# Investigation of the Microbiological Profile of Oral Leukoplakia

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

**Abdrazak Ali Amer**

**March 2019**

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

## Table of Contents

## List of Figures

## List of Tables

## Abbreviations

DDUH……Dublin Dental University Hospital
CFU/ml …… Colony-forming unit per milliliter
16S rRNA……16S ribosomal gene
ACDH…… Acetaldehyde dehydrogenase genes
ACH…… Acetaldehyde
ATP……Adenosine triphosphate
BHI-A……Brain heart Infusion-Agar
BHI-B……Brain heart Infusion-broth
BLAST……The basic local alignment search tool
bp…….. Base pairs
$CO_2$…….. Carbon dioxide
CRC… ……Colorectal cancer
DAPI……4′,6-diamidino-2-phenylindole
DCF-DA……Dihyrodichlorofluorescein diacetate
$dH_2O$…… Double distilled water
DIT…….. Dublin Institute of Technology
DMEM……Dulbecco's modified eagle medium (DMEM)
DMSO…… dimethyl sulfoxide
DNA……Deoxyribonucleic acid
EDTA……Ethylendiaminetetaacetic acid
FCS……..... Foetal calf serum
GI …… Gastrointestinal
h……Hour
$H_2O_2$..... ……..... …….....Hydrogen peroxide
HBV ……Hepatitis B
HCV……Hepatitis C
HOMD……Human oral microbiome
HPV …… Human Papilloma Virus
LDA……linear discriminant analysis
LEfSe…… Linear discriminant analysis effect size
LPS……Lipopolysaccharides
M……Molar
mg……Milligram
min……Minute
ml……Millilitre
mM……Millimolar
n ……Number
NAC…… ……N-acetylcysteine
NDK…… Nucleoside diphosphate kinase
NGS…… Next Generation Sequencing
NOC ……N-nitroso compounds
NMDS ……...Non-Metric Multidimensional Scaling

OD……Optical density
OSCC …… Oral Squamous Cell Carcinoma
OTU…… Operational Taxonomic Unit
*p*……*p* value (to determine significant and non-significant differences)
PCR……Polymerase chain reaction
PTFE……Polytetrafluoroethylene
qPCR……Quantitative Real-Time PCR
ROS ……Reactive oxygen species
spp……Species
subsp……Subspecies
v/v……Volume per volume (ml/100ml)
V1……Region 1 on the 16S rRNA
V2……Region 2 on the 16S rRNA
V3……Region 3 on the 16S rRNA
V4……Region 4 on the 16S rRNA
WHO…… Word Health Organization
µM……Micromole

## Acknowledgments

This study was supported by the Ministry of Higher Education and Scientific Research Libya.

I would like to express my sincere thanks to Trinity College Dublin and the Dublin Dental University Hospital (DDUH) for all facilities and help that was provided to me from all departments. I would like to thank my supervisor **Gary Moran** for all his time, and who was always there and giving advice, guidance, and steering me in the right direction whenever I had trouble or had any questions about the project.

This study could not have been completed without the help of **Dr. Claire Healy** who carried out sample collection and sample diagnostics, thank you for the great job.

I must express my very profound gratitude to all my family and friends for providing me with unfailing support and continuous encouragement to me during my study years.

To my wife (**Wafa),** my kids (**Raged, Ali, Omar** and **Mohab**)

**Thank you all**

**Abdrazak Ali Amer**

## Declaration

I declare that this thesis, which I submitted to Trinity college Dublin (TCD) for examination in consideration of the award of a higher degree **< Degree of Doctor of Philosophy >** that this PhD thesis is purely my own personal effort.
I give a permission to the TCD to allow this thesis for what the university sees in public interest.

Signed _____

Student number _____

Date 29/3/2019

## Summary

Oral leukoplakia (OLK) presents as a white patch on the oral mucosa and is recognised as a potentially malignant lesion. Although colonisation of these patches with yeast such as *Candida albicans* has been reported, little is known about the bacterial microbiome of these lesions. Recent studies have indicated the microbiome of oral squamous cell carcinoma (OSCC) is different from healthy mucosa and we wished to determine if similar changes occurred in OLK prior to malignant progression. In the current study, analysed of the OLK microbiome in 36 patients compared to contralateral healthy mucosal tissue from the same patients and 36 healthy control subjects to determine if specific microbial enrichments could be identified early in the malignant process that could play a role in the progression of the disease. This was firstly carried out by sequence analysis of the V1–V2 region of the bacterial 16S rRNA gene using the Illumina MiSeq. Oral leukoplakia exhibited increased abundance of Fusobacteria and reduced levels of Firmicutes ( $p < 0.01$). *Candida* colonisation was also more prevalent in leukoplakia patients relative to healthy controls. Bacterial colonisation patterns on oral leukoplakia were highly variable and 5 distinct bacterial clusters were discerned. These clusters exhibited co-occurrence of *Fusobacterium, Leptotrichia*, and *Campylobacter* species (Pearson $p < 0.01$), which is very similar to the microbial co-occurrence patterns observed on colorectal cancers. Severe dysplasia was associated with elevated levels of *Leptotrichia* spp., and *Campylobacter concisus* ($p < 0.05$). Further sequence analysis using primers specific for the V3-V4 region of the 16S rRNA confirmed these data. In general, it was found that V1-V2 sequencing was more effective and could identify more OTUs than V3-V4 sequence analysis.

Smoking, *Candida* spp. carriage and alcohol consumption were all shown to have specific effects on the oral microbiome. *Candida* carriage was associated with increased

carriage of *F. nucleatum* whereas smoking was negatively associated with carriage of *F. nucleatum*. Alcohol consumption was associated with increased levels of *Campylobacter* spp.

An increasing of abundance some species including *Rothia mucilaginosa* and other Gram-positive were observed on OLK samples. *R. mucilaginosa* spp., were isolated from several OLK sites and assessed its capacity to generate acetaldehyde (ACH) *in vitro* using aldehyde detection kits and Gas-Chromatography Mass-Spectrometry (GC-MS). *R. mucilaginosa* was shown to produce ACH from ethanol at similar levels to *C. albicans* and *Neisseria mucosa*, with only one isolate (41C2.2) producing >150 μM ACH from 10 mM ethanol. These levels of ACH were shown to induce oxidative stress in cultured oral keratinocytes.

In summary, OLK exhibits an altered microbiota that has similarities to the microbiome of colorectal cancer. *R. mucilaginosa* has the potential to drive malignant change through the production of mutagenic levels of ACH.

# Chapter 1

General introduction

## 1.1 Human disease

### 1.1.1 Human microbiome

The study of the human microbiome began with the work of Antonie van Leewenhoek in the early 1680s, when he recognised different types of microbes in a comparison of his oral and faecal matter and gave these entities the name of "Dierken", meaning small, lively object (Gao *et al.,* 2018). However, the word microbiome was not coined until much later in the 20$^{th}$ century by Joshua Lederberg, who defined the microbiome as all microorganisms living on or in the human body. There are an estimated 10 to 100 trillion symbiotic microbial cells in the human body, including different types such as bacteria, archaea, viruses and fungi. Studies of the human microbiome have sharply increased in recent decades and reached more than 10,000 publications in 2018. Microbes associated with the human body are found in various habitats such as the skin, gut, vagina and the oral cavity (Fig. 1.1). However, some of these microbes are necessary for a healthy body, but some are a cause of disease. The Human Microbiome Project (http://nihroadmap.nih.gov/hmp) was launched in 2007 with the goal of establishing the constituents of the healthy microbiome in order to better understand their effects on human health and disease.

### 1.1.2 Gastrointestinal (GI) Microbiome

The GI tract of humans is an intense area for microbiome research and is associated with a diverse range of microbes with roles in health and disease (Blaut and Clavel, 2007; Mobeen, 2014). The first member of the GI microbiome to be characterised was *Escherichia coli*, first cultured in 1885 (Shulman *et al.,* 2007). The GI tract is one of the richest habitats in the human body; however, this diversity remained uncharacterised until

improvements in anaerobic culture techniques in the 1960s allowed the culture and characterisation of the largely anaerobic microbiota of the GI tract (Rajilic *et al.,* 2014).

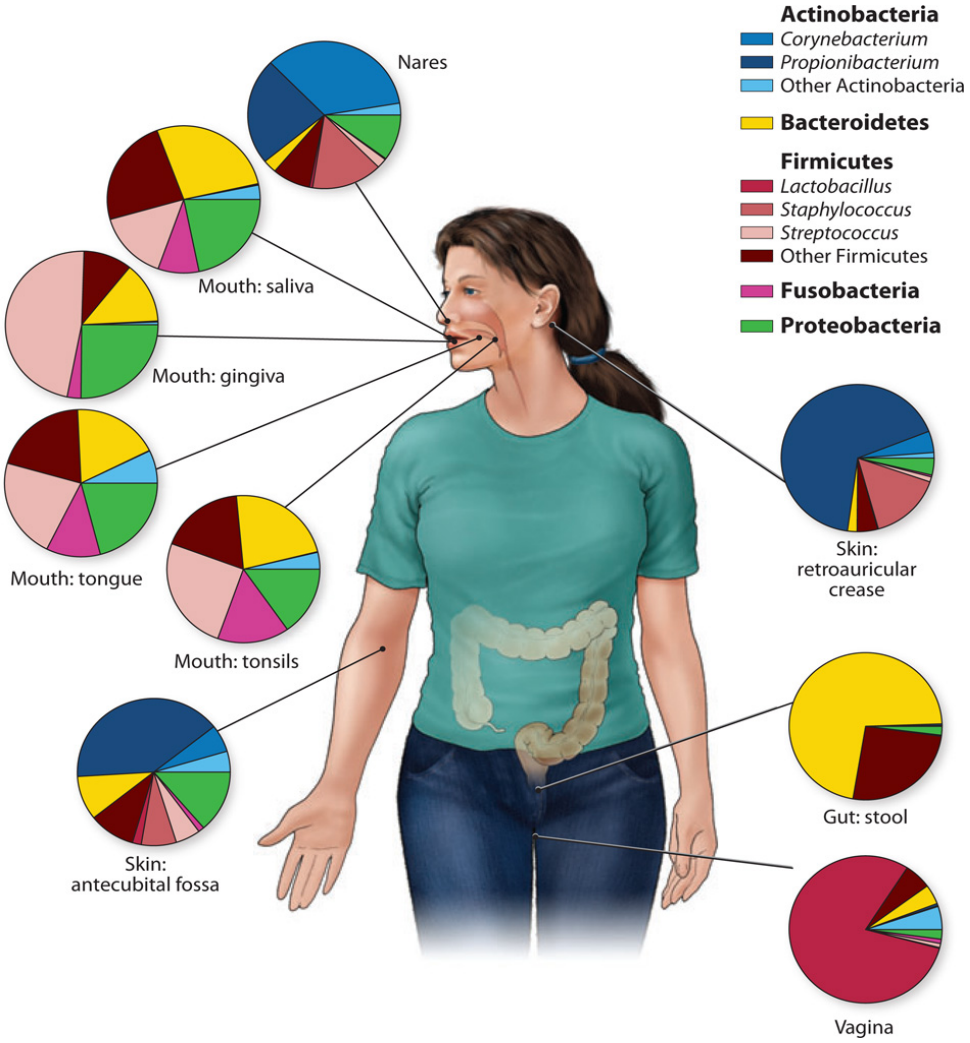



Figure 1.1 The different habitats and constituents of the human microbiome

Figure reproduced with permission from Grice and Serge, (2012).

More recently, new technologies such as the polymerase chain reaction (PCR), whole genome sequencing and next generation sequencing (NGS) have greatly improved our understanding of the microbiome of the GI tract. These technologies have not only increased the number of known microorganisms in the GI tract, they have also allowed

identification of microbes that have not yet been cultured, which were identified on the basis of target genes such as the 16S rRNA, which is present in almost in all bacterial species (Janda and Abbott, 2007). The microbiome of the GI tract is dominated by anaerobic Bacteroidetes, Clostridia and Proteobacteria. These organisms have an important role in maintaining the health of the GI tract in terms of preventing colonisation with enteric pathogens (i.e. colonisation resistance) and in the generation of essential nutrients, such as vitamin K2. In addition to bacteria, many fungi are to be found in the gastrointestinal tract and analysis of 18S rRNA sequences has shown the presence of phyla such as the Ascomycota, Mucorales, Scleroderma and Basidiomycota in the GI tract (Tang *et al*., 2015).

### 1.1.3 Skin microbiome

Human skin is an important barrier covering and protecting our bodies from microbial infection. However, it is a common habitat for different microbes, including some that benefit the human body and some that cause disease (Findley and Grice, 2014). This skin microbiome also comprises viruses, fungi and bacteria. Bacterial species on the skin mainly belong to 4 dominant phyla, the Proteobacteria, Actinobacteria, Bacteroidetes and Firmicutes phyla (Nakatsuji *et al.,* 2017). The dominant genera are *Staphylococcus*, *Propionibacterium*, and *Corynebacterium* and the distribution of these types is different from site to site on the skin. For example, the bacterium *Staphylococcus aureus* is a common resident of the skin and can in susceptible individuals become pathogenic and is usually associated with patients with atopic dermatitis (AD) (Nakatsuji *et al.,* 2017). Another example is the bacterium *Propionibacterium acnes,* a normal resident that can become pathogenic with the onset of adolescence and is associated with acne (Findley and Grice, 2014).

### 1.1.4 Vaginal microbiome

The human vagina is a rich habitat for different microbes, including viruses, fungi and bacteria. *Lactobacillus* spp. are common bacterial residents and are associated with vaginal health (Donders *et al.,* 2000), and may play a role in protecting from other pathogenic species such as *Neisseria gonorrhoeae* (Graver and Wade, 2011). Also, fungi can be found in the vagina, such as *C. albicans* which is an important cause of vaginal infection. *C. albicans* is found in approximately in 20% of women without causing any disease symptoms. However, vaginal infection can occur following use of antibiotics, due to hormonal changes or pregnancy. Most infections are easily treated with antifungals but some women will suffer from recurrent infections (Ma *et al.,* 2012).

### 1.1.5 Oral microbiome

The oral microbiome is another microbially rich environment with over 700 bacterial species identified to date (He and Shi, 2009). These species include both cultivated and non-cultivated species identified by DNA sequencing. The oral cavity also contains other microbes such as *Candida* spp. and viruses. Approximately one gram of the oral biomass typically contains about 100 million microbes and these microbes have the ability to double their numbers every 20 minutes under optimal conditions. Although rich in microbial life, the oral cavity has several natural barriers to stop microbes penetrating the body, including saliva which contains antibodies and lysozyme and the immune cells of the mucosa (Fabian *et al.,* 2012) . The oral cavity presents numerous different habitats for microbes to colonise and a knowledge of these niches is important to understand the diversity of the oral microbiome .

## 1.2 Structure of the oral cavity

The oral cavity is commonly affected by many diseases including oral cancer, and many of these diseases have microbial aetiologies. The oral cavity is the link between the internal and the external compartments of the body and it has the ideal conditions for microbial growth including moisture, a warm environment and the availability of plentiful dietary nutrients. The oral cavity is inhabited by different microbes including viruses, fungi, archaea, and bacteria with representatives of more than 13 bacteria phyla and 199 genera. The 6 most abundant bacterial phyla (Table 1.1) account for approximately 99% of all bacteria species present. The Human Oral Microbiome Database (HOMD; http://www.homd.org/index.php) was established by researchers at the Forsyth Institute to catalogue and curate the DNA sequences (both 16S fragments and whole genomes) associated with this microbiome, and this database now represents over 770 microbial species.

Within the oral cavity, there are numerous different micro environments with their own unique conditions allowing for the growth of different microbial communities. The exposed mucosal surfaces are generally covered in saliva and allow the growth of numerous aerobic species such as streptococci, *Neisseria* spp. and *Haemophilus* spp., Some areas, such as the floor of the mouth, may have reduced levels of oxygen and have greater levels of anaerobes such as *Veillonella* spp. and *Prevotella* spp. The hard surfaces are also colonised by streptococci and *Actinomyces* spp., which act as the foundation for the development of dental plaque.

A study examining the microbiomes of the dorsum tongue, lateral tongue, buccal cavity, supra-gingival plaque, sub-gingival plaque, hard palate, soft palate, maxillary anterior vestibule and tonsils found that each site had different bacterial species present. Species common to all oral sites included the genera *Gemella*, *Granulicatella*, *Streptococcus* and

*Veillonella*, with *Streptococcus mitis* being the species most commonly found on all sites. Some species were located specifically in certain niches such as *R. dentocariosa* on the tongue, *Streptococcus sanguinis* and *Streptococcus gordonii* on the tooth surface and *Campylobacter* and *Capnocytophaga* in sub gingival plaque (Aas *et al.,* 2005).

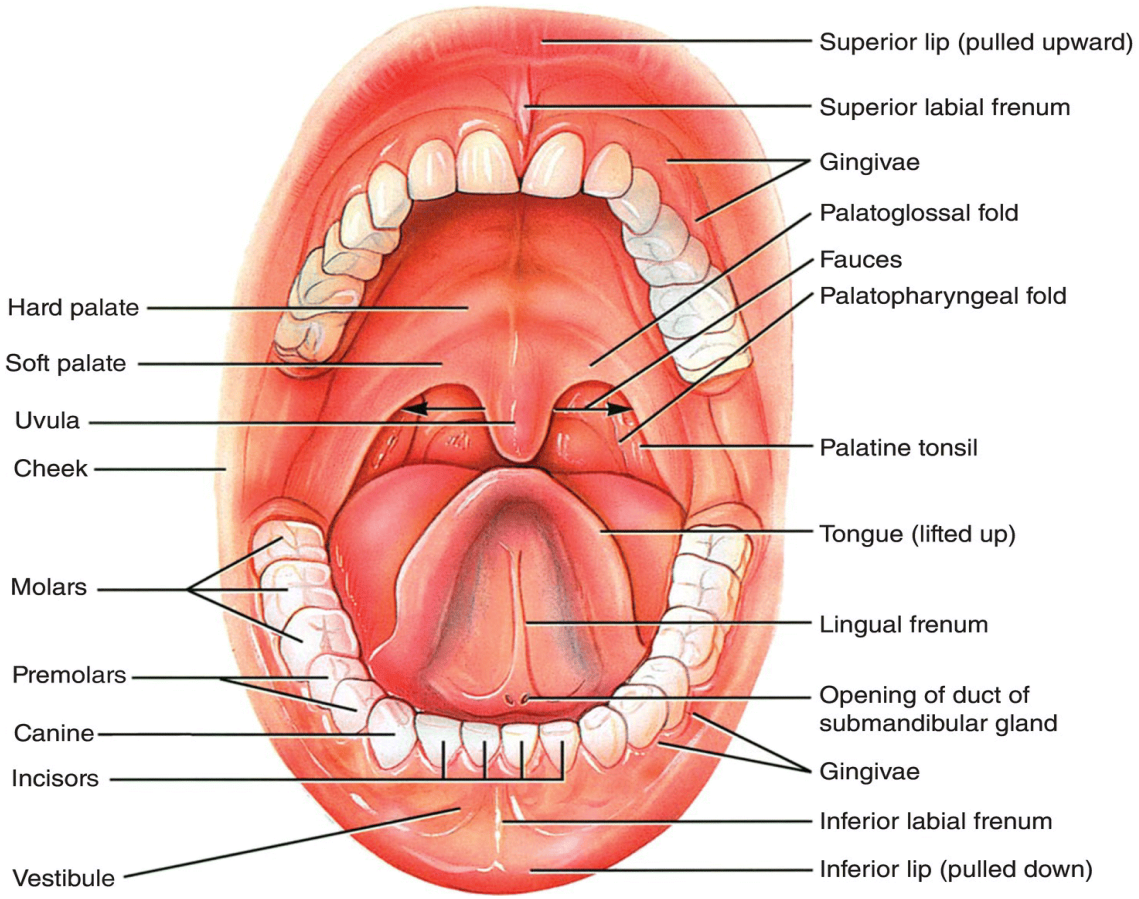



**Figure 1.2** The structure of the oral cavity

Figure reproduced with a permission of John Wiley and Sons Tortora, G.J. and Grabowski, S.R. 2000. Principles of Anatomy and Physiology, 9th edition, Wiley, New York

Table 1.1 The six common oral bacteria phyla and example species

| Phylum name | Species name | Gram-stain |
|---|---|---|
| **Firmicutes** | ***Streptococcus parasanguinis*** | **Gram-positive coccus** |
| | ***Streptococcus mitis*** | **Gram-positive coccus** |
| | ***Staphylococcus epidermidis*** | **Gram-positive coccus** |
| | ***Viellonella dispar*** | **Gram-negative coccus** |
| | ***Gemella sanguinis*** | **Gram-positive coccus** |
| **Proteobacteria** | ***Haemophilus parainfluenza*** | **Gram-negative coccobacillus** |
| | ***Lautropia mirabilis*** | **Gram-negative coccus** |
| | ***Eikenella corrodens*** | **Gram-negative bacillus** |
| | ***Kingella oralis*** | **Gram-negative coccobacillus** |
| | ***Neisseria mucosa*** | **Gram-negative coccus** |
| **Fusobacteria** | ***Fusobacterium periodonticum*** | **Gram-negative fusiform** |
| | ***Fusobacterium nucleatum*** | **Gram-negative fusiform** |
| | ***Leptotrichia hongkongensis*** | **Gram-negative bacillus** |
| | ***Leptotrichia shahii*** | **Gram-negative bacillus** |
| | ***Sneathia sanguinegens*** | **Gram-negative bacillus** |
| **Bacteroidetes** | ***Capnocytocphaga ochracea*** | **Gram-negative bacillus** |
| | ***Porphyromonas endodontalis*** | **Gram-negative bacillus** |
| | ***Porphyromonas gingivalis*** | **Gram-negative coccobacillus** |
| | ***Prevotella melaninogenica*** | **Gram-negative bacillus** |
| | ***Tannerella forsythia*** | **Gram-negative bacillus** |
| **Actinobacteria** | ***Actinomyces naeslundii*** | **Gram-positive bacillus** |
| | ***Actinomyces georgiae*** | **Gram-positive bacillus** |
| | ***Actinomyces graevenitzii*** | **Gram-positive bacillus** |
| | ***Corynebacterium accolens*** | **Gram-positive bacillus** |
| | ***Rothia mucilaginosa*** | **Gram-positive coccus** |
| **Spirochaetes** | ***Treponema socranskii*** | **Gram-negative spiral** |
| | ***Treponema denticola*** | **Gram-negative spiral** |

## 1.3 Oral disease

Treatment and prevention of oral disease is an important public health issue (Myburgh *et al.,* 2004). A recent survey of 150 individuals aged from 5 to 45 years carried out by the World Health Organisation (WHO) in Nigeria showed that 91.3% of the 150 individuals had an oral health condition, most commonly accumulations of dental plaque (66.0%), enamel wear (15%) and dental caries (13%) (Tobin and Ajayi, 2017). The oral cavity microbiome composition can vary greatly due to oral hygiene. Dental hygiene has an important effect on oral health and in the absence of regular hygiene the surface of a single tooth can contain between 100 million and one billon adherent bacteria. Although these bacteria are not all harmful, changes in the normal structure of the microbiome, in particular the build-up of dental plaque, can lead to disease. Dental plaque is a complex biofilm and is associated with more than 500 different bacterial species (Lamont *et al.,* 2002). Supragingival plaque is predominantly composed of Gram-positive bacteria such as *S. gordonii* and *Actinomyces* spp., whereas subgingival plaque is associated with a Gram-negative microbiome including *Prevotella* spp. and *Porphyromonas* spp. (Socransky and Haffajee, 1992). Subgingival plaque presents a new environment for colonisation with Gram-negative anaerobes such as *Porphyromonas gingivalis*, which thrives in areas with reduced oxygen tension and rich sources of protein nutrition (Lamont *et al.,* 2002). The most common plaque related diseases include dental caries, gingivitis and periodontitis. Dental caries is one of the most common childhood diseases and is generally the result of poor oral hygiene combined with a diet rich in sucrose. A sugar rich diet leads to acidification of dental plaque due to conversion of sucrose to lactic acid by oral streptococci and these acids can rapidly demineralise the dental enamel. These carious lesions are often enriched in acid resistant species such as *S. mutants* and *S. sobrinus* which produce copious biofilm and acid in sucrose rich environments.

