388.
- [33] Barnett D, Arts I, Penders J. microViz: an R package for microbiome data visualization and statistics. *J Open Source Softw*. 2021;6(63):3201. doi: [10.21105/joss.03201](https://doi.org/10.21105/joss.03201)
- [34] Mallick H, Rahnavard A, McIver LJ, et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol*. 2021;17(11):e1009442. doi: [10.1371/journal.pcbi.1009442](https://doi.org/10.1371/journal.pcbi.1009442)
- [35] Mager DL, Ximenez-Fyvie LA, Haffajee AD, et al. Distribution of selected bacterial species on intraoral surfaces. *J Clin Periodontol*. 2003;30(7):644–654. doi: [10.1034/j.1600-051X.2003.00376.x](https://doi.org/10.1034/j.1600-051X.2003.00376.x)
- [36] Welch JLM, Dewhirst FE, Borisy GG. Biogeography of the oral microbiome: the site-specialist hypothesis. *Annu Rev Microbiol*. 2019;73(1):335–358. doi: [10.1146/annurev-micro-090817-062503](https://doi.org/10.1146/annurev-micro-090817-062503)
- [37] Segata N, Haake SK, Mannon P, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol*. 2012;13(6):R42. doi: [10.1186/gb-2012-13-6-r42](https://doi.org/10.1186/gb-2012-13-6-r42)
- [38] Socransky SS, Haffajee AD. Periodontal microbial ecology. *Periodontol 2000*. 2005;38(1):135–187. doi: [10.1111/j.1600-0757.2005.00107.x](https://doi.org/10.1111/j.1600-0757.2005.00107.x)
- [39] Diaz PI, Dupuy AK, Abusleme L, et al. Using high throughput sequencing to explore the biodiversity in oral bacterial communities. *Mol Oral Microbiol*. 2012;27(3):182–201. doi: [10.1111/j.2041-1014.2012.00642.x](https://doi.org/10.1111/j.2041-1014.2012.00642.x)
- [40] McLean AR, Torres-Morales J, Dewhirst FE, et al. Site-tropism of streptococci in the oral microbiome. *Mol Oral Microbiol*. 2022;37(6):229–243. doi: [10.1111/omi.12387](https://doi.org/10.1111/omi.12387)
- [41] Uchibori S, Tsuzukibashi O, Kobayashi T, et al. Localization of the genus *Rothia* in the oral cavity. *Int J Oral-Medical Sci*. 2013;11(3):207–210. doi: [10.5466/ijoms.11.207](https://doi.org/10.5466/ijoms.11.207)
- [42] Halboub E, Al-Ak'hali M S, Alamir A H, Homeida H E, Baraniya D, Chen T and Al-Hebshi N Noor. (2020).

- Tongue microbiome of smokeless tobacco users. *BMC Microbiol*, 20(1), doi: [10.1186/s12866-020-01883-8](https://doi.org/10.1186/s12866-020-01883-8)
- [43] Velmurugan S, Gan JM, Rathod KS, et al. Dietary nitrate improves vascular function in patients with hypercholesterolemia: a randomized, double-blind, placebo-controlled study. *Am J Clin Nutrition*. 2015;103(1):25–38. doi: [10.3945/ajcn.115.116244](https://doi.org/10.3945/ajcn.115.116244)
- [44] Teughels W, Christof G, Quirynen M, Jakubovics, N Biofilm and Periodontal Microbiology. Carranza's Clinical Periodontology. Elsevier Health Sciences. 2012; 132–169.
- [45] Liu X, Tong X, Jie Z, et al. Sex differences in the oral microbiome, host traits, and their causal relationships. *iScience*. 2023;26(1):105839. doi: [10.1016/j.isci.2022.105839](https://doi.org/10.1016/j.isci.2022.105839)