Inflammation of the gingiva is a very common condition in individuals with poor oral hygiene and is the result of irritation of the gingiva by accumulated plaque. If poor oral hygiene continues into middle age and old age, accumulated plaque often extends subgingivally and can lead to inflammation of the periodontal tissues (periodontitis). These plaque communities are dominated by Gram-negative anaerobes which can induce a chronic inflammatory response that gradually destroys the periodontal tissues. Periodontal disease is the major cause of tooth loss in adults.

The oral cavity is also comprised of more than 100 fungal species (Ghannoum *et al.,* 2010). *Candida* spp., *Malasezzia* spp., and *Aspergillus* spp. are the most prevalent fungi in the oral cavity (Peters *et al.,* 2017). *Candida* spp. have been associated with periodontal disease and dental caries but their roles in these diseases are poorly understood (Canabarro *et al.,* 2013). Oral thrush is the most common oral fungal disease and is often a sign of immunosuppression due to HIV infection, old age or diabetes. Broad-spectrum antibiotics that disturb the bacterial microbiome can also lead to *Candida* overgrowth and oral thrush.

## 1.4 Oral cancer

Cancers are still one of the leading causes of death in the world. The incidence of various cancers differs throughout the world and these incidences are greatly affected by life-style factors such as a healthy diet, smoking and tobacco use, exercise and alcohol consumption.

Head and neck cancers are a group of malignancies that commonly arise on the mucosal surfaces of the upper aerodigestive tract. Head and neck cancers commonly account for approximately 3% and 4% of all cancers diagnosed in the UK and USA, respectively. Head and neck cancers can be classified into different types, including laryngeal cancer,

nasal cavity and paranasal sinus cancer, nasopharyngeal cancer, oral and oropharyngeal cancer and salivary gland cancer.

Approximately 90% of head and neck cancers are squamous cell carcinomas (SCCs) affecting the mucosal surfaces. Strong evidence has been presented that Human Papilloma Viruses (HPV) infection is a risk factor for squamous cell carcinoma of the head and neck, specifically nasopharyngeal carcinomas (Mork *et al.,* 2001). However, the role of HPV in oral SCCs (OSCC) is still a matter of debate as the prevalence of HPV DNA in OSCCs is lower than that identified in nasopharyngeal carcinomas. Nasopharyngeal cancer is a common cancer which affects the head and neck areas. The incidence of nasopharyngeal carcinoma world-wide is 1.2 per 100,000, (https://www.wcrj.net/article/1046). This incidence of OSCC can be different between countries, with a higher prevalence in China, Indonesia, Vietnam, India and Malaysia.

In general, the oral cancer rate is higher in developed countries compared to developing countries. However, the highest rates of oral cancer are found in some Asian countries such as India, Pakistan and Sri Lanka, where oral cancers contribute approximately 25% of all cancers. Approximately 90% of these cases can be attributed to tobacco consumption (Gupta *et al.,* 2016). Within Europe, oral cancer rates differ from country to country. France, Slovakia, Slovenia, the Russian Federation and Hungary have reported high incidences in men whereas countries such as Iceland and Cyprus have some of the lowest rates (Warnakulasuriya, 2009). In 2012, the total number of cases in European countries was 99,630 new cases of oral and pharyngeal cancers, comprising of 73,860 and 25,770 in men and women respectively (Ferlay *et al.,* 2013). In the USA oral cancers and oropharynx cancer constitute approximately 3% and 2% of all malignancies in men and women respectively (Neville and Day, 2002). In North America, oral cancer is considered to be the 11$^{th}$ most common cancer in the USA, while it is the 12$^{th}$ most

common in Canada and the 13th most common in Mexico (International Agency for Research on Cancer; (http://globocan.iarc.fr/Pages/fact_sheets_population.aspx). Oral cancer is the 15th most common cancer in Africa, however, it was the 7th most common cancer in Central Africa in 2012 with an estimated number of 17,276 cases diagnosed (Gupta *et al.,* 2016). Oral cancer is generally higher in men compared to women and this is generally attributed to the higher rates of smoking, and alcohol consumption in men. Smoking and alcohol consumption are considered as the most common risk factors associated with oral cancer. Tobacco smoke contains almost 7000 chemicals, more than 60 of which are well known carcinogenic compounds such as carbon monoxide, ammonia and acetone. It was estimated that in Ireland in 2013, approximately 5,950 premature deaths were smoking related (https://www.cancer.ie). Oral cancer has begun to show an increasing rate in women and this is likely due to the changes in lifestyle patterns related to increasing rates of alcohol consumption and smoking in women. The risks associated with alcohol consumption and smoking may be greater in women. A study by Muscat *et al.* (1996) compared men and women who had a history of smoking and alcohol intake and showed there was increasing risk of developing oral cancer in women who smoked and drank alcohol compared to men, also in older women, some changes such as hormonal changes and nutritional deficiencies in iron can increase the risk of oral cancer. However, oral cancer has risen globally in recent decades in all age groups, especially in the younger age groups (Brown *et al.,* 2012).

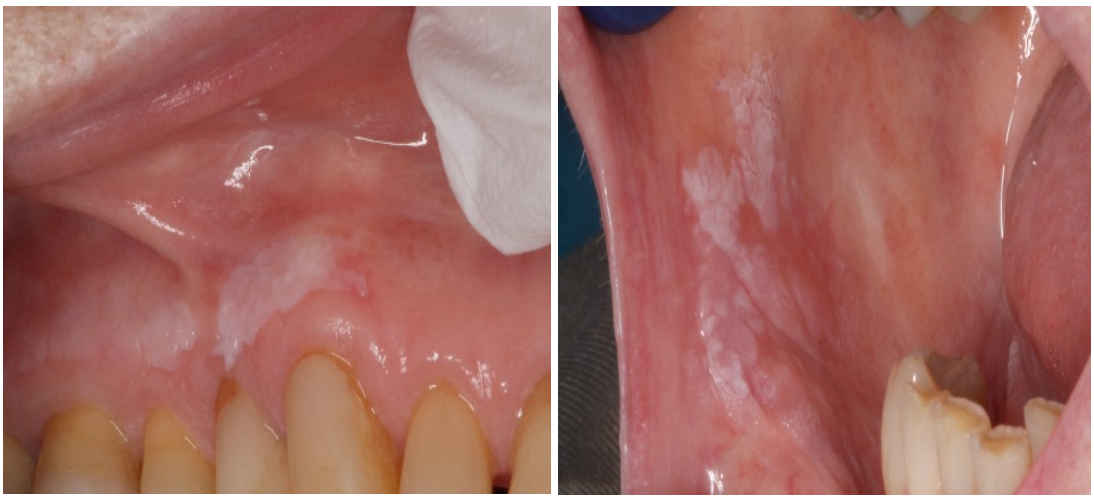
## 1.5 Oral leukoplakia

### 1. 5.1 Oral leukoplakia: definition and clinical features

Oral leukoplakia (OLK) refers to a white patch on the oral mucosal surface. Leukoplakia is derived from Greek meaning white (leuko) patch (plakos). In 1877, Erno Schwimmer was the first person who described the condition of oral leukoplakia (Starzynska *et al.,* 2015). Leukoplakia normally presents on the oral mucosa as a thick white layer which cannot be removed by brushing. The clinically accepted definition of OLK has been revised on several occasions. The World Health Organization originally defined OLK as a white patch or plaque which cannot be characterised pathologically or clinically. Subsequently, the first international conference on OLK in Malmo, Sweden in 1978 this definition was modified to include white patches that were not associated with any physical or chemical agents, with the exception of tobacco (Balasundaram *et al*., 2014). OLK has also been defined as an area of high malignant potential (Starzynska *et al.,* 2015).

The most common area of the oral cavity affected by OLK is the buccal mucosa; however, all mucosal areas of the oral cavity can be affected including the gums, lips, floor of mouth, tongue and palate. Generally, OLK can be classified into different subtypes, leukoplakia could be classified as simplex (homogenous), verrucosae (raised, red or white patches), and erosiva leukoplakia (with extensive proliferation above the mucosa). According to the WHO (Tandon *et al.,* 2017), OLK could also be classified into homogeneous and non-homogeneous types. The homogeneous type has a smooth surface with a uniform colour and is more commonly found in men and appears on the buccal mucosa (Fig. 1.2a). The non-homogeneous type can be classified as predominantly white or red-white (Erythroleukoplakia) and the surface is usually nodular and xerophytic (Fig. 1.2b).

Oral leukoplakia itself is not a serious condition; however, in some cases it is considered to be an initial stage in the development of oral squamous cell carcinoma (OSCC). OLK is estimated to occur in about 0.7% to 4% of the population. The prevalence of the disease is significantly affect by age, whereas 1% of men under 30 years old can be affected, while this percentage can increase to 8% in those aged 70 years or older. In addition, the malignant transformation rate of OLK varies between 0.7% and 2.9% of cases (Feller and Lemmer, 2012). However, this percentage can differ in relation to gender, life style factors (smoking and alcohol consumption) and is also higher in developing countries compared to developed countries. Although smoking is a risk for the development of OLK, non-smokers can still progress to OSCC. However, the aetiology of OSCC in those patients with a non-smoking history is still unknown (Kruse *et al.,* 2010).



(a) b

Figure 1.3a Photographs showing typical homogenous leukoplakia on the gingival mucosa and (b) Showing non-homogenous leukoplakia on the buccal mucosa (photographs courtesy of Dr. Claire M. Healy, DDUH, Trinity College Dublin).

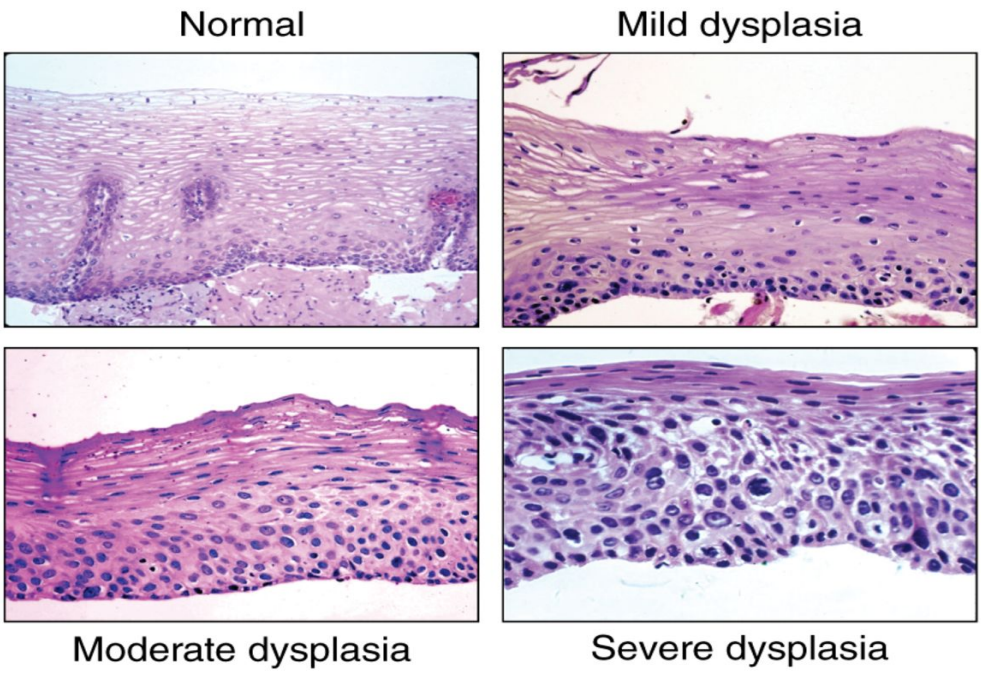



Figure 1.4 Microscopic appearance of tissues exhibiting different degrees of dysplasia compared to normal epithelial cells. Figure reproduced with permission from Taylor *et al*. (2013).

### 1.5.2 Diagnosis of leukoplakia

Proper diagnosis and assessment of the degree of dysplasia generally requires an excisional biopsy of the lesion (Pindborg *et al.,* 1980). Biopsy and assessment by microscopic analysis by an experienced pathologist is still the 'gold standard' for assessment of the degree of dysplasia. Leukoplakia is commonly classified as exhibiting three levels of dysplasia; mild, moderate and severe (Fig. 1.3). There are no treatments available for OLK and observation of the disease for changes in appearance is required for early detection of malignant transformation. Early detection of malignant transformation is essential for good patient outcomes. If OSCC is detected at an early stage, survival rates will approach 80%, whereas cancers detected in the later stages (T3 -T4), have survival rates of 20 to 30% (Sinha *et al*., 2014).

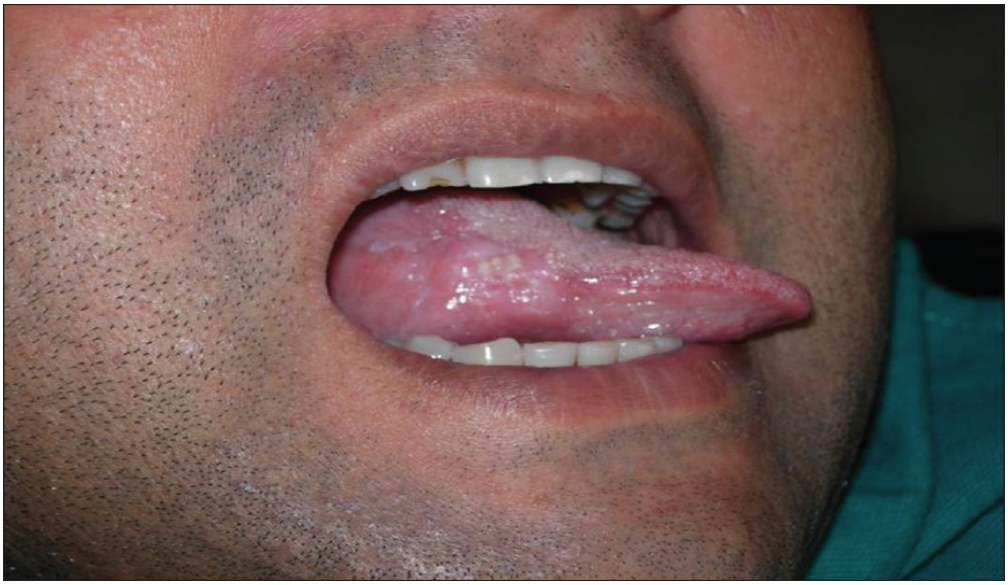


Figure 1.5 Shows a photograph of oral squamous cell carcinoma on the right border of the tongue, represented from (Vaidya *et al.,* 2013). Open access doi: 10.4103/0976-433X.116830

### 1.5.3 Human cancers and microorganisms

There are an increasing number of studies investigating the relationship between cancers and bacteria or other microorganism. The relationship between microorganisms and cancers was reinforced following the recognition in 1994 of *Helicobacter pylori* by the WHO as a bacterium that can cause human cancer. *Helicobacter pylori* is commonly linked with human gastric cancer and is associated with 60% to 90% of all gastric cancers (Pushalkar *et al.,* 2011). The bacterium *Salmonella typhi* is also associated with human cancer of the gall bladder and the bacteria *Streptococcus bovis* and *Chlamydia pneumoniae* are associated with colon cancer and lung cancer, respectively. Stronger associations exist between viruses and human cancers including Epstein Barr viruses (EBV) (Burkitt's lymphoma), Hepatitis B virus (liver cancer) and Human Papilloma

Virus (HPV) (oropharyngeal and cervical cancer). It is estimated that around 20% of all fatal cancers are caused by microorganisms (Pushalkar *et al.,* 2011).

### 1.5.4 Bacteria and human cancers

In recent years there have been several studies of the microbiome associated with human cancers of the GI tract. These studies have highlighted a possible role for Fusobacterium spp., in malignant transformation. *Fusobacterium* spp., can be opportunistic pathogens associated with human disease, for example cerebral abscesses or a cause of gastrointestinal symptoms such as heartburn. *Fusobacterium nucleatum* is a Gram-negative, anaerobic, fusiform, classified into 5 sub-species (*polymorphum*, *nucleatum, animalis*, *fusiforme* and *vincentii* (Nie *et al.,* 2015). *Fusobacterium nucleatum* is commonly found in dental plaque and is associated with common oral infections including chronic periodontitis, gingivitis and oral abscesses. In 2012, two concurrently published studies identified high levels of *F. nucleatum* on colorectal carcinomas, (Castellarin *et al.,* 2012; Kostic *et al.,* 2012). In a comparison between colorectal carcinoma samples and normal colonic samples using RNA sequencing by Castellarin *et al.* (2012) found that *F. nucleatum* sequences were higher in tumour samples relative to control samples. Kostic *et al.* (2012) also found that *Fusobacterium* spp., were more abundant in samples from colorectal carcinomas compared to normal tissues using qPCR and 16S rRNA sequencing.

Subsequent studies have also shown that *Fusobacterium* spp., are found at higher levels oesophageal cancers compared to normal oesophageal mucosa by real-time PCR *F. nucleatum* was also more commonly associated with recurrent colon cancers and resistance to chemotherapy by a mechanism that involved activation of autophagy (Yamamura *et al.,* 2016; Bullman *et al.,* 2017). In the same study *F. nucleatum* was

detected in colon cancers and metastatic lesions. Following xenograft transfer of these tumours to mice, *F. nucleatum* was also transferred. However, treatment of the mice with the antibiotic metronidazole resulted in reduced *Fusobacterium* load, cancer cell proliferation and tumour growth (Bullman *et al.,* 2017).

A study by Mehta *et al.* (2017), found that the level of *F. nucleatum* associated with colorectal cancer can be influenced by diet, with high-fibre diets associated with a lower risk of *F. nucleatum* colonisation.

### 1.5.5 Microbes associated with oral squamous cell carcinoma (OSCC)

The association of the microbiome and oral cancers has also been under intense investigation in recent years. Studies of OSCC have identified enrichments of many oral bacteria in the saliva of cancer patients, however the results of these studies have been inconsistent indicating that saliva may not be a useful diagnostic for OSCC. A study by Mager, (2005) identified six common bacterial species at high levels in saliva in OSCC using DNA-DNA checkerboard, including *Capnocytophaga ochracea, prevotella melaninigenica, Eubacterium Saburreum, Leptotrichia buccalis, C. gingivalis,* and *S. mitis.* Lower levels of these bacteria were identified in control samples that were OSCC-free, and this study concluded that OSCC had more effect than smoking and periodontal infections on the salivary microbiome (Mager *et al.,* 2005). Pushalkar *et al.* (2011), sequenced the V4-V5 region of the salivary microbiota in 3 patients with OSCC and found that the most prevalent genera in OSCC were *Gemella, Peptostreptococcus, Rothia, Lactobacillus,* and *Porphyromonas.* However, in the control samples, the genera *Neisseria, Gemella, Leptotrichia*, *Actinobacillus* and *Capnocytophaga,* were present at high levels (Pushalkar *et al.,* 2011).

Direct examination of the mucosal surface of tissue biopsy has produced more consistent results. Nagy *et al.* (1998) carried out direct culture of OSCC lesions from 21 patients and identified high levels of *Porphyromonas* spp. *Prevotella* spp., and *Fusobacterium* spp. Schmidt *et al.* (2014) analysed the microbiome of OSCC in 27 patients compared to 17 patients with pre-cancer and 18 healthy control patients by sequencing the V4 region of bacterial DNA recovered from swabs, this study showed that OSCC patients exhibited reduced levels of *Streptococcus* spp., and *Rothia* spp., and exhibited higher levels of *Bacteroidetes* and *Fusobacterium* spp. This study also showed that *Fusobacterium* spp. levels were enriched on pre-cancerous lesions compared to healthy mucosa. More recently Al-Hebshi *et al.* (2017), investigated the microbiome of biopsy tissue including 20 biopsies of OSCCs and 20 healthy epithelium controls by sequence analysis of the V1-V3 regions of the 16S rRNA, this study showed that *F. nucleatum* subsp. *polymorphum* was one of the most significantly enriched species in the tumour samples, as well *P. aeruginosa* and *Campylobacter* spp., control samples in contrast exhibited higher levels of *S. mitis*, *R. mucilaginosa* and *H. parainfluenzae*.

Investigation of the oral microbiome in 197 OSCC patients exhibiting different tumour stages (I to IV) and 51 healthy individuals showed that the abundance of *Fusobacterium* spp., increased significantly with the progression of the oral cancer, from the healthy controls where this species accounted for 2.98% of the microbiome, to 4.35% in stage I OSCC and 7.92% in stage IV of OSCC (Yang *et al.,* 2018).

### 1.5.6 Functional analysis of the OSCC microbiome

Although studies of the microbiome composition can be informative (Al-Hebshi *et al.,* 2019) studies of the functional capacity of the microbiome may be more important. Analysis of the oral metagenome in OSCC using functional prediction tools shows that

genes encoding functions in flagellar assembly, lipopolysaccharide (LPS), biosynthesis and bacterial chemotaxis were higher in tumours samples compared to healthy controls (Al-Hebshi *et al.,* 2017). However, in control samples genes responsible for the DNA repair and recombination were more common. In addition, genes encoding components of bacterial flagella and lipopolysaccharide (LPS) may have a role in inducing cancer-promoting inflammatory reactions by activation of the TLR/MyD88/NF-NF-κB pathway (Ikebe *et a*l., 2009).

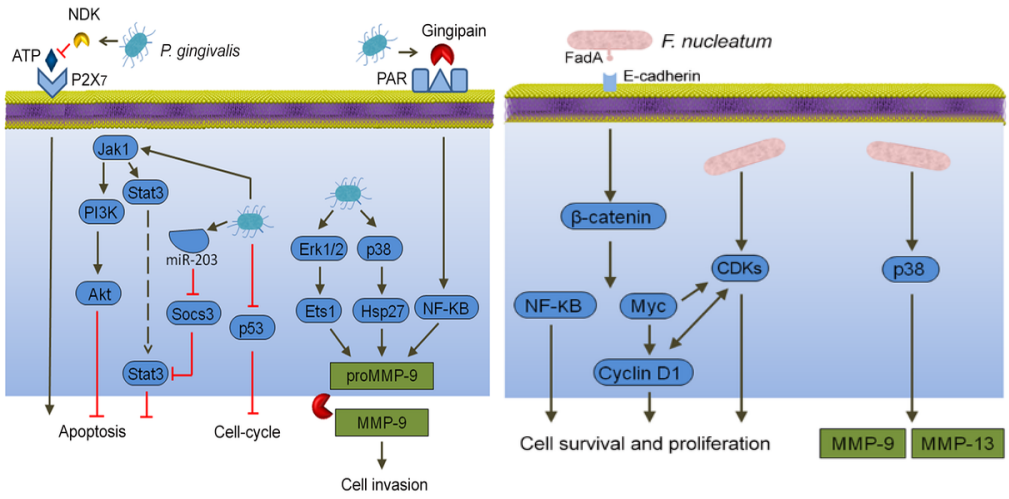
### 1.5.7 Interactions of oral microbes and host cells

The oral cavity microbes have begun to receive more attention in order to understand the relation between oral microbes and oral cancers. The oral cavity is comprised of hundreds of species (Aas *et a*l., 2005). However, microbes colonisation have been linked to the transformation of premalignant lesions into OSCC (Perera *et al*., 2016). Changes in the oral epithelial tissue may be linked with changes in bacterial communities on the surface, by selecting and enriching for species able to stick to and colonise to the abnormal epithelial tissue. The surface glycoconjugate layer acts as a receptor for specific species and determines which species are able to adhere to this layer and colonise; for example, *S. salivarius* can adhere to and colonise the oral mucosa; whereas *S. gordonii* is found less commonly on mucosal surfaces and has a preference to adhere and grow on the tooth surface (Kolenbrander *et al*., 2002).

Changes to the mucosal microbiome that can result in the acquisition of a more invasive microbiota may have undesired effects on the host. Infection with invasive microorganisms could affect the cell cycle by disrupting host signalling factors or by inducing uncontrolled cell growth by inhibiting apoptosis (Whitmore *et al.,* 2014). Virus infection is often associated with effects on tumour suppressor genes such as BCL (B-cell

lymphoma protein) and pRB which are major anti-apoptotic proteins, resulting in uncontrolled cell growth. With regards to oral bacteria, the effects of *F. nucleatum* and *Prohormones gingival is* on host cells are probably the most intensively studied (Fig. 1.5). This is due to the ability of *P. gingivalis* to invade gingival cells and due to the association of *F. nucleatum* with colorectal cancer. Interestingly, both *Fusobacterium* and *Porphyromonas* spp*.,* have been recovered at higher rates from OSCC compared to healthy mucosal tissue (Nagy *et al.,* 1998).

The bacterium *P. gingivalis* can affect cells in two different ways. During cell division, *P. gingivalis* can produce a nucleoside diphosphate kinase (NDK) which inhibits the ATP (adenosine triphosphate ), activation of the epithelial cell receptor p2X7 which results in inhibition of apoptosis (Fig. 1.5a ). Also *Porphyromonas gingivalis* can also affect the levels of the p53 tumour suppressor, thereby increasing the rate of progression of the cell cycle. On dendritic cells, NDK activation of p2X7 receptor can also stop the activation of the inflammasome, resulting in reduced IL-1ß which is important for activating anti-tumour $CD8^+$ T-cells. *F. nucleatum* has also been studied intensely to explain its association with colorectal cancer (CRC) (Whitmore and Lamont, 2014). The *F. nucleatum* adhesion FadA binds to E-cadherin on colon cancer cells and activates ß-catenin signalling that results in activation of the Cyclin D1 protein, increasing the rate of progression of the cell cycle (Fig. 1.5b). *F. nucleatum* is also associated with activation of p38 and production of matrix-metalloproteinases (MMP-9 and MMP-13) which play an important role in tumour invasion. Whitmore and Lamont, 2014).




(a) (b)

Figures 1.6 Schematic showing interactions between (a) *P. gingivalis* (b) *F. nucleatum* and epithelial cells that could promote oncogenic phenotypes. Figures from Whitmore and Lamont (2014). Open access, doi:10.1371/journal.ppat.1003933

### 1.5.8 Products of microbial metabolism and OSCC

In addition to invading host cells, the microbiome may produce metabolites that could negatively affect host cells. Acetaldehyde is generated by the oxidation of ethanol and can be produced by different oral microbes (Fig. 1.6). Acetaldehyde is toxic and mutagenic and can induce point mutations in DNA and promote the formation of DNA adducts resulting in chromosomal aberrations. Saliva does not normally contain any acetaldehyde; however, carcinogenic levels can be found after alcohol consumption (Moritani *et al.,* 2015).

An investigation of some oral microbes and the ability to produce acetaldehyde *in vitro*, found that *Neisseria* spp., had the greatest capacity to generate toxic levels of acetaldehyde also found that *Neisseria* had a high capacity to generate acetaldehyde *in vitro. Candida albicans* has also been shown to be capable of generating toxic levels (>100 μM) of acetaldehyde. A recent study has examined the salivary microbiome in individuals with a high and low capacity to generate acetaldehyde *in vitro,* and showed

that *R. mucilaginosa* was more abundant in the saliva of acetaldehyde producers. Whereas, unexpectedly, *Neisseria* spp., were negatively associated with acetaldehyde production (Yokoyama *et al.,* 2018).

Another group of toxic metabolites is produced by nitrosation, which involves the conversion of nitrites and amides to N-nitroso compounds (NOC), which are known carcinogens (Jones *et al.,* 2016). Nitration activity has been shown in many gut bacteria including *E. coli* and *Clostridium spp.* and in the oral yeast *C. albicans*.

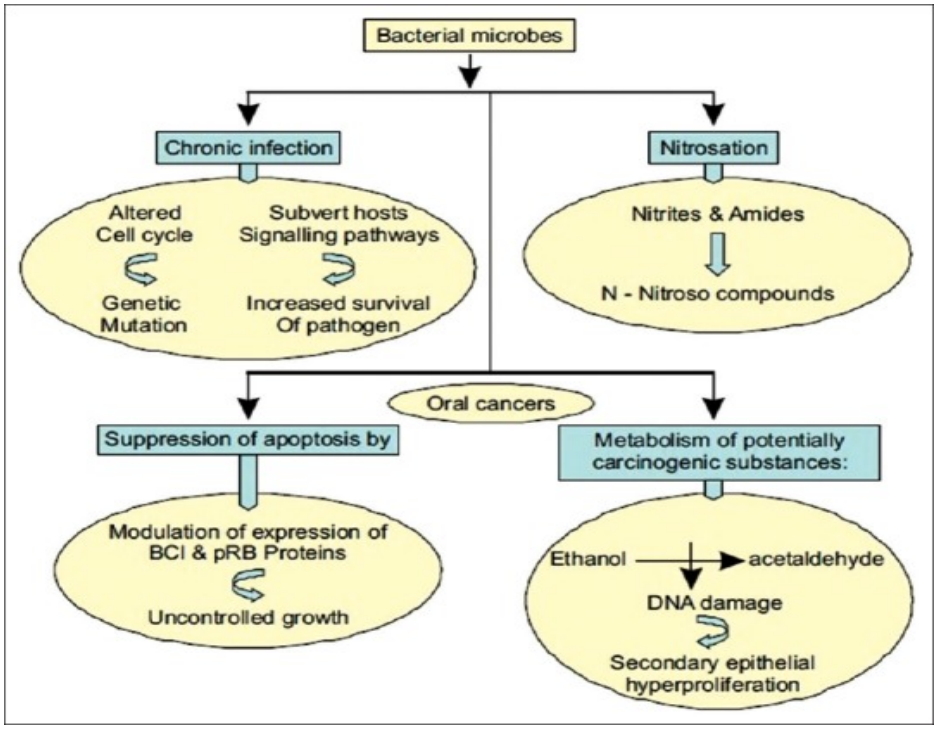



Figure 1.7 Schematic showing potential mechanisms by which bacteria could induce carcinogenesis. Figure from Srinivasprasad and Dineshshankar. (2015). Open access, doi: 10.4103/0975-7406.163451

## 1.6 Fungi and OSCC

*Candida* spp. are commonly found in the oral cavity; however, the most common species is *C. albicans*, which can be recovered from about 50% of healthy individuals, (Mobeen, 2014). *Candida* spp., are considered to be parasitic organisms and an etiological agent of OLK. Some *Candida* infections in the oral cavity are termed as candidal leukoplakia, which is characterised by hyphal invasion of the superficial layers of the mucosa (Pushalkar *et al.,* 2011; Abdulrahim *et al.,* 2013).

A study by Mohd Bakri *et al.* (2010), showed that *C. albicans* colonised pre-cancerous lesions and suggested a role in the development of OSCC. However, the relationship between *Candida* spp.*,* OLK and OSCC is not fully understood, (Vuckovic *et al.,* 2004). Some studies have indicated that the presence of *Candida* spp., on OLK is associated with the degree of the dysplasia (McCullough *et al.,* 2002) .While *C. albicans* has the ability to colonise leukoplakia tissues and causes serious damage of tissues, its contribution to malignant transformation is not known (Sankari*,* 2015).

An examination of potentially malignant and malignant oral disorders for evidence of fungal growth, reported that 8 /15 patients with potentially malignant lesions were colonised by *C. albicans* and that *C. albicans* could be recovered from 20 of 30 patients with OSCC (Saigal *et al.,* 2011). A study by Sarkar and Rathod (2014) examined 40 patients with OLK and found approximately 50% were positive for *Candida* spp., in contrast only 3/21 healthy controls were *Candida*-positive.

## 1.7 Human cancer and human papilloma viruses (HPV)

Human papilloma viruses (HPVs), Hepatitis B virus (HBV) and Hepatitis C virus (HCV) are viruses that are commonly linked with human cancers (Ha *et al.,* 2019). Studies have been carried out to understand the involvement of HPV in OSCC (Srinivasprasad *et al.,*

2015). HPV has been identified as an important cause of cancer of the cervix, but the role of HPV in cancers of the head and neck is less clear (Curry *et al.,* 2018). HPV has been associated with cancers of the oropharynx and tonsils and HPV-16 and -18 tend to be the most common viruses associated with cancers, (Herrera and Goepfert, 2009) . Studies of the prevalence of these viral subtypes (HPV-16 and -18) in OSCC have produced conflicting results. (Gupta, 2015) found that 22% of 167 OLKs analysed were positive for HPV. However, the presence of the virus was not significantly associated with malignant transformation. In contrast Bouda *et al.* (2000) reported that over 90% of 53 oral dysplastic and cancerous tissues were HPV DNA positive with 0% positivity in healthy control tissues. Although a role for HPV in OSCC has not been definitively proven, the involvement of the HPV virus in a proportion of OSCCs is certainly possible. In addition, the role of different HPV subtypes or novel viral pathogens cannot be discounted and a thorough analysis of the virome in malignant and premalignant tissues is required.

## 1.8 Bacterial identification

The oral microbiome has become an area of intense investigation due to its role in common oral diseases, such as periodontal disease and dental caries. As a result of this, approximately 54% of the oral microbiota has been characterised, while 14% of oral bacterial species remain unnamed but have been cultivated. The remaining 32% are uncultivated species. Information on the content of the oral microbiome is curated in the Human Oral Microbiome Database (HOMD), most of these uncultivated species have been identified based on the 16S rRNA sequence information. The Human Oral Microbiome Database (HOMD) provides scientific community with a database of

sequenced and identified oral microorganisms, with a catalogue of more than 700 species found in the oral cavity, including un-cultivated species.

### 1.8.1 Culture-dependent techniques

Identification of bacteria to the species or the subspecies level is an important process in medical microbiology, and can guide the treatment of infection and facilitate epidemiological studies. Microbial culture is still the primary method identification. Most common oral bacteria can be cultured on blood agar either aerobically or anaerobically under standard laboratory conditions; however, some species have specific growth requirements. Morphological features and metabolic characteristics are commonly used to identify bacteria; however, some different species share the same or similar features and these characteristics can only provide a presumptive identification. For this reason, molecular identification has begun to replace culture-based methods for microbes identification.

## 1.9 Molecular Technology

### 1.9.1 PCR and DNA sequencing

Traditional methods of bacterial identification can take several days and are often inaccurate and cannot be used to identify bacteria that cannot be cultured *in vitro*. However, polymerase chain reaction (PCR), is now routinely used to identify and quantify bacterial species, using End-point PCR and Real-Time PCR. PCR is quicker method than the culture and can also offering more accurate classification, however, controls to avoid false positive results are very important. A comparison between bacteria culture and Real-Time PCR for identification of the oral bacterium *Porphyromonas gingivalis* in 259 adult patients with severe periodontitis, showed that *P. gingivalis* was

observed in 111 samples (43%) by culture; however, 138 samples (53%) were shown to be positive for *P. gingivalis* by PCR (Khalil and Boutaga, 2003). PCR is not only more accurate and more sensitive, but it is also suitable for early diagnostic examination with bacterial infection, when bacterial load may be very low.

Recently, DNA sequencing has become the method of choice for bacterial detection and identification and is widely used in clinical laboratories. Databases of DNA sequences have become more widely available to assist bacterial detection and identification, including cultivated and un-cultivated species. The 16S rRNA gene contains many conserved regions that can be used to design pan-bacterial primers. These primers can be used to amplify internal variable portions (V1-V9) that contain sufficient sequence diversity to allow bacterial classification up to species and sub-species levels (Fig. 1.8).

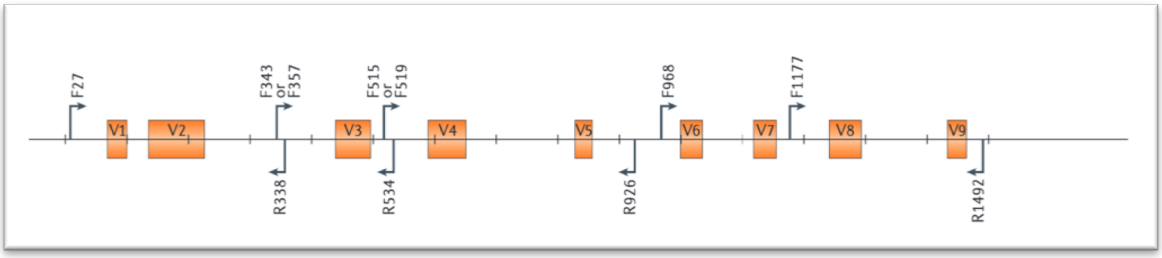



Figure 1.8 Shows the position of sequencing targeted by common pan-bacterial primers are indicated by arrows (Kuczynski *et al.,* 2011). Open access, doi: 10.1038/nrg3129

### 1.9.2 Next generation sequencing

The first next-generation sequencing (NGS) sequencing platform was launched in 2005, and since then NGS technologies, has developed dramatically and revolutionized the field of genomics. There are currently several different NGS technologies in common use, produced by different companies including 454 pyrosequencing developed by Roche Applied Science, (Basel, Switzerland), Illumina/Solexa technology developed by Illumina (San Diego, CA, USA), SOLiD technology, developed by Applied Biosystems (Foster City, CA, USA), Single Molecule Real Time technology (SMRT), developed by Pacific Biosciences (Menlo Park, CA, USA).

The development of these technologies and the relatively cheap cost of sequencing reagents has led to an explosion in studies of the human microbiome. Illumina sequencing technology is a commonly used technology for bacterial identification, which allows for parallel sequencing of the DNA molecules on a massive scale. This facilitates sequencing of short fragments of the DNA (100 -300 bp) in complex samples and applied to DNA from mixed populations of bacteria, it can allow identification and quantification of individual bacteria in a whole community population. For such studies, the 16S rRNA gene is the target molecule of the choice.

The Illumina MiSeq platform is commonly used in these studies. Sequencing on the Illumina MiSeq consists of 4 basic steps: sample preparation, cluster generation, sequencing, and data analysis. Sample preparation involves the amplification of the 16S rRNA regions, in addition, Illumina primers are used to bind DNA products (amplicon PCR) to allow multiplexing of the reaction. Cluster generation occurs on the flow cell once samples have been loaded on the Illumine MiSeq. The Illumina adaptors facilitate hybridization of the samples to the flow-cell. A polymerase then creates a compliment of the hybrid fragment. This double stranded DNA molecule is denatured and the original

template is washed away. The remaining strand arches over to bind to the complimentary oligo nucleotide on the flow cell and then the polymerase generates a complementary strand, resulting in a double bridge strand (Fig. 1.8). This binding strand is denatured resulting in a single strand and the process of synthesis is then repeated over and over, resulting in millions of clusters. At this stage the reverse strands are washed away leaving only the forward strands and at this stage the sequencing by extension commences. Fluorescent nucleotides are added during synthesis of the complimentary strand and detection of these fluorescent signals is used to determine the sequence of the newly synthesised DNA strand.

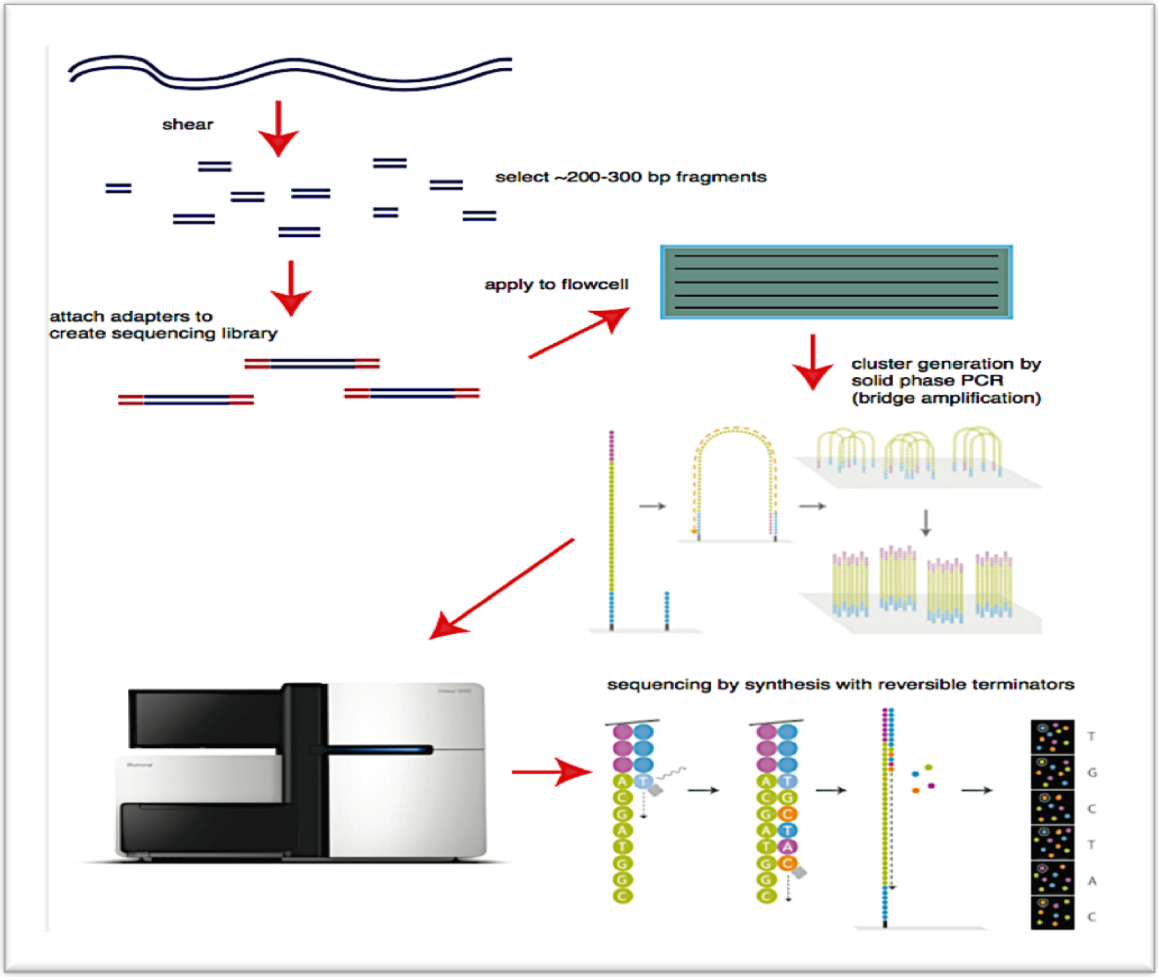



Figure 1.9 A diagram outlining the process of Illumina DNA sequencing (https://www.illumina.com/systems/sequencing-platforms/miseq.html)

## 1.10 Aims and objectives of the study

The aim of this study was to characterise bacterial communities colonising OLK compared to communities recovered from contralateral normal sites from the same patient. In addition, healthy control samples were included. We aimed to determine whether these patches harboured an altered microbiome that could contribute to the process of malignant transformation. Specific questions to be addressed include:

- Whether using parts of the 16S rRNA such as V1-V2 or V3-V4 regions are more appropriate for oral analysis.
- Whether the bacterial species identified in OSCC in previous studies are enriched in OLK.
- The influence of smoking, alcohol drinking and *Candida* spp. on the microbiome.
- Determine if potentially carcinogenic levels of acetaldehyde can be produced by the oral microbes associated with OLK patients in this study.

# Chapter 2

Materials and Methods

## 2.1 Reagents and equipment

### 2.1.1 Chemicals, buffers and molecular biology reagents

All the chemicals and buffers used in this study are listed in Table 2.1. Unless otherwise stated all solutions, buffers and all PCR reactions, all buffers were sterilised before use.

Table 2.1. List of the chemicals, buffers and reagents used in this study.

| **Reagent/Buffer** | **The function** |
|---|---|
| $dH_2O$ (RNase, DNase and protease- free; Sigma-Aldrich) | PCR reaction and buffer preparation |
| Breaking Buffer (2% Triton x-100, 1 mM EDTA, 100 mM NaCl and 10 mM Tris-HCl) | Used in DNA extractions to break down the cell membrane |
| 10x TBE buffer (1 M Tris, 1 M Borate, 0.02 M EDTA, pH 7.8) | Gel electrophoresis, used at 0.5x concentration |
| 10x Dulbecco's Phosphate Buffered Saline (DPBS; 1.47 mM $KH_2PO_4$, 2.67 mM KCl, 1.37 mM NaCl, 8.1 mM $Na_2HPO_4$) | Washing buffer (cell line), used at 1x |
| Acetaldehyde (ACS reagent, ≥ 99.5%) | Used to induce cell stress |
| Dimethyl Sulfoxide (DMSO) | Cell culture storage media |
| Hydrogen peroxide solution (30% w/v in $H_2O$) | Used as positive control for cell stress |
| Dulbecco's modified Eagle Medium (DMEM) with 4.5g/l glucose, L-glutamine and sodium bicarbonate. | Base cell culture media |
| Complete DMEM (above with 10% v/v Foetal Calf serum, 100 100 U penicillin, 0.1 mg/ml streptomycin. | Complete cell culture media |
| Brain Heart Infusion Media (Bacto) | Used as liquid and solid bacterial growth media. |

## 2.1.2 General plastics and laboratory equipment

Commonly used laboratory equipment used during the study listed in (Table 2.2). General plastics (Table 2.3).

**Table 2.2** List of the equipment used during the study.

| **Equipment** | **Manufacturer** | **Model design** | **Rotor number** |
|---|---|---|---|
| Micro centrifuge | Eppendorf (Hamburg, Germany) | 5810R | FA45-30-11 |
| Large centrifuge | Hittich (Tuttlingen, Germany) | Universal 320 | 5810R |
| Real-Time thermocycler | Applied Biosystems (CA, USA) | 7500 Fast Real-Time PCR Applied Biosystems | -- |
| Spectrophotometer | Thermo Scientific (MA, USA) | NanoDrop 200c | -- |
| Spectrophotometer | Thermo Scientific (MA, USA) | Evolution 60s | -- |
| Water bath | Nickel-Electro LTD, (Weston Super-Mare, UK) | Clifton NE5-28 | -- |
| Water bath (cell culture room) | Nickel-Electro LTD (Weston Super-Mare, UK) | Clifton NE4-p | -- |
| Cell disrupter | Biospectrometer (Bartlsville, OK, USA) | Mini-bead-beater (24 sample) | -- |

Table 2.3 List of the disposable plastic ware used during the study.

| Plastic ware | Size/ volume | Manufacturer |
|---|---|---|
| Microcentrifuge tube | 1.5 ml | Eppendorf (Hamburg Germany) |
| Micro tube (screw cap) | 2 ml | Starsted (Hildesheim Germany) |
| PCR tube | 0.2 ml | Molecular bioproducts (CA, USA) |
| Conical centrifuge tubes | 50 ml | Greiner bio-one (Kremsmunster, Austria) |
| Cell culture flask | 75 $cm^2$ | Corning (Corning, NY, USA) |
| 96 well (cell culture) | 0.1 ml | Greiner bio-one (Kremsmunster, Austria) |
| 96 well plate (PCR) | 0.1ml | Applied biosystems (Carlsbad, CA, USA) |

## 2.2 Oral sample collection

A total of 112 samples were collected at The Dublin Dental University Hospital (DDUH), Trinity College Dublin (TCD), using the Catch-all swab kit (Epicentre, Madison, WI). 78 samples were collected from patients attending the Oral Mucosal Dysplasia Clinic (Appendix I). Patients were excluded if they had cancer, Crohn's disease or diabetes or had received antibiotics or steroids in the preceding three months. Eligible patients were swabbed at mucosal surface to sample the site of OLK, also at contralateral normal sites where available. In detail, 23 Samples were recovered from buccal leukoplakia, with 18 corresponding buccal contralateral normal, while, 27 from lingual leukoplakia and 10 recovered from lingual contralateral normal site. As controls, 23 samples were recovered from buccal sites in healthy individuals and 11 samples from lingual sites in healthy individuals. At the time of sampling, information about the site of sample collection was recorded (leukoplakia or contralateral), along with information on denture wearing, smoking frequency, alcohol consumption and mouthwash use. The age of the patients

included ranged from 40 to 70 years old, average age was 55 years old, healthy control samples ranged in age from 30 to 70 years old (average 45.53 years old).

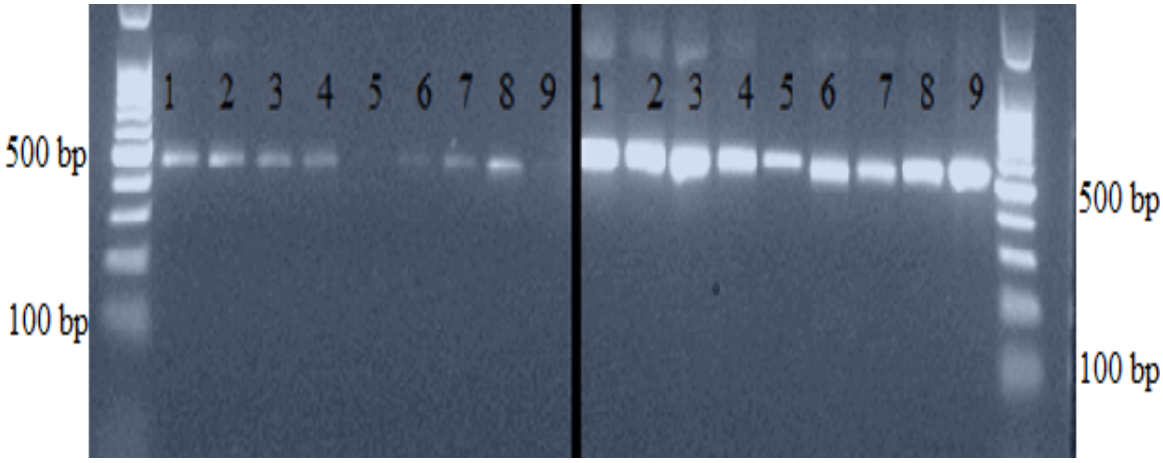
## 2.3 DNA techniques

### 2.3.1 DNA extraction and DNA purification

Using the Catch-all collection swab kit ( Epicentre, Madison, WI, United States) to collect sample from each patient, DNA extractions were carried out using the MasterPure DNA Purification Kit (Epicentre, Madison, WI). Ready-lyse lysozyme (Epicentre, Madison, WI) was diluted to 250 U/µl in TES buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, and 100 mM NaCl). Each sample was mixed with 300 µl of the TES buffer in a screw-top tube and 2 µl of the diluted Ready-lyse lysozyme (Epicentre, Madison, WI, United States) was added and samples were incubated at room temperature for 15 min with occasional mixing. Proteinase K (included in the kit) was diluted in 2x T and C lysis buffer (1 µl of proteinase K to 150 µl of buffer) and 150 µl of this solution was added to the tube and then incubated at 65℃ for 15 min. Next, 0.25 ml glass beads (sterilised in an autoclave for 15 min at 120˚C) was added to the tube and the sample was disrupted in a Mini bead beater (Biospec Products, Bartlesville, OK, United States) for 30 s. The top layer was transferred to a new 1.5 ml tube and 1 µl of RNase was added and incubated for 30 min at 37˚C. Samples were placed on ice and then 175 µl of the MPC protein precipitation buffer (included in the kit) was added and vortexed and centrifuged for 10 min at 4˚C and 18,000 x *g*. The supernatant was removed and placed in a 1.5 ml Eppendorf tube and mixed with 500 µl isopropanol, gently mixed and centrifuged for 10 min at 4˚C at 18,000 x *g*. The liquid was then removed and the pellet dried and suspended in 35 µl of TE buffer. DNA concentration was measured by using a Nano Drop 2000 C spectrophotometer.

### 2.3.2 PCR amplification

The V1-V2 regions of the 16S rRNA gene were amplified using the primers 27-FYM and 338R-R and the V3-V4 regions were amplified using primers V3-V4-F and V3-V4-R (Table 2.4). These primers contained regions homologous to the relevant variable regions of the 16S rRNA gene and sequences to facilitate the addition of Illumina index primers for Illumina sequencing. KAPA HiFi Hot start Master Mix (Roche Diagnostics Ltd., West Sussex, UK) was used in all amplifications. A 2.5 µl volume of microbial genomic DNA (5 ng/µl), 5µl of forward and reverse primers (10 mm) and 12.5 µl of 2x KAPA HiFi Hot start ready mix were mixed together in a total final volume of 25 µl. PCR reactions were pre-heated at 95°C for 5 min, followed by PCR conditions consisting of 30 cycles of denaturation at 95°C for 45s, annealing at 60°C for 45 s, and elongation at 72° for 90 min, followed by a final elongation step at 72°C for 7 min, followed by a holding step at 4°C. In order to visualise the PCR products, 2 µl from the reaction was loaded on a 1% agarose and visualised by UV transilluminator (Agilent DNA 1000 kit catalogue number 5067-1504) (Fig. 2.1).




(a) (b)

Figure 2.1 Shows a photograph of a 1% agarose gel containing (a) PCR amplimers generated using primers 27F and 338R from template DNA of 9 OLK (b) index PCR generated with Illumina indexing primers (FC-131-1001 or FC-131-1002), using the samples shown in panel (a) as template DNA. Outermost lanes contains DNA ladders 100bp ladder Promega )

### 2.3.3 PCR clean-up

Clean-up of the PCR products was performed in a 96-well plate format using magnetic beads. A 20 µl volume of AMPure XP beads (Beckman Coulter, Brea, CA, USA) was added to each sample and mixed. 52.5 µl volume of 10 mM Tris pH 8.5 and 400 µl fresh 80% (v/v) ethanol were used to wash the bead-bound DNA and remove traces of buffer and salts. A magnetic rack (96 wells) was used to isolates the beads bound to the purified DNA, according to the manufacturer instructions.

Table 2.4 Showing the PCR primers used in this study.

| **Primer name** | **primer sequencing** | **Reference** |
|---|---|---|
| 27F | GAGTTTGATCMTGGCTCAG | (Frank *et al.,* 2008) |
| 1492R | CGGTTACCTTGTTACGACTT | (Frank *et al.,* 2008) |
| 27F-YM | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG AGTCAGTCTGTCAGAGTTTGATYMTGGCTCAG | (Frank *et al.,* 2008) |
| 338R-R | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACA GTATGGTAATTCATGCTGCCTCCCGTAGRAGT | (Frank *et al.,* 2008) |
| V3V4-F | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG CCTACGGGNGGCWGCAG | https://support.illumina.com/documents/documentation/chemistry |
| V3V4-R | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACA GGACTACHVGGGTATCTAATCC | https://support.illumina.com/documents/documentation/chemistry |
| QPCR-16S-F | TCCTACGGGAGGCAGCAGT | (Kraneveld *et al.,* 2012) |
| QPCR-16S-R | GGACTACCAGGGTATCTAA | (Kraneveld *et al.,* 2012) |
| ITS F | CCTGTTTGAGCGTCRTTT | (Kraneveld *et al.,* 2012) |
| ITS R | TTCTCCGCTTATTGATAT | (Kraneveld *et al.,* 2012) |

### 2.3.4 Second PCR (index PCR)

The purified V1-V2 and V3-V4 amplicons were used as templates for index PCR that was carried out using primers containing a unique 8 bp sequence tag. The indexing primers were included in the Illumina Nextera DNA library preparation kit and the unique 8 bp indices in each primer used are shown in Table 2.5. Each reaction contained, 5 µl of the amplicon PCR product, 25 µl of 2 x KAPA HiFi per sample, 10 µl of PCR-grade water and 5 µl from the indexes forward and reverse primers. Each sample was indexed with a unique combination of the index primers to allow the sequences to be identified according to each sample. Index PCR was carried out with an initial denaturation at 95˚C for 3 min, followed by eight cycles of denaturation at 95˚C for 30 s, 55˚C for 30 s and 72˚C for 30 s with a final extension for 5 min at 72˚C. After that PCR reactions were then held at 4˚C. Index PCR clean-up was also performed as described above and the products were visualised by agarose gel electrophoresis (Fig. 2.1).

**Table 2.5** Showing eight bp indices contained in the index primers.

| Primer name | Index Sequence | Primer name | Index Sequence |
|---|---|---|---|
| N701 | CGTACTAG | S501 | TAGATCGC |
| N702 | AGGCAGAA | S502 | CTCTCTAT |
| N703 | TATCCTCT | S503 | TATCCTCT |
| S504 | TCCTGAGC | S504 | AGAGTAGA |
| N705 | GGACTCCT | S505 | GTAAGGAG |
| N706 | TAGGCATG | S506 | ACTGCATA |
| N707 | CTCTCTAC | S507 | AAGGAGTA |
| N708 | CAGAGAGG | S508 | CTAAGCCT |
| N709 | GCTACGCT | | |
| N710 | CGAGGCTG | | |
| N711 | AAGAGGCA | | |
| N712 | GTAGAGGA | | |

## 2.3.5 Library quantification, normalisation, and pooling

All samples were quantified using the bioanalyser (Agilent Technologies) to verify the size and quantity of the indexed PCR products. The concentration of the index PCR products were adjusted to 4 nM, using 10 mM Tris pH 8.5, following this, pooling of the library was carried out by mixing 4.0 nM from each index PCR using the below format calculated.

$$\frac{(\text{concentration in ng/µl}) \times 10^6}{(660\text{g/mol} \times \text{average library size})} = \text{concentration in nM}$$

Each library was diluted to the final concentration of 4 nM, then 5 µl from each library was pooled in a 1.5 ml Eppendorf tube and combined with PhiX control DNA at 5% and loaded at a concentration of 6 pM. The library was denatured and then loaded into the MiSeq cartridge and analysed on the MiSeq machine according to the manufacturer's guidelines. Sequencing was carried out using the Illumina MiSeq 600-cycle kit (2 x 300 bp; Catalogue number MS-102-3003).

## 2.4 Mothur software analysis

In this study, the Mothur software package developed by Patrick Schloss was used for sequence analysis (Schloss *et al*., 2009). This software is freely available at https://www.mothur.org/. This study used version 1.35.1, the latest version at the time of analysis, was used to classify the sequences up to the genus and species levels. A total number of 19 commands were used to generate a unique sequence profile of the operational taxonomic units (OTUs) for each sample. Initially, the 'make contigs' command was used to align the over-lapping DNA sequence reads. Sequences were then aligned to the V1-V2 region using a dataset of sequences from the Human Oral Microbiome Database (HOMD). After alignment, over-hanging sequences were removed resulting in a gapped alignment. A pre-clustering command was used to merge sequences that differed by 1 base that are likely due to sequencing errors. Commands were removed next to remove singletons and to remove chimeras generated during PCR using UChime (Edgar *et al*., 2011).

The first step of the OTU assignment is to split the sequences into groups at the taxonomic level of order and cluster within the groups. The sequences were then classified in Mothur using the classify seqs command and the HOMD database. Any OTUs that were identified in subsequent statistical analysis as being significantly enriched were further classified

using BLAST (Altschul *et al.*, 1990), analysis of the 16S database at HOMD (http://www.homd.org) using the consensus sequence for that OTU. Sequences were also classified at the phyla level and species level for phylotype analysis.

### 2.4.1 Alpha Diversity

#### 2.4.1.1 Inverse Simpson Index

Inverse Simpson index of biodiversity was used to compare the diversity of species and their proportion in one site. The Inverse Simpson value was generated using the summary. Single command in Mothur.

#### 2.4.1.2 Rarefaction curve

The Rarefaction curve is a technique used to estimate species richness, by calculation of the number of individual species or OTUs as a function of sampling, developed in 1968 (Sanders, 1968). The number of OTUs in each sample was determined as sequencing depth increases in Mothur. The number of OTUs identified was plotted against sequencing depth for each sample type.

### 2.4.2 Beta Diversity

Various tests were used to measure the dissimilarity between different communities. Some commonly used beta diversity indices include the Bray-Curtis dissimilarity index and the Jaccard index. Bray-Curtis compares the composition and abundance of species in two samples and the Jaccard index largely measures composition.

### 2.4.2.1 Jaccard Index

The Jaccard index, also known as 'Intersection over Union' and the Jaccard similarity coefficient was originally created by Paul Jaccard is a statistical method that can be used to compare the similarity between different groups. In the current study, the Jaccard index was used to determine the differences between leukoplakia disease and contralateral and healthy controls. In this following example, 4 species are identified in two hypothetical samples. The Jaccard index shows that the two hypothetical groups were 50% similar.

| | Species 1 | Species 2 | Species 3 | Species 4 |
|---|---|---|---|---|
| Group 1 | 1 | 1 | 1 | 0 |
| Group 2 | 1 | 0 | 1 | 1 |

The similarity between these two groups can be measured by using this Jaccard format:

$$J = a/(a+b+c)$$

a = number of common species

b = number of species unique group 1

c = number of species unique group 2

Species 1 and 3 were present in both groups, the intersection (a) is 2. Species 2 and 4 are unique to group 1 and 2 respectively, so the union is 2+1+1. Therefore:

$$J= 2/2+1+1$$

$$J= 2/4= 0.5 \text{ or } 50\% \text{ similarity}$$

### 2.4.2.2 Bray-Curtis dissimilarity

While the Jaccard index measures dissimilarity depending on the presence or absence of species, the Bray-Curtis dissimilarity index assesses quantity as well as presence. The Bray-Curtis dissimilarity index is a common way to determine the differences between samples or groups in ecological studies. An example of how this is calculated is shown below where two groups contain species A to E. The sum of the difference in abundance

of each species is divided by the total sum of all entities. Where the dissimilarity value is 1, the difference between two groups is 100%, meaning they are completely different.

| | A | B | C | D | E | sum |
|---|---|---|---|---|---|---|
| Group 1 | 11 | 0 | 7 | 8 | 0 | 26 |
| Group 2 | 24 | 37 | 5 | 18 | 1 | 85 |

$$\frac{|11-24| + |0-37| + |7-5| + |8-18| + |0-1|}{26+85} = \frac{13+37+2+10+1}{111}$$

Bray Curtis Dissimilarity equal 63/111= 0.56 or 56%

### 2.4.2.3 Phylogenetic analysis

Based on the Jaccard index and Bray-Curtis dissimilarity values, phylogenetic analysis was carried out in Mothur to generate trees showing the relationships of the samples. These trees were used to determine if the clustering of samples was significantly different between all samples. The Parsimony test was used to determine if clustering of samples was significantly different between the study groups.

### 2.4.2.4 Non-Metric Multidimensional Scaling (NMDS)

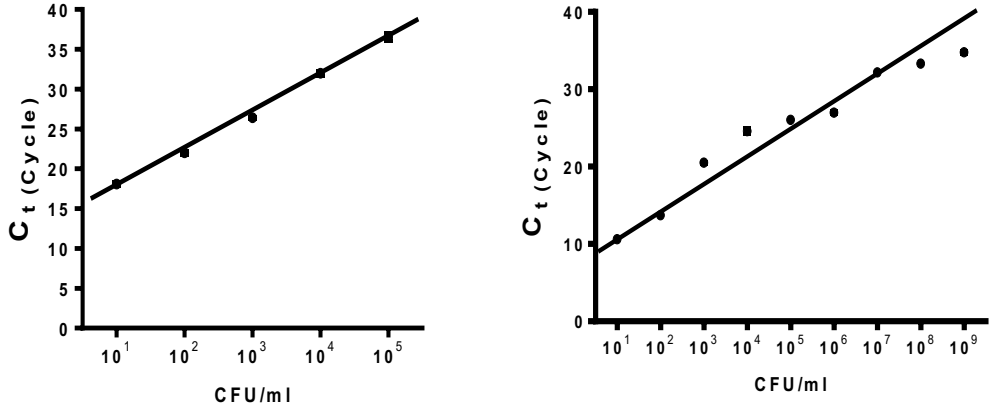
Using a matrix of Jaccard or Bray-Curtis dissimilarity values, the major trends in variation in the datasets were identified using the non-metric multidimensional scaling (NMDS) tool in Mothur. This analysis expressed the variation in population structure across 3 main axes which could be visualised on a 3-D plot generated in R. The analysis of molecular variance (AMOVA) test was used to determine if the distribution of samples was significantly different.

## 2.5 Linear Discriminant Analysis Effect size (LEfSe)

LEfSe was applied to identify any significant enrichments in specific species or OTUs in all samples. LEfSe is a tool that can be used to identify significantly different entities in large biological data sets, here the relative abundance of species or OTUs (White *et al*., 2009; Segata *et al*., 2011). LEfSe was performed using the Huttenhower lab Galaxy web application. LEfSe first performs a Kruskall-Wallis test between the two groups under investigation. It then confirms these results by performing a paired Wilcoxon test between pairs of samples from each group. Finally, a Linear Discriminant Analysis (LDA) model is generated where the features are ranked based on the effect size with which they differentiate the classes.

## 2.6 *Candida* species determined using quantitative Real-Time PCR (qPCR)

The standard curves were generated to estimate *candida* levels using a serial dilution of *C. albicans* SC5314 (Gillum *et al.*, 1984). Positive DNA ranging in concentration from 10 to $10^8$ CFU/ml. Parallel to this, the concentration of bacteria in each sample was estimated using a standard curve generated from a serial dilution of *S. mitis* NCTC12261 DNA (10 to $10^9$ CFU/ml). Amplifications were carried out using primers of qPCR-16S-F and qPCR-16S-R (Table 2.4). The number of CFU/ml were estimated by interpolation of the Ct values determined in each sample using the standard curves below in GraphPad Prism 7.1(GraphPad Software, La Jolla, CA, United States).




(a) (b)

Figure 2.2a Showing standard curve of the 16S rRNA gene of *S. mitis* NCTC1226, (b) Showing standard curve of *C. albicans* SC5314 generated by 18S amplification.

## 2.7 Culture and identification of *Rothia* spp.

### 2.7.1 DNA isolation

DNA extraction from presumptive *Rothia* spp. isolates was performed using the phenol: chloroform: isoamyl-alcohol method. A single colony was used to inoculate 6 ml of Brain heart infusion broth (BHI; Oxoid, Fisher Scientific, Ireland) and grown overnight at 37℃ in a shaking incubator at 200 rpm. The next day, cells were harvested in a 1.5 ml Eppendorf tube by centrifugation at 18,000 × *g* for 1 min. The media was discarded, and the pellet resuspend in 300 µl of breaking buffer and incubated at room temperature for 5 min. The suspension was then transferred into a screw-capped tube containing 250 *g* of glass beads (beads were previously sterilised by autoclaving at 120℃ for 15min) and 200 µl of phenol: chloroform: isoamyl-alcohol (24:24:1) was added. Samples then were shaken for 1 min at setting 4.5 in the bead beater (Sartorius, Goettingen, Germany), after which the samples were centrifuged at 18,000 × *g* for 10 min. In a fresh tube the aqueous phase was extracted with the same amount of chloroform: isoamyl-alcohol (24:1). Samples were centrifuged for 3 min at 18,000 × *g*, and the aqueous phase was transferred into a new tube. To precipitate the DNA, 20 µl of sodium acetate solution (3 M) was added followed by 400 µl of ethanol (100%). Precipitated DNA was centrifuged at 18.000 × *g* for 10 min. The pellet was washed using 400 µl of 70% (v/v) ethanol and air dried. Pellets were suspending in 50 µl of $dH_2O$ (Millipore), and stored at -20℃.

### 2.7.2 PCR amplification

Polymerase chain reaction (PCR) experiments were performed in order to identify presumptive isolates of *Rothia* spp., using primers 27F and 1492R to amplify the 16S rRNA gene, PCR conditions were 95℃ for 45 s, followed by 30 cycles of denaturation at

95℃ for 45 s, annealing at 55℃ for 45 s and an extension at 72℃ for 30 s, with a final incubation at 72 ℃ for 8 min.

After that, DNA purification was carrying out using AMPure XP magnetic purification beads (Section 2.3.3). DNA products were then diluted to 10 ng/µl and sent to Source Biosciences (Units 1 & 2 Riverstown 5 Complex, Riverstown Industrial Estate, Tramore, Co. Waterford Ireland) for Sanger Sequencing with primers 27F and 1492R as required. Sequences were BLAST searched against the HOMD database in order to identify the bacterial species.

## 2.8 Growth curve analysis

The growth of bacterial strains in the presence of different concentrations of ethanol was assessed by growth curve analysis. After reactivating bacterial strains using Brain-Heart infusion agar at 37℃ overnight, a single colony was growing for each strain in BHI broth and incubated overnight at 37℃ in a shaking incubator. Next day, 1 ml of the culture was centrifuged and resuspend in 1 ml of PBS buffer. In order to measure the optical density (OD) at 600 nm, a spectrophotometer (UV-Spectrophotometer Evolution 60s, Thermo Scientific, Waltham MA, USA) was used to determine the bacterial optical density using 1:10 dilution of the bacterial suspension in PBS buffer.

Following this, bacterial cultures were diluted to 0.035 OD at 600 nm in 6 ml of BHI. For negative control, tubes with BHI broth without ethanol added, ethanol was then added into other two tubes at concentration 2% (v/v) and 4% (v/v) of ethanol final concentration (3 replicates of each concentration were investigated). Cultures were incubated in a shaker incubation, at 37˚C. The growth curves were then determined measuring the OD at 600 nm every hour for 8 h. Growth curves values and graphs were then generated using GraphPad Prism 7.1.

## 2.9 Measurement of Acetaldehyde Produced by oral Microbes

### 2.9.1 Acetaldehyde reaction

Bacterial species were grown overnight in BHI broth (at 37 ˚C and centrifuged at 10,000 x *g* for 1 min. The pellets were then washed and resuspend in 1x PBS buffer. Optical density (OD) was measured at 600 nm using the UV spectrophotometer and each suspension was then adjusted to 0.5 OD. 450 µl volume of each strain was transferred into new screw-capped tubes (2 ml), 3 of these samples were used as negative controls (10 mM glucose was added). To the remaining tubes, 50 µl of ethanol at a concentration 1 M was added (final concentration at 100 mM). Samples were then incubated for 2 h at 37 ˚C, and then frozen at -80 ˚C to stop the reaction.

### 2.9.2 Acetaldehyde detection using colorimetric kits

A colorimetric aldehyde detection assay (Sigma-Aldrich) was used for aldehyde detection. Frozen reaction vials were incubated at room temperature until completely thawed, then the vials were centrifuged briefly. To prepare the aldehyde standard (glyceraldehyde), 1 ml of dilution buffer was added to the aldehyde standard to generate a 10 mM aldehyde standard stock solution. The stock was mixed by pipetting and stored at -20 ˚C if not immediately used. To prepare a working concentration at 1 mM, 100 µl of the 10 mM stock was mixed with 900 µl of PBS buffer. A series of standards were prepared at 1 µM, 50 µM, 100 µM, 250 µM, 500 µM, 750 µM, 1000 µM. For the assay reaction, a 96-well plate was used. A 50 µl volume from the reaction vials (positive and negative) was added to triplicate wells. A 50 µl volume of the freshly prepared colorimetric detection reagent (prepared fresh) was added in each well (including standards) and incubated for 1 h in a dark place at room temperature. The absorbance was measured at 540 nm in a Genios 96-well plate reader (Tecan, Männedorf, Switzerland).

The absorbance values were used to estimate the concentration of aldehydes generated in each reaction by interpolation of a standard values of absorbance values generated by the aldehyde standards using GraphPad Prism 7.1.

### 2.9.3 Acetaldehyde detection by gas-chromatography mass spectrometry (GC-MS)

Gas-chromatography mass spectrometry was used in order to specifically measure the acetaldehyde produced by microorganisms we used GC-MS. Samples for analysis were sent to the Mass Spectrometry laboratory at Dublin Institute of Technology (DIT). The standard aldehyde solutions (1 μM, 5.0 μM, 100 μM, and 200 μM) were prepared from a stock solution of acetaldehyde (0.01 M) in dichloromethane (99%, GC grade, Sigma-Aldrich). A 1 ml volume of each standard solution was added to individual polytetrafluoroethylene (PTFE) capped glass vials. The aldehydes were extracted using a 65 μm PDMS/DVB fibre (Sigma-Aldrich). Prior to each use, the fibre was blanked in the GC injection port for 2 min at 230°C to remove any contaminants. The fibre was then exposed to the headspace of a PTFE-capped 4 ml glass vial containing 1 mL of a 10 mg/ml aqueous solution of O-2,2,4,5,6-(pentafluorobenzyl) hydroxylamine (PFBHA, 98%; Sigma-Aldrich). The fibre was then exposed to the acetaldehyde standard for 10 min. The PFBHA formed oxime adducts (cis- and trans- isomers) with the acetaldehyde. The fibre was then inserted into the injection port of the GC-MS and left to desorb for 2 min. The instrument used for this analysis was the varian 3800 gas chromatograph with the Varian Saturn 2000 GC/MS/MS detector using the parameters in Table 2.6. For analysis of the bacterial samples, the reactions were thawed. A blanked fibre was exposed to PFBHA for 10 min, then the fibre was exposed to the headspace of the reaction samples for 10 min. The fibre was then desorbed in the injection port for 2 min and the analysis

was left to run. Any acetaldehyde presents in the samples then formed the same oxime adducts as previously mentioned.

**Table 2.6** Materials and conditions of the GC-MS protocol

| **Parameter** | **Setting** | | |
|---|---|---|---|
| Injection port temperature | 230$^{o}$C | | |
| Carrier gas and flow rate | He, 1 mL/min | | |
| Split Ratio | 100 :1 | | |
| Solvent delay | 2.5 min | | |
| Column | Chrompack capillary column CP-SIL 8 CB 25m x 250 µm x 0.25 µm # CP7451 | | |
| Oven temperature | Gradient | Temperature | Hold Time |
| | - | 40˚C | 1 min |
| | 10˚C/min | Until 140˚C | 2 min |
| Run time | 13 min | | |

## 2.10 Tissue Culture

Human buccal epithelia cells (TR146) was grown in culture and a humid environment at 37 ˚C and 5% (v/v) $CO_2$ using Dulbecco's modified Eagle's Medium (DMEM), supplemented (Sigma-Aldrich, Co. Wicklow, Ireland) with bovine serum (FCS; Sigma-Aldrich) at 10% (v/v) final concentration and 1% of penicillin-streptomycin (Complete DMEM). Culture media was routinely changed daily. Cells were incubated until they reached 80% to 90% confluence, at which time cells were treated with Trypsin-EDTA to

detach the cells from the plate. Harvested cells were then transferred into a centrifuge tube (50 ml) and the cells were spun at 250 x *g* to create a pellet. Pelleted cells were then resuspended in complete DMEM medium and the cell number was then counted using an improved neubauer haemocytometer slide.

### 2.10.1 Oxidative Stress Assay

Harvested cells were adjusted to $2\times10^4$ cells per ml and 0.1 ml of the cell suspension was added into the wells in a 96-well plate. Cells were incubated from 24 to 48 h, after that, cell media was removed, cell monolayers then were washed with 1x D-PBS buffer. To prepare the cells for the assay, 0.1 ml of 1x DPBS buffer containing 100 µM of DCF-DA (dihyrodichlorofluorescein diacetate) was added to each assay well. A series of wells was also prepared containing 1x DPBS with 100 µM DCF-DA and 100 µM of the ROS scavenger N-acetylcysteine (NAC), and incubated for 1 h.

Following incubation, the cells were washed in 1x of DPBS buffer. In order to induce oxidative stress on the cells, acetaldehyde at two concentrations (50 µM and 100 µM) was added. For positive control, $H_2O_2$ was used at concentration of 30 mM in 1 x DPBS buffer was added. For negative control by using 1 x of DPBS buffer. Cells were incubated for 3 h under standard conditions to induce stress. Fluorescence was measured in a Genios 96-well plate reader (Tecan) using an excitation wavelength of 485 nm and a detection wavelength of 530 nm. Fluorescence data were blanked and reported as a percentage of the negative control.

For microscopic analysis, cells were visualised using a Zeiss AX10 epifluorescence microscope. The cells were in preincubated in 100 µM of DCF-DA only, and control cells in 100 µM of DCF and 100 µM of NAC for 1 h. In order to induce cell stress, cells were incubated in 50 µM and 100 µM of ACH and also in 30 mM of $H_2O_2$ as a positive control.

Following incubation for 2 h, cells were then washed with 1x DPBS, stained with 4′,6-diamidino-2-phenylindole (DAPI), (fluorescent stain) (Sigma-Aldrich), visualised the nucleus using fluorescence microscope.

# Chapter 3

Investigation of the microbiome of oral leukoplakia by using high-throughput sequencing of 16S rRNA genes.

## 3.1 Introduction

There has been a significant increase in the study of oral microbes and their potential role in the development of OSCC in recent decades. Oral cancer accounts for ~3% of all malignancies world-wide, with approximately 300,000 patients annually diagnosed with oral cancers worldwide (Markopoulos, 2012). Oral cancer is the fourth most common cancer in men globally. Approximately 90% of oral cancers are oral squamous cell carcinoma and these cancers may arise *de novo*, or from malignant precursors such as OLK (Don *et al*., 2015). Studies have shown that OLK may undergo malignant transformation, however, the rate can vary by study population from between 0.6% to 20% (Abidullah *et al*., 2014). However, the mechanism of malignant transformation of OLK to OSCC is still poorly understood (Dong *et al*., 2015; Meurman, 2010). Oral leukoplakia was defined by WHO as a potentially malignant disorder (Mortazavi *et al.*, 2014).

There are different hypotheses proposed to explain the possible relationship between oral microbes and the development of oral cancer (Perera *et al*., 2016). For example *N*-nitroso compounds (Mills and Alexander, 1976) and acetaldehyde (Marttila *et al*., 2013; Moritani *et al*., 2015) have been proposed as microbial carcinogens that can induce OSCC. Identification of the bacterial species colonising malignant tissues and the relationship between these organisms and malignant transformation has become an intense area of scientific research due to the increasing evidence linking oral microbes with cancers of the GI tract. An important element of this research is the identification of the oral microbes colonising premalignant and malignant tissue in order to investigate their possible role in this transformation. Approximately 50% of oral bacteria are non-cultivable (Paster *et al*., 2001), and cannot be identified by laboratory culture and phenotypic tests. For culture independent studies, 16S rRNA sequence analysis has

become the method of choice for bacterial identification, the 16S rRNA gene consists of 1500 bp (Janda and Abbott, 2007), and contains 9 variable regions (V1 to V9). The gene can be amplified as a full-length sequence or, more commonly, one or more of the variable regions are selected for bacterial identification. The selection of variable regions varies between studies and depends on the species under analysis and the sequencing technology being used. This study was carried out using the V1-V2 regions of the 16S rRNA gene which is approximately 364 bp in length, making this region suitable for analysis using the Illumina MiSeq. Diaz *et al*. (2014) showed that the V1-V2 region could provide accurate taxonomic classification of most oral taxa with similar accuracy and discrimination as the entire 16S rRNA gene.

## 3.2 Material and Methods

Ethical Approval for this study was granted by the Joint Hospitals' Research Ethics Committee (REC), Tallaght Hospital, Dublin. All laboratory experiments starting with sample collection up to the library DNA preparation were implemented at the DDUH microbiology laboratory. The DNA sequence analysis was performed using the Illumina MiSeq at St James's Hospital, Dublin.

After swab samples were collected, bacteria were harvested by dispersion of the sample in 300 µl of TES buffer. DNA extraction was performed using the MasterPure DNA purification kit (Section 2.3.1). Amplicon PCR (first PCR), was carried out targeting the V1-V2 region of the 16S rRNA gene using primers 27F-YM and 338R (Table 2.4) and the KAPA HiFi Master Mix (KAPA Biosystems). Following this, an index PCR (second PCR) was carried out using index primers for 8 cycles (Nextera XT Kit, Illumina, catalog number FC-131-1001). Libraries were prepared by pooling 5 $\mu$l from each index reaction at concentration 4 nM. Sequencing was carried out using the Illumina MiSeq 600-cycle kit (2 x 300 bp; Catalogue number MS-102-3003).

.

## 3.3 Results

### 3.3.1 Sequence Assembly

Bacteria from a total number of 112 oral swab samples were sequenced using the Illumina MiSeq. Approximately 15,000,000 sequences were obtained. A total of 8 samples yielded less than 3,000 reads and were excluded from the analysis. After assembling contigs from both forward and reverse sequences and alignment to the V1-V2 region, a total of 99,000 unique sequences were identified. For beta diversity measurements, data sets were normalised by subsampling, subsample equivalent to the smallest data set, consisting of 3,967 reads were sampled from each dataset.

### 3.3.2 Phylum level analysis

#### 3.3.2.1 Phyla identified in all samples (buccal and lingual).

Figure 3.1 shows the distribution of sequences classified to the phylum level in all samples, buccal and lingual and controls. In total, 58 % of the sequences were classified to the Firmicutes. The second largest phylum was the Proteobacteria, representing 14% of the sequences. The third most abundant phylum was Bacteroidetes, representing11% of sequences. The Actinobacteria and Fusobacteria accounted for less than 10% of sequencing reads, at 9% and 6%, respectively. Other phyla, including Spirochaetes, TM7, SR1, Tenericutes and Synergistetes accounted for a very small proportion of sequences (~2%). Approximately 1.1% of reads were unclassified.

Samples recovered from buccal and lingual sites were analysed in order to determine if the distribution of the main phyla differed between these two niches.

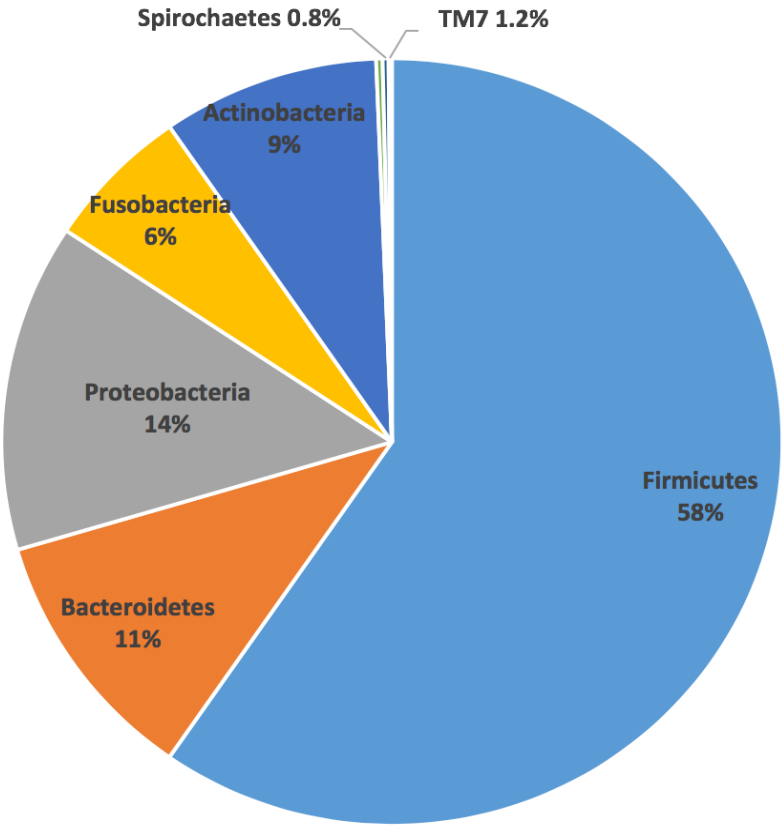



Figure 3.1 Shows the most abundant phyla identified in all buccal and lingual samples

### 3.3.2.2 Phyla level analysis in all buccal samples

Figure 3.2 shows the most abundant phyla identified in buccal samples from leukoplakia patients and healthy controls. The most abundant phylum was the Firmicutes with 60% of the total reads, and the second most abundant was the Proteobacteria with 13%, followed by the Bacteroidetes with 11%. Actinobacteria and Fusobacteria had the same proportion (7%). The remaining sequences were classified among the Actinobacteria, Spirochaetes, TM7, SR1, Tenericutes and Synergistetes (2% of sequences). However, 1% of reads remained unclassified.

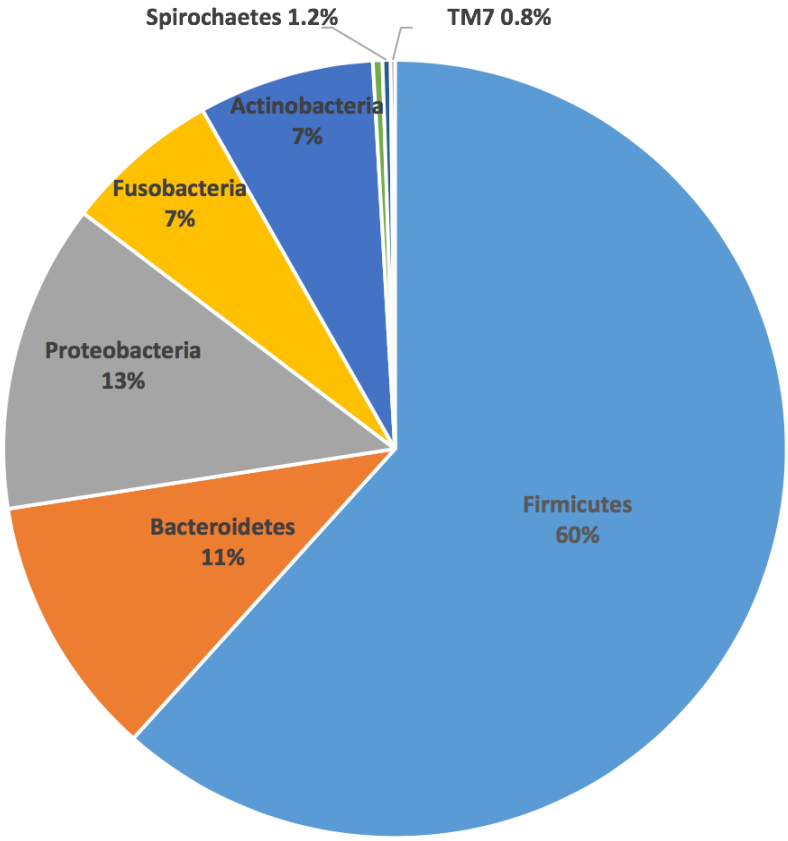



Figure 3.2 Shows the most abundant phyla in buccal samples from leukoplakia patients and healthy controls.

The bar graph in Fig. 3.3 shows the number of reads classified to the 12 phyla identified in buccal samples including samples recovered from patients (both leukoplakia and contralateral normal sites) and samples from healthy control patients.

Overall Firmicutes was the dominant phylum in all three sample types. In contralateral normal samples and healthy controls samples, higher levels of Firmicutes were observed compared to the leukoplakia samples. However, leukoplakia samples had significantly greater levels of Bacteroidetes and Fusobacteria compared to healthy controls and contralateral normal tissue.

Statistical analysis using Metastats showed that the level of Bacteroidetes was significantly higher in samples from leukoplakia compared with healthy controls ($p$ = 0.009) and that the level of Fusobacteria was significantly higher in samples from leukoplakia compared to contralateral normal samples ($p$ = 0.04).

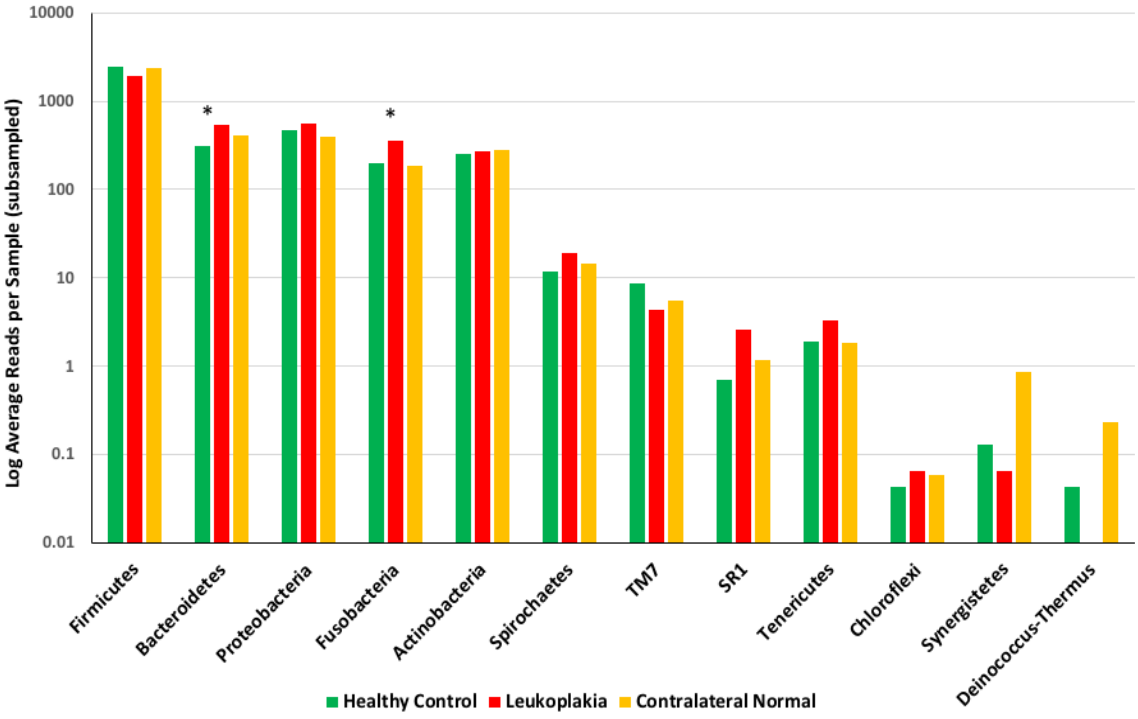



Figure 3.3 Shows the distribution of phyla in buccal samples from patients and healthy controls. Asterisks indicate significant differences ($p < 0.05$).

### 3.3.2.3 Phylum level analysis of lingual samples

Analysis of the lingual samples from all patients showed that the majority of reads were classified at (57%) to the Firmicutes phylum. Proteobacteria was the second most abundant phylum with 15% of the reads and 11% of the reads were classified to the phylum Actinobacteria and Bacteroidetes, with the remaining 6% of reads classified to the phylum Fusobacteria. Spirochaetes accounted for 1.3% of the reads and 0.7% of the reads were classified to TM7 phylum. Approximately 0.9% of the reads were unclassified (Fig 3.4).

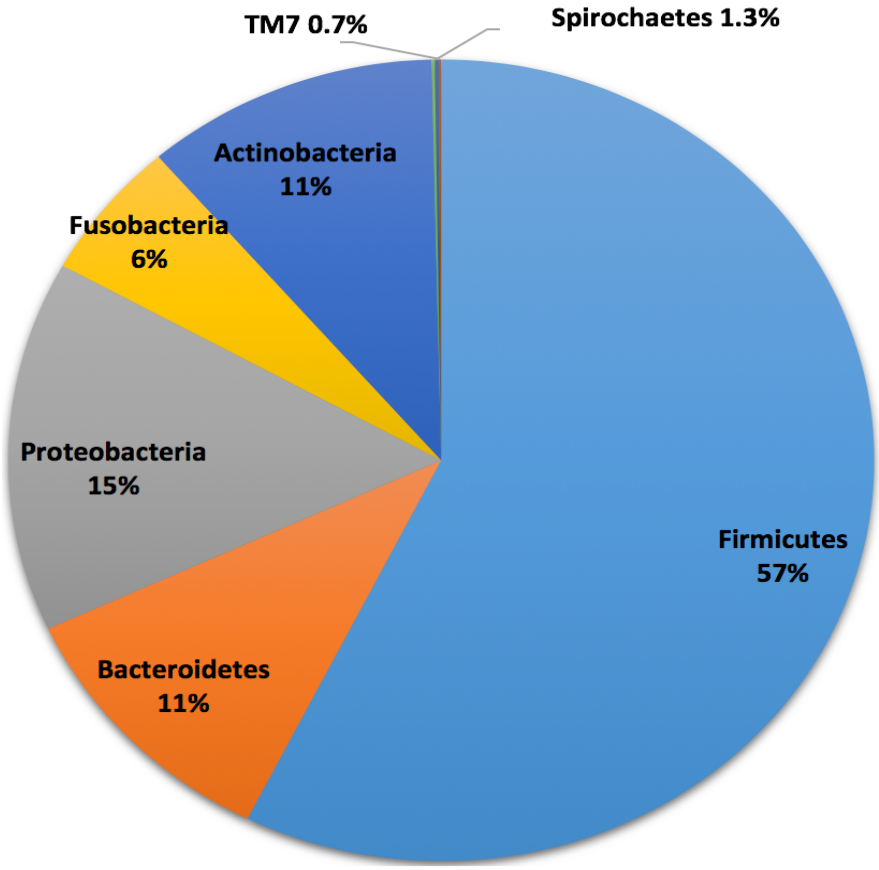



Figure 3.4 Shows the distribution of phyla in lingual samples from patients and healthy controls.

Figure 3.5 shows the distribution of the phyla in the lingual samples from leukoplakia patients (both leukoplakia and contralateral normal sites) and samples from healthy control patients. Contralateral normal sites had the highest level of Firmicutes and exhibited significantly higher levels of Firmicutes compared to leukoplakia sites ($p$ = 0.0099). The phylum Proteobacteria exhibited significantly higher levels in healthy tissue compared to contralateral normal tissues from patients ($p$ = 0.014). The phylum Fusobacteria had higher levels in samples from leukoplakia sites compared to contralateral normal tissues ($p$ = 0.099).

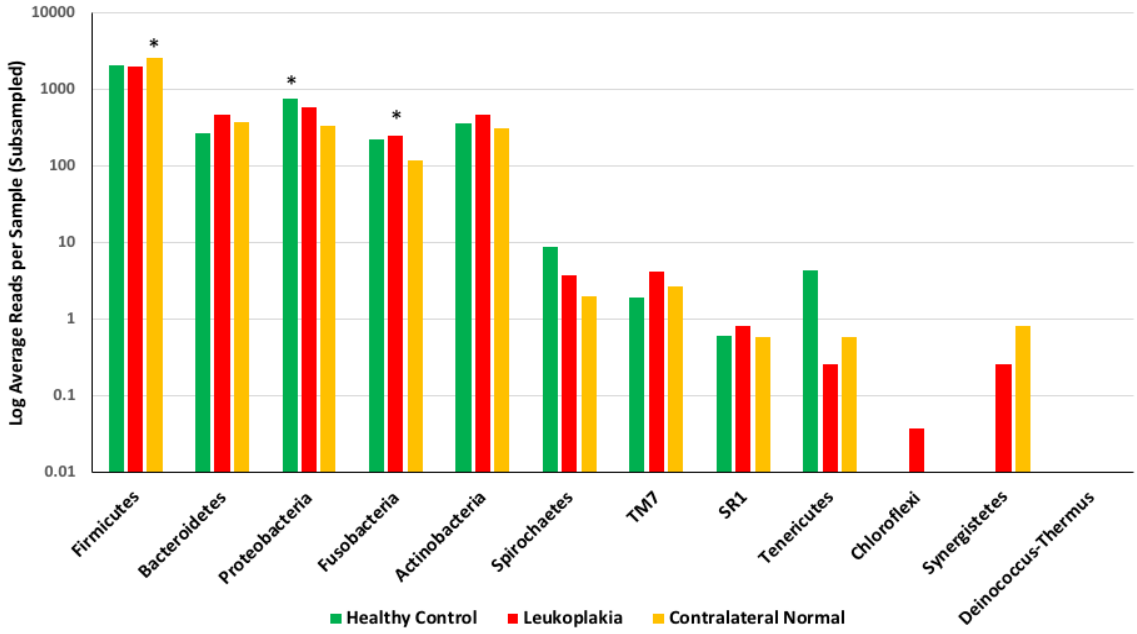



Figure 3.5 Shows the distribution of the phyla in all lingual samples. Asterisks indicate significant differences ($p < 0.05$).

### 3.3.3 Analysis of Operating Taxonomic Units (OTUs)

Further processing of the sequences in Mothur allowed classification of the sequence reads to the level of OTUs at the level of 2% difference in nucleotide sequence. A total of 2030 OTUs were identified. The lowest number of OTUs (64) was found in a lingual leukoplakia sample, whereas the highest number of OTUs (530) was identified in a lingual healthy control sample.

Table 3.1 Showing the average number of OTUs in all samples

| Sample type | OTUs range | OTUs average |
|---|---|---|
| Buccal leukoplakia | 132 - 521 | 287 |
| Buccal contralateral normal | 111 – 403 | 234 |
| Buccal healthy control | 124 - 437 | 243 |
| Lingual leukoplakia | 64 - 392 | 231 |
| Lingual contralateral normal | 68 - 504 | 204 |
| Lingual healthy control | 172 - 530 | 329 |

### 3.3.4 Alpha diversity measurements

#### 3.3.4.1 Inverse Simpson index

The inverse Simpson index was calculated for each sample and the average value determined for each type of sample, both buccal and lingual (Fig. 3.6). Overall, buccal samples exhibited lower levels of biodiversity compared to the lingual samples. In buccal samples, the difference between buccal leukoplakia, buccal contralateral normal and buccal healthy controls were not significant (t-test $p > 0.05$). However, the lingual samples from healthy controls had significantly greater biodiversity compared to lingual leukoplakia and lingual contralateral normal samples (t-test $p < 0.05$).

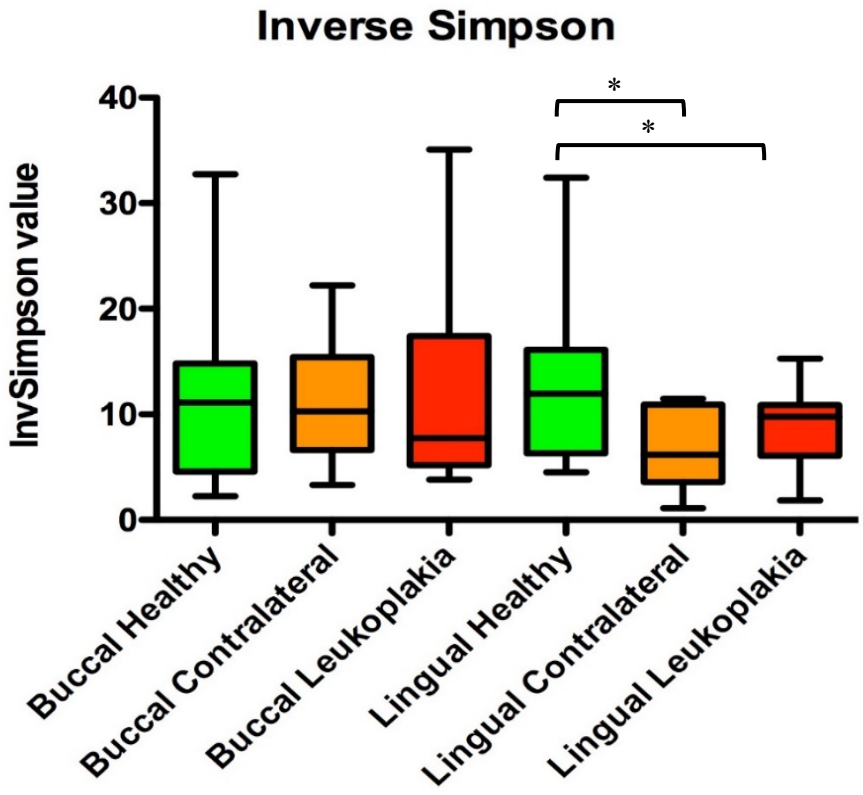
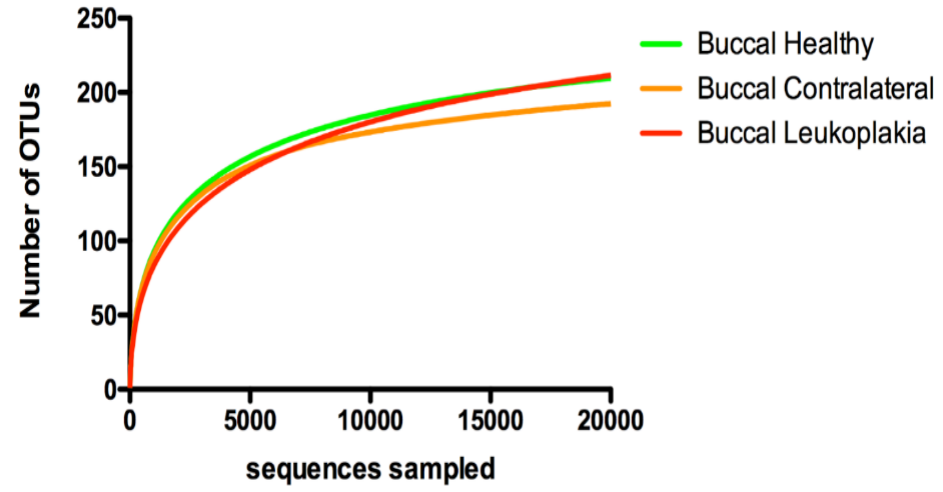


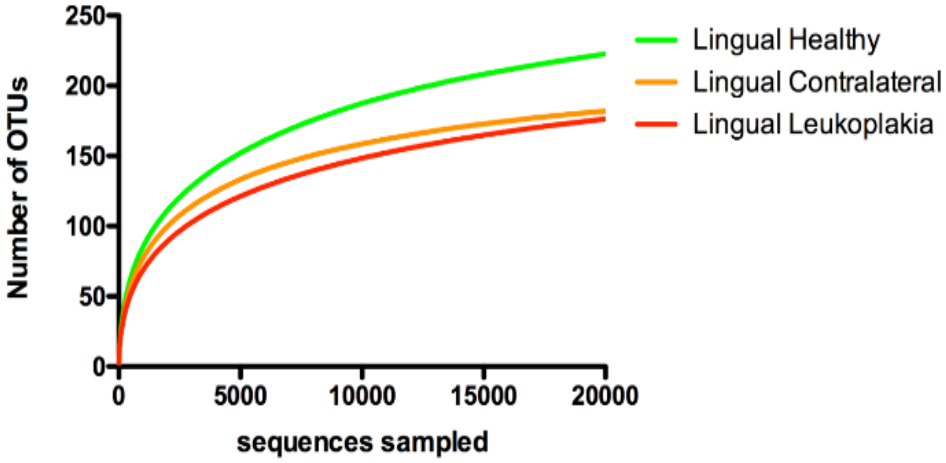

Figure 3.6 Box plot showing the average of the inverse Simpson index in all samples, buccal and lingual. (*= $p < 0.05$).

### 3.3.4.2 Species Richness (Rarefaction curve)

Species richness was estimated by generating a rarefaction curve, illustrating the number of OTUs in each specimen with increasing sampling of sequences. Figure 3.7a shows a rarefaction curve generated from the data for all buccal samples (healthy control, leukoplakia and contralateral normal). Up to 5,000 reads, leukoplakia samples exhibited a lower number of OTUs compared to healthy controls and contralateral normal samples. However, as sampling depth approached 20,000 reads, similar richness in OTUs was observed in leukoplakia disease samples. Figure 3.7b shows a rarefaction curve for all lingual samples showing the increase in the number of OTUs detected as sequence sampling increases. A greater number of the OTUs were detected in lingual healthy control samples compared to leukoplakia and contralateral normal samples.




**(a)**




**(b)**

Figure 3.7 Shows rarefaction curves illustrating species richness in (a) all buccal samples and (b) all lingual samples.

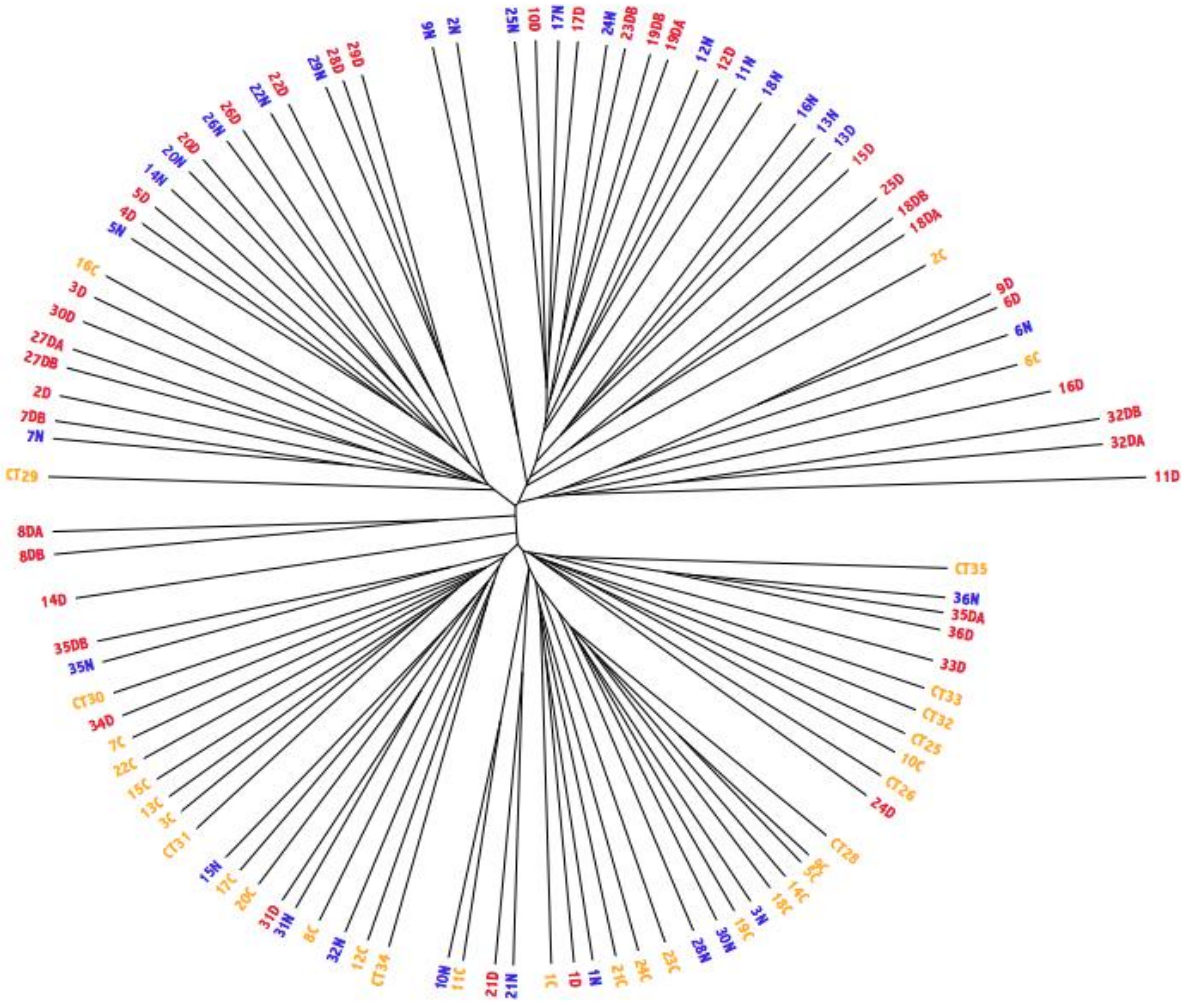
### 3.3.5 Beta diversity

Next, Beta diversity was carried out to measuring the variation in community structure and membership between samples.

#### 3.3.5.1 Phylogenetic analysis

Phylogenetic trees can be used to visualise the relationships between communities and to test whether these communities are significantly different. Trees can be generated using different similarity coefficients, which are used determine the relationships between samples and the branch lengths. In this study, the Jaccard index (examining community membership) and the Bray-Curtis dissimilarity index (examining community membership and structure) were used to determine the similarity of the communities in each sample. Figure 3.8 shows the phylogenetic trees for all samples generated using the Jaccard distance matrix. In order to determine whether samples from all healthy controls, all leukoplakia and all contralateral normal samples clustered differently, the Parsimony statistical test was used, showing that buccal healthy control samples were significantly different compared to buccal leukoplakia samples ($p = 0.001$; Table 3.2). Buccal healthy control and buccal contralateral normal samples were not significantly different ($p = 0.157$). In general, the parsimony statistical showed that lingual samples were not significantly different in phylogenetic trees generated using the Jaccard distance matrix (Table 3.2). A phylogenetic tree was also generated using the Bray-Curtis dissimilarity index (Fig. 3.9).

The parsimony test was used to determine whether samples from healthy controls, leukoplakia and contralateral normal samples were significantly different (Table 3.2). The results indicate that there are no significant differences in buccal samples from patients and controls. However, lingual samples from healthy controls were significantly different compared to leukoplakia ($p < 0.001$) and contralateral normal samples ($p = 0.007$).




Leukoplakia disease  Contralateral normal  Healthy control sample

Figure 3.8 Shows a phylogenetic tree of all samples generated using a distance matrix calculated using the Jaccard similarity coefficient. Legend indicates colour code used to label samples as leukoplakia, contralateral normal or healthy control ($p = 0.05$).

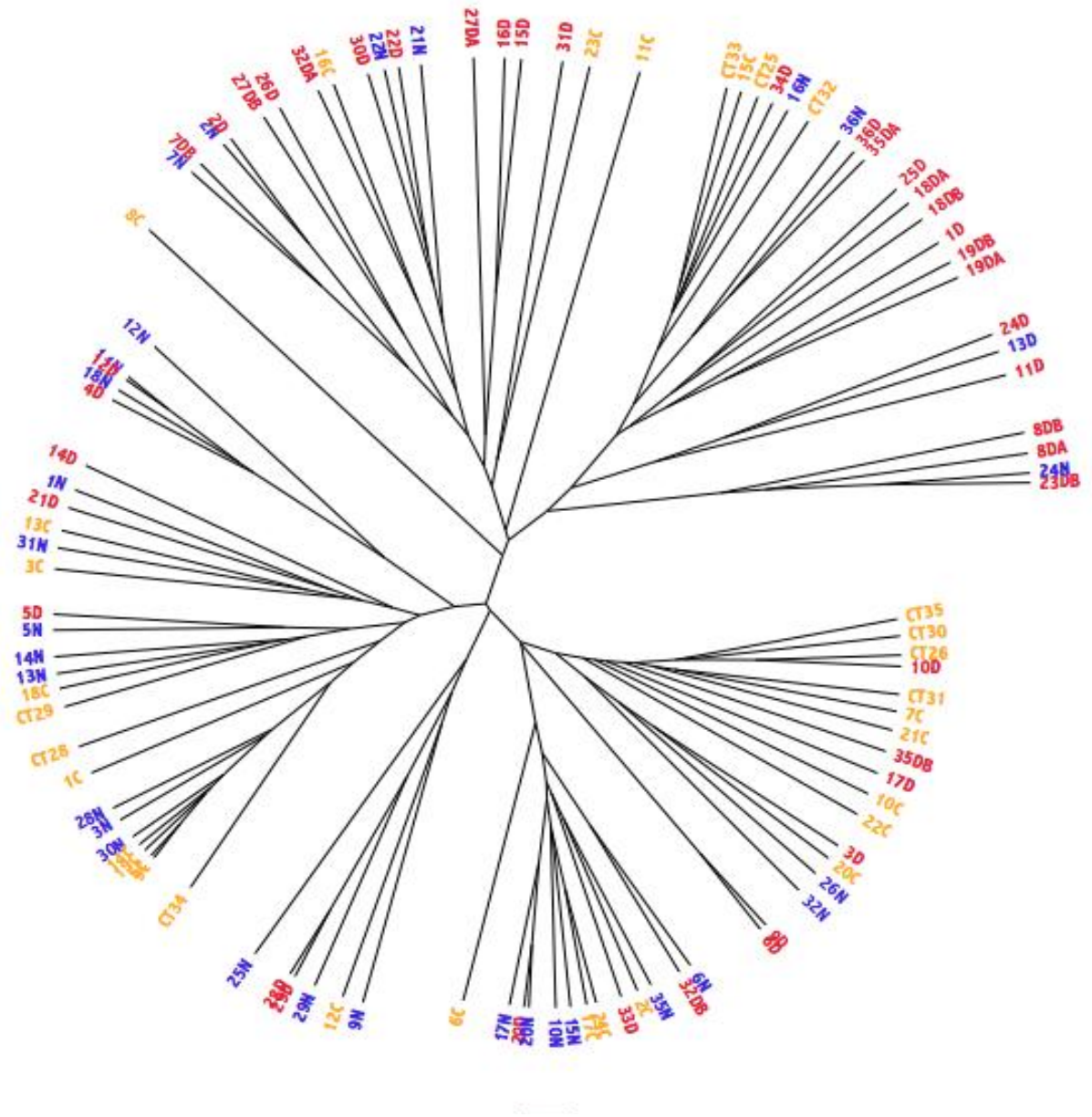



Figure 3.9 Shows phylogenetic tree of all samples generated based on the Bray-Curtis dissimilarity index. Legend indicates colour code used to label samples as leukoplakia, contralateral normal or healthy control ($p$ = 0.04).

**Table 3.2** Shows *p* values of parsimony test analysis of phylogenetic trees generated with Jaccard index and Bray-Curtis dissimilarity values of all samples (buccal and lingual).

| **Parsimony test: Jaccard** | ***p* values** | | **Parsimony test: Bray-Curtis** | ***p* values** |
|---|---|---|---|---|
| BC-BD | $p < 0.001$* | | BC-BD | $p = 0.837$ |
| BC-BN | $p = 0.157$ | | BC-BN | $p = 0.079$ |
| BD-BN | $p = 0.643$ | | BD-BN | $p = 0.369$ |
| LC-LD | $p = 0.075$ | | LC-LD | $p < 0.001$* |
| LC-LN | $p = 0.053$ | | LC-LN | $p = 0.007$ * |
| LD-LN | $p = 1.0$ | | LD-LN | $p = 0.847$ |

BC=buccal healthy control BD=buccal leukoplakia BN=buccal contralateral normal
LC=lingual healthy control LD= lingual leukoplakia LN=lingual contralateral normal
(*) = $p < 0.05$

### 3.3.5.2 Non-metric multidimensional scaling (NMDS)

Non-metric multidimensional scaling is a data visualisation tool that can be used to simplify complex data sets such as 16S rRNA community profiling data. To generate three dimensional NMDS plots the Bray-Curtis distance matrices (Fig. 3.10) and Jaccard index (Fig. 3.11) were employed. Initially all samples were analysed to investigate the major sources of variation in the data. Analysis of molecular variance (AMOVA) showed that the major source of variation across 'axis 1' was the region swabbed (buccal or lingual). This indicated a significant difference in bacterial community structure between the buccal and lingual sites and for this reason, subsequent beta diversity analysis was carried out separately on buccal and lingual samples.

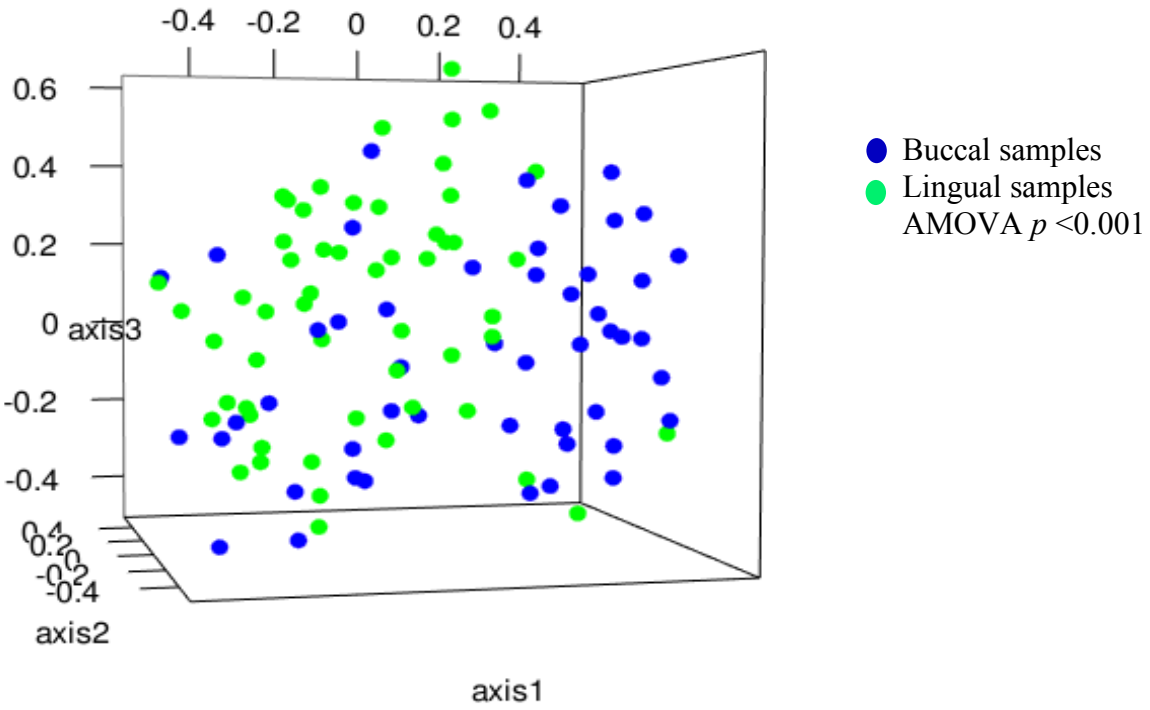



Figure 3.10 NMDS plots generated from distant matrices generated using the Bray-Curtis dissimilarity index

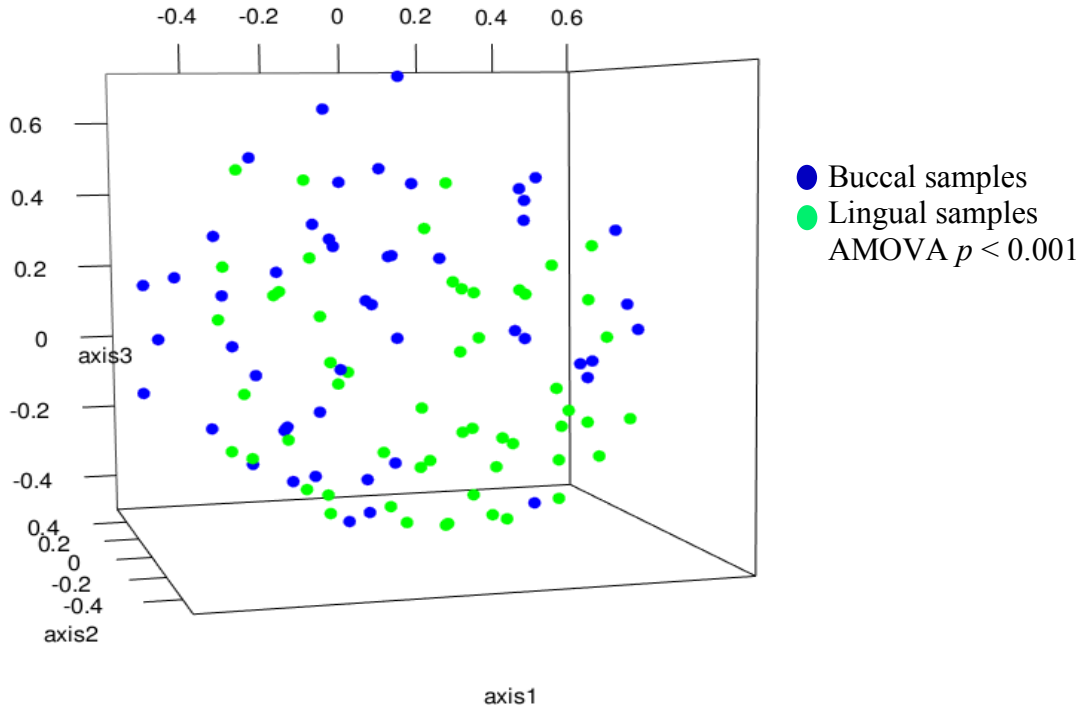



Figure 3.11 NMDS plots generated from distant matrices generated using the Jaccard calculator of dissimilarity ($p$ <0.001).

### 3.3.5.3 NMDS of buccal samples

The Jaccard and Bray-Curtis distance values were used to generate NMDS plots of all buccal samples. Analysis of molecular variance (AMOVA) was used to determine whether these plots could discriminate samples from healthy controls, leukoplakia and contralateral normal sites. Analysis of both Jaccard index (Fig. 3.12) and Bray-Curtis plots by AMOVA (Fig. 3.13) showed that samples from healthy controls and leukoplakias were significantly different ($p < 0.024$) (Table 3.3). Samples from healthy controls were not significantly different from contralateral samples from patients.

**Table 3.3** Summary of *p* values from AMOVA analysis of Jaccard and Bray-Curtis distance values in all buccal samples.

| **Jaccard index** | ***p* values** | | **Bray-Curtis index** | ***p* values** |
|---|---|---|---|---|
| BC-BD-BN | 0.023* | | BC-BD-BN | 0.038* |
| BC-BD | 0.014* | | BC-BD | 0.024* |
| BC-BN | 0.055 | | BC-BN | 0.284 |
| BD-BN | 0.358 | | BD-BN | 0.118 |

BC=Buccal Healthy Control BD=Buccal Leukoplakia BN=Buccal Contralateral Normal

(*) = $p < 0.05$

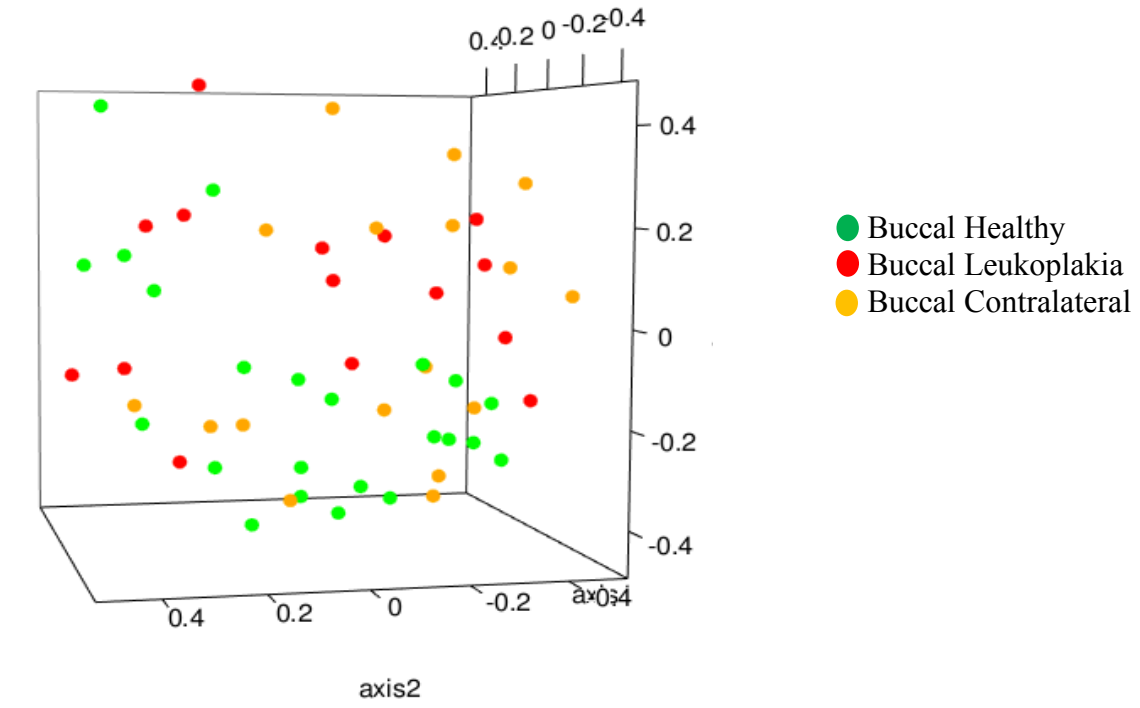



Figure 3.12 NMDS plots of all buccal samples generated using the Jaccard distance calculator.

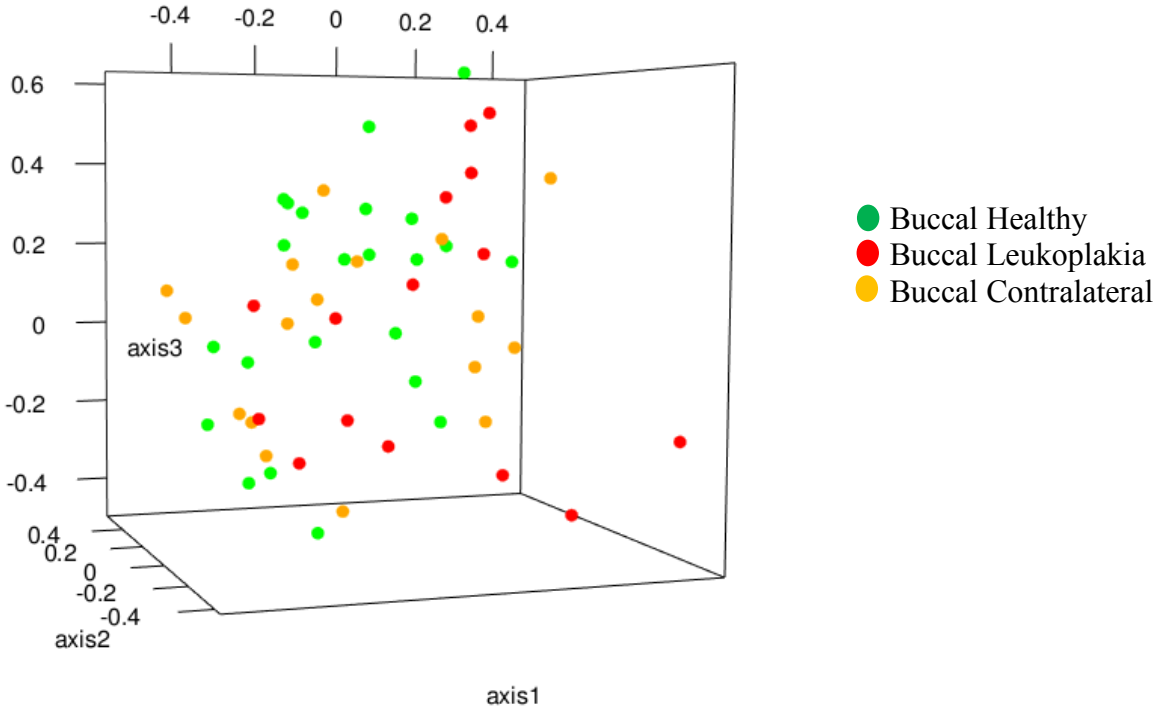



Figure 3.13 Shows NMDS plots of all buccal samples generated using the Bray Curtis dissimilarity index.

### 3.3.5.4 NMDS of lingual samples

The Jaccard index and Bray-Curtis dissimilarity values were used to generate NMDS plots of all lingual samples (Figs 3.14 1nd 3.15). AMOVA was used to determine if communities from healthy controls, leukoplakia and contralateral normal samples could be differentiated (Table 3.4). Analysis of the Bray-Curtis data showed that contralateral normal samples were significantly different from healthy controls ($p = 0.024$). However, no significance was seen between leukoplakia samples and healthy controls samples.

The Jaccard index shows that community memberships of the samples from healthy controls were significantly different compared to lingual contralateral samples ($p < 0.001$) and leukoplakia samples ($p = 0.02$). However, leukoplakia samples were not significantly different compared to contralateral normal samples ($p = 0.414$).

**Table 3.4** A summary of *p* values from AMOVA analysis of Jaccard and Bray-Curtis distance values in all lingual samples.

| **Bray-Curtis** | ***p* values** | | **Jaccard index** | ***p* values** |
|---|---|---|---|---|
| **LC-LD-LN** | $p = 0.031$ | | LC-LD-LN | $p = 0.004$* |
| **LC-LD** | $p = 0.147$ | | LC-LD | $p = 0.02$* |
| **LC-LN** | $P = 0.024$* | | LC-LN | $p = 0.001$* |
| **LD-LN** | $p = 0.072$ | | LD-LN | $P = 0.414$ |

LC=lingual control LD= lingual leukoplakia LN= lingual contralateral normal

(*) = $p < 0.05$

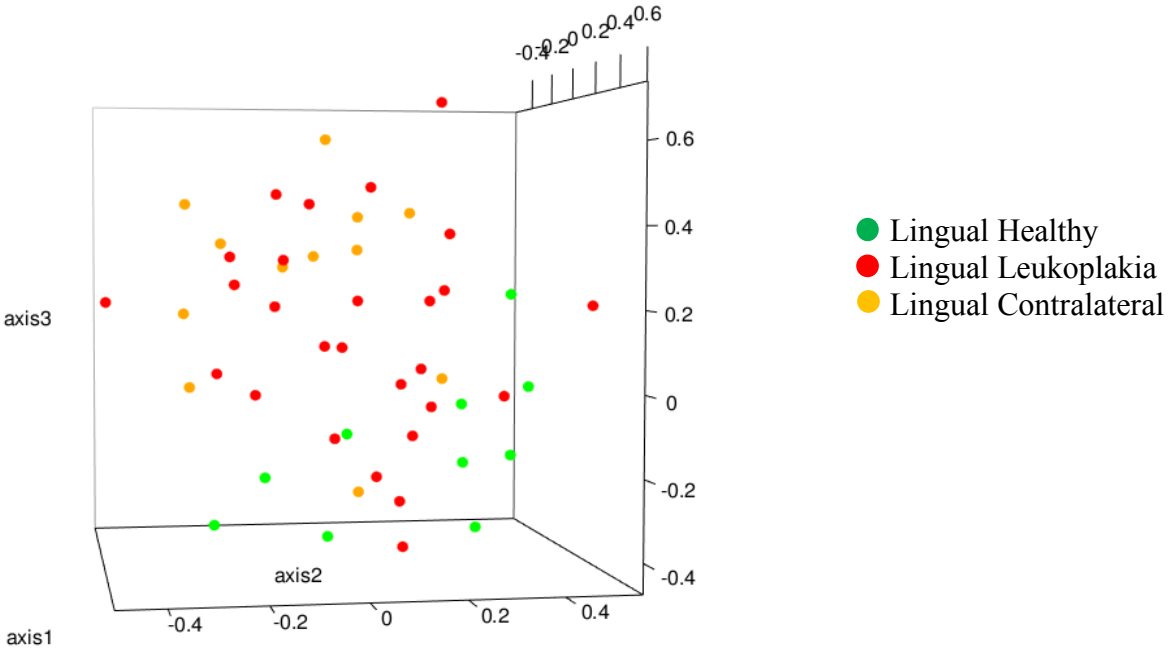



Figure 3.14 NMDS plots of all lingual samples generated with the Jaccard index distance Calculator

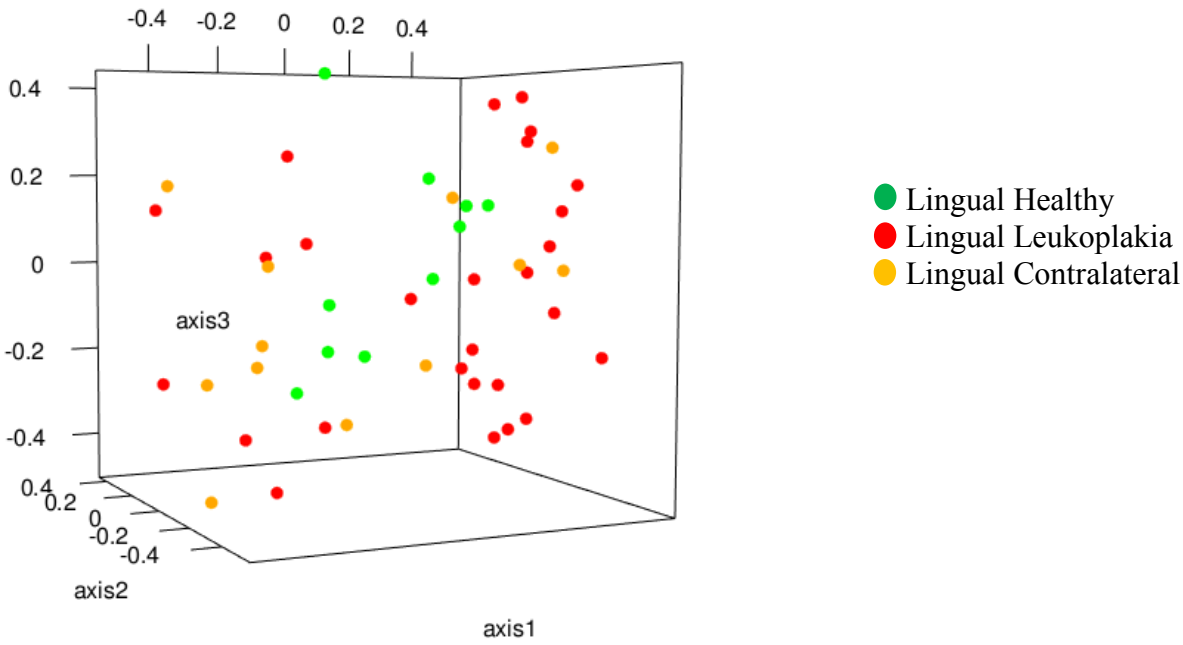



Figure 3.15 NMDS plots of all lingual samples generated with the Bray-Curtis dissimilarity index

### 3.3.6 Linear discriminant analysis effect size (LEfSe)

Linear discriminant analysis effect size was created by the Huttenhower group at Harvard University to identify significant trends in large biological data sets. LEfSe was applied to identify significant enrichments in specific species or OTUs in patients with leukoplakia (using both contralateral and leukoplakia samples) compared to healthy controls. The identity of OTUs with significant enrichments in the following sections were confirmed by BLAST searches using the OTU consensus sequence against the 16S rRNA database hosted by (HOMD).

#### 3.3.6.1 Linear discriminant analysis effect size comparison of all healthy control and all patient samples

Linear discriminant analysis effect size was carried out on all samples from leukoplakia patients compared to all samples from healthy control subjects (buccal and lingual). Patients with leukoplakia harboured significantly greater levels of *Alloprevotella* spp., *Neisseria meningitidis* (OTU0050) and *Leptotrichia* spp. (OTU0081). Healthy controls showed enrichment for members of the *Burkholderiales* among others (Fig. 3.16).

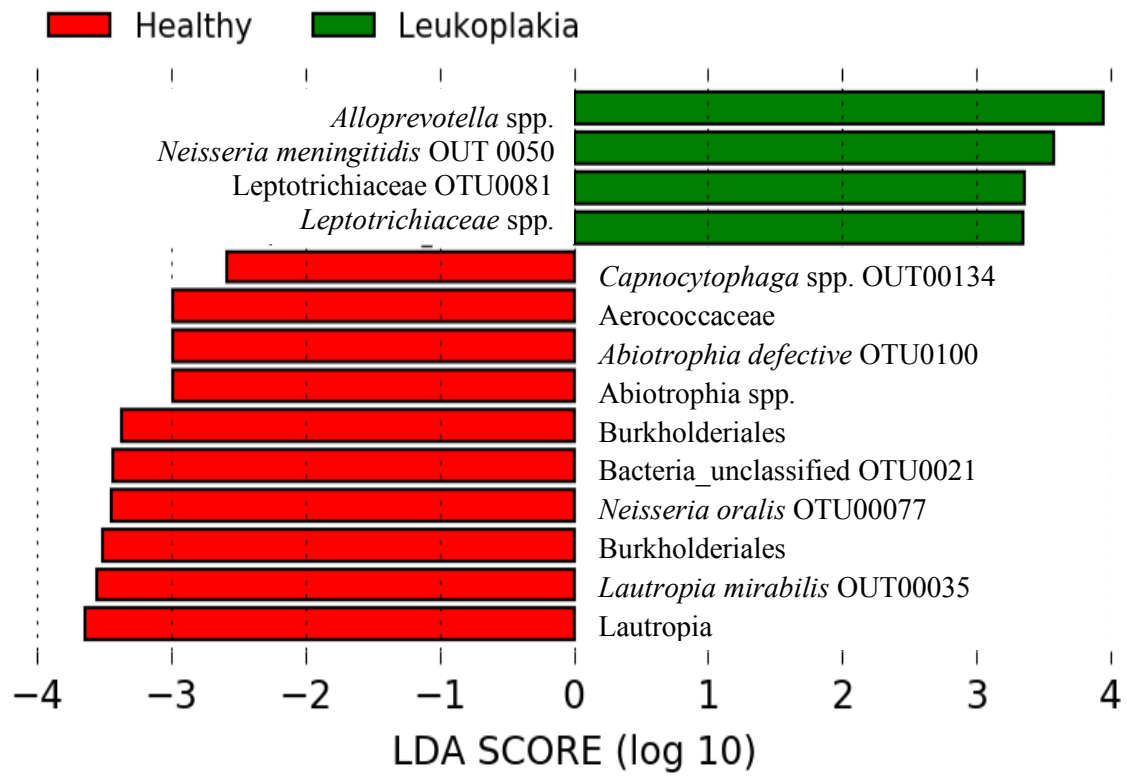



Figure 3.16 LEfSe plot showing significantly enriched organisms ($p < 0.05$) identified in LEfSe analysis of all patient samples (leukoplakia and contralateral normal) compared to all healthy control samples. Only OTUs that gave LDA scores > 2.5 were selected.

### 3.3.6.2 LEfSe comparison of all leukoplakia and all contralateral samples

LEfSe was then used to compare all leukoplakia samples with the contralateral samples from the same patient. This showed enrichment for *Fusobacterium spp*. and the family Fusobacteriaceae in leukoplakia samples, Contralateral samples showed greater enrichment of streptococci among others (Fig. 3.17).

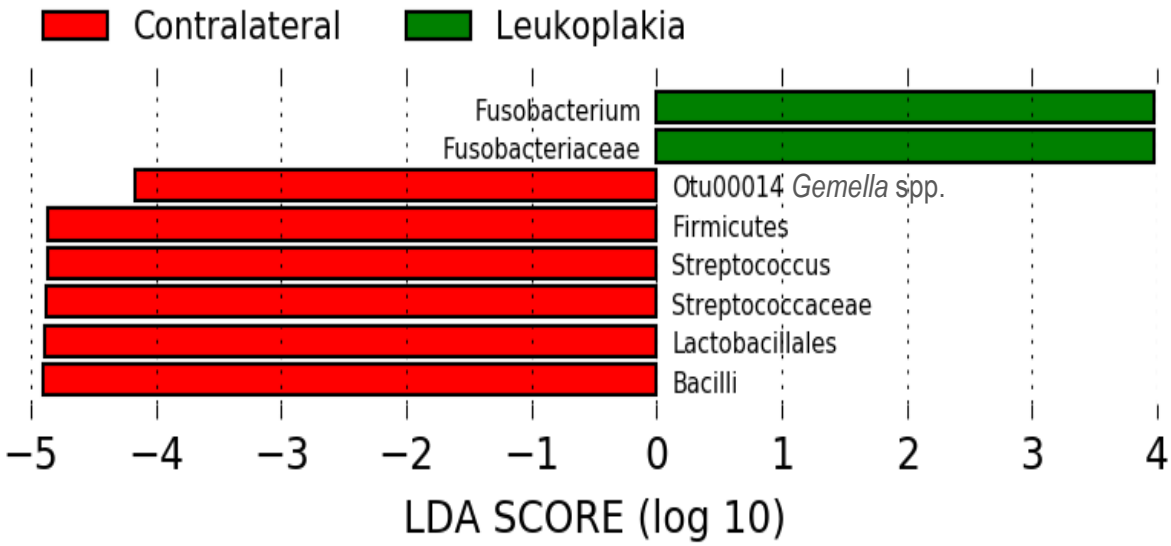



Figure 3.17 LEfSe plot showing significantly enriched organisms ($p < 0.05$) identified in LEfSe analysis of all leukoplakia and all contralateral normal samples. Only results that gave linear discriminant analysis (LDA), only scores greater than or 2.5 were selected.

### 3.3.6.3 LEfSe analysis of buccal samples

Next, we carried out LEfSe analysis on samples recovered from buccal swabs only (patient and healthy control) in order to identify enrichments specific for this niche.

Figure 3.18 shows the OTUs identified by LEfSe when buccal healthy control samples were compared to buccal samples from patients (both leukoplakia and contralateral normal samples). In buccal patient samples, nine OTUs in total were enriched ($p$ <0.05, LDA > 2.5). The OTUs exhibiting the highest enrichment in patients included *Leptotrichia* spp. *N. meningitidis* OTU0050, *R. mucilaginosa* OTU0004, *S. parasanguinis* OTU0005 and *Alloprevotella* species. In buccal healthy control samples, three OTUs were enriched including *Neisseria mucosa* OTU0022 and *Neisseria oralis* OTU0077.

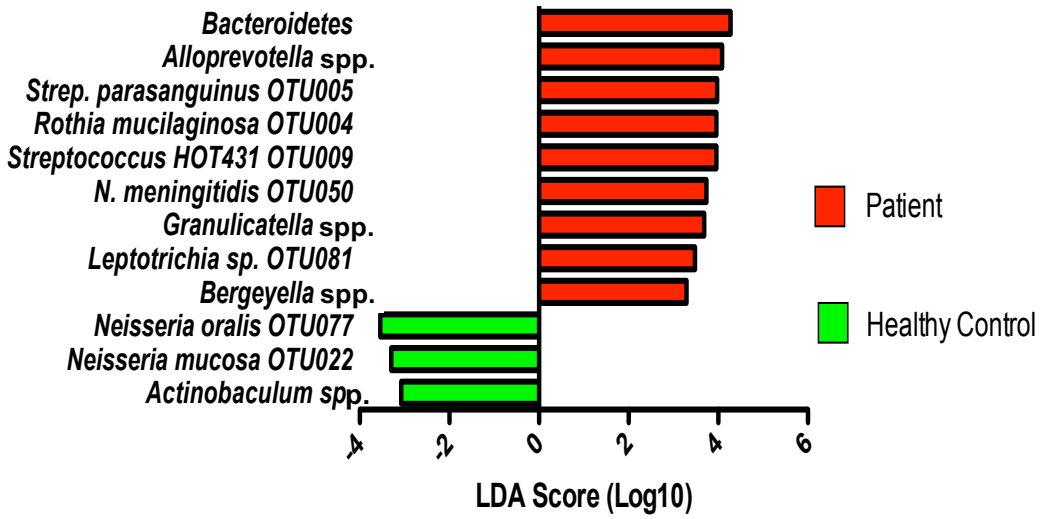



Figure 3.18 Shows a plot of enriched OTUs ($p$ < 0.05, LDA > 2.5) in patient buccal samples from both leukoplakia and contralateral compared to healthy buccal samples.

LEfSe was also applied to compare the buccal leukoplakia and contralateral normal swabs from the same patients. In figure 3.19 shows the OTUs enriched in buccal leukoplakia samples compared to buccal contralateral normal samples. A total number of 3 taxonomic

groups were observed to be enriched in leukoplakia samples and 3 in contralateral normal samples. *Fusobacterium* spp. *Campylobacter* spp., and the family Flavobacteriaceae were enriched in the buccal leukoplakias. Contralateral buccal samples exhibited enrichments for *Gemella haemolysans* (OTU0014), *Streptococcus* spp. and *Neisseria flavescens* (OTU0010).

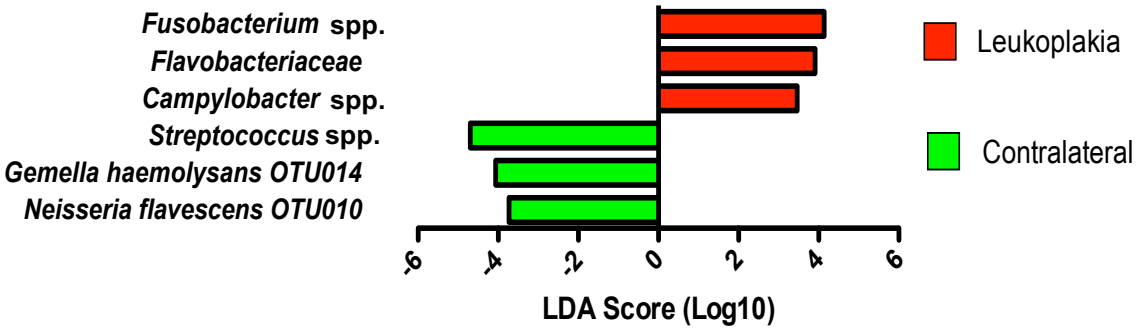



Figure 3.19 Shows a plot of enriched OTUs ($p < 0.05$, LDA > 2.5) in buccal leukoplakia compared to buccal contralateral normal sites.

### 3.3.6.4 LEfSe analysis of lingual samples

Figure 3.20 shows the results of LEfSe analysis of all lingual samples from healthy controls compared to all lingual samples from patients (leukoplakia and contralateral normal). The LEfSe plot shows that 13 OTUs were enriched in lingual healthy control and 3 OTUs in lingual patient samples (leukoplakia and contralateral normal), OTUs were ranked according to LDA score > 2.5. In lingual samples from patients, *Leptotrichia* spp. OTU0081, *N. meningitidis* OTU0050 and *Alloprevotella* spp. OTU0029 were enriched relative to healthy controls. The enriched taxa in healthy controls were dominated by Gram-positive species including streptococci and *Gemella sanguinus* OTU0023, in addition to *N. oralis* OTU0077 and *N. flavescens* OTU0010.

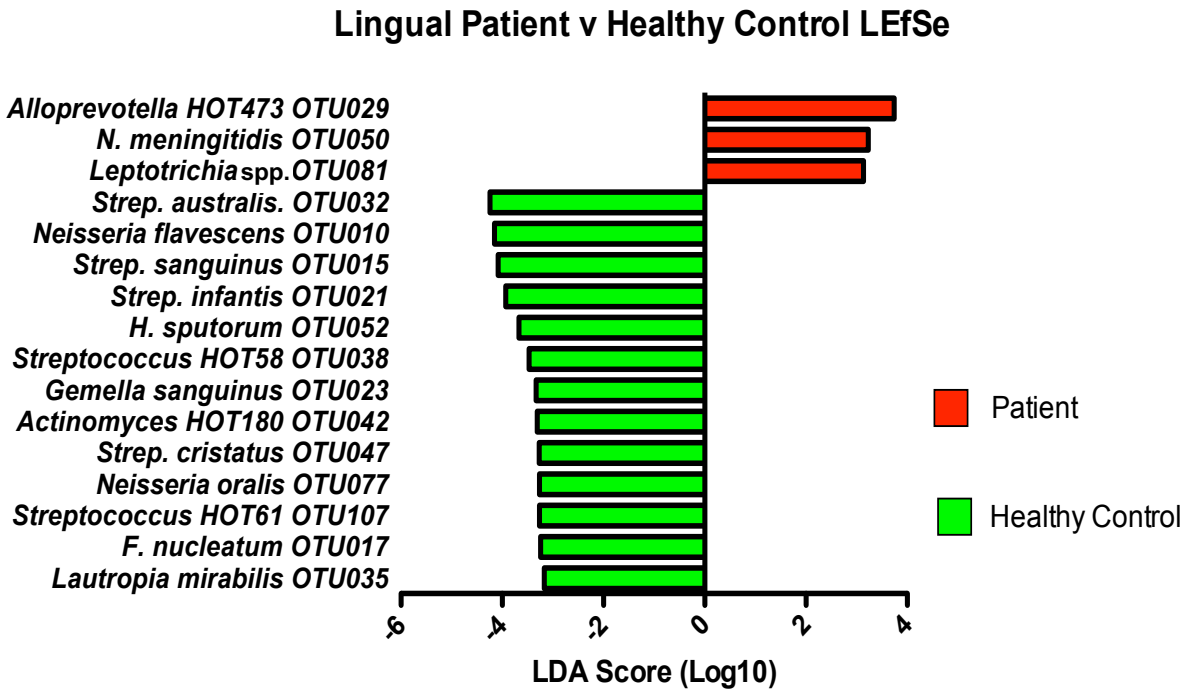



Figure 3.20 LEfSe plot showing enrichments in lingual samples from healthy controls compared to all lingual samples from patients.

Subsequently, LEfSe was used to analyse lingual samples from patients, comparing leukoplakia samples to contralateral normal sites. This analysis showed enrichment for 3 OTUs in lingual leukoplakia samples including *Fusobacterium* spp. *Streptococcus* spp. HOT66 OTU0025 and *Capnocytophaga* spp., while in lingual contralateral normal samples 6 taxa were enriched, LDA > 2.5. Interestingly *Leptotrichia* spp. *N. meningitidis,* and *Alloprevotella* spp. which were previously shown to be patient enriched (Fig. 3.21) enriched only on the contralateral normal tissues of these patients.

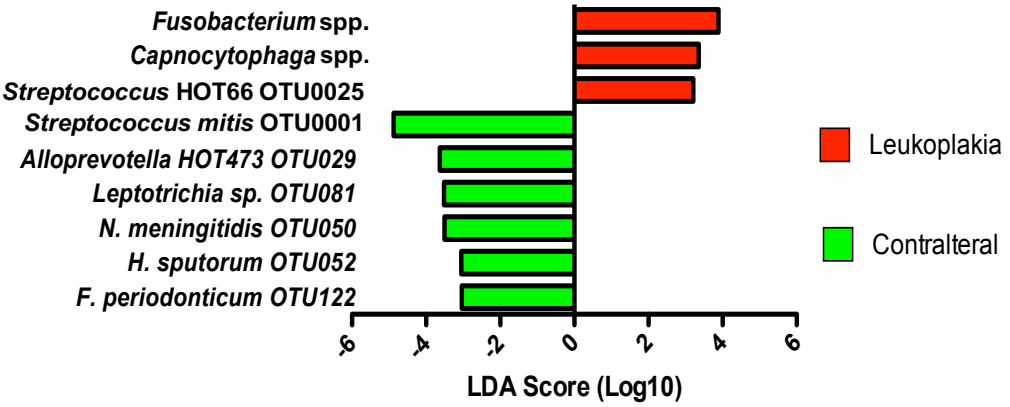



Figure 3.21 Shows LEfSe plot of enrichments OTUs in lingual leukoplakia sites and lingual contralateral normal sites in the same patients ($p < 0.05$, LDA > 2.5).

### 3.3.7 Confirmation of specific enrichments using the Wilcoxon matched-pairs test

In order to confirm the results of LEfSe analysis, further investigation of the levels of selected species and OTUs was carried out using the Wilcoxon matched-pairs test, comparing pairs of leukoplakia and matched contralateral normal samples.

*Fusobacterium* spp. OTU016 was found to be significantly enriched in leukoplakia relative to matched contralateral normal samples $p = 0.039$, (Fig. 3.22). This OTU was confirmed to be most closely related to *F. nucleatum subsp. polymorphum* by BLAST analysis of the consensus sequence. The other major *Fusobacterium* OTU, *Fusobacterium* spp., OTU017, was found to be most similar to *F. nucleatum subsp. vincentii* by BLAST analysis but did not exhibit increased levels in leukoplakia compared to contralateral samples (Fig. 3.22). However, *Leptotrichia* spp., were observed to be significantly enriched ($p = 0.039$) in a proportion of leukoplakia samples (Fig. 3.22). Enrichment for *Campylobacter* spp. and *R. mucilaginosa* were also confirmed using the Wilcoxon matched-pairs test (Fig. 3.23). However, *Capnocytopga* spp., which was identified as leukoplakia enriched by LEfSe analysis, exhibited few examples of

enrichment and were not found to be significantly enriched in leukoplakia using the Wilcoxon matched-pairs test (Fig. 3.23).

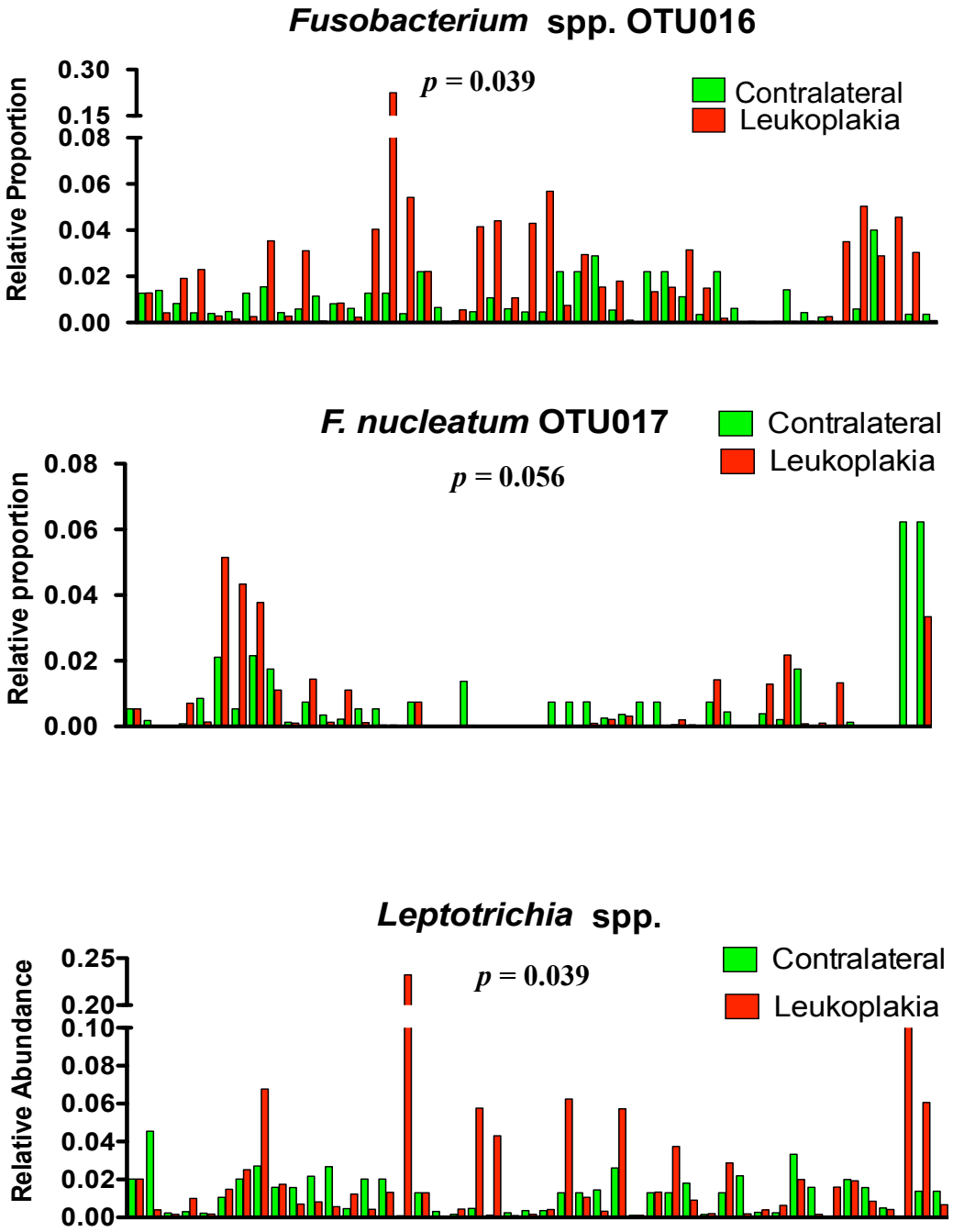



Figure 3.22 Relative proportion of each species in the microbiomes of individual patients. Matched contralateral healthy and leukoplakia samples are shown side by side. The *p value* was calculated using the Wilcoxon matched-pairs test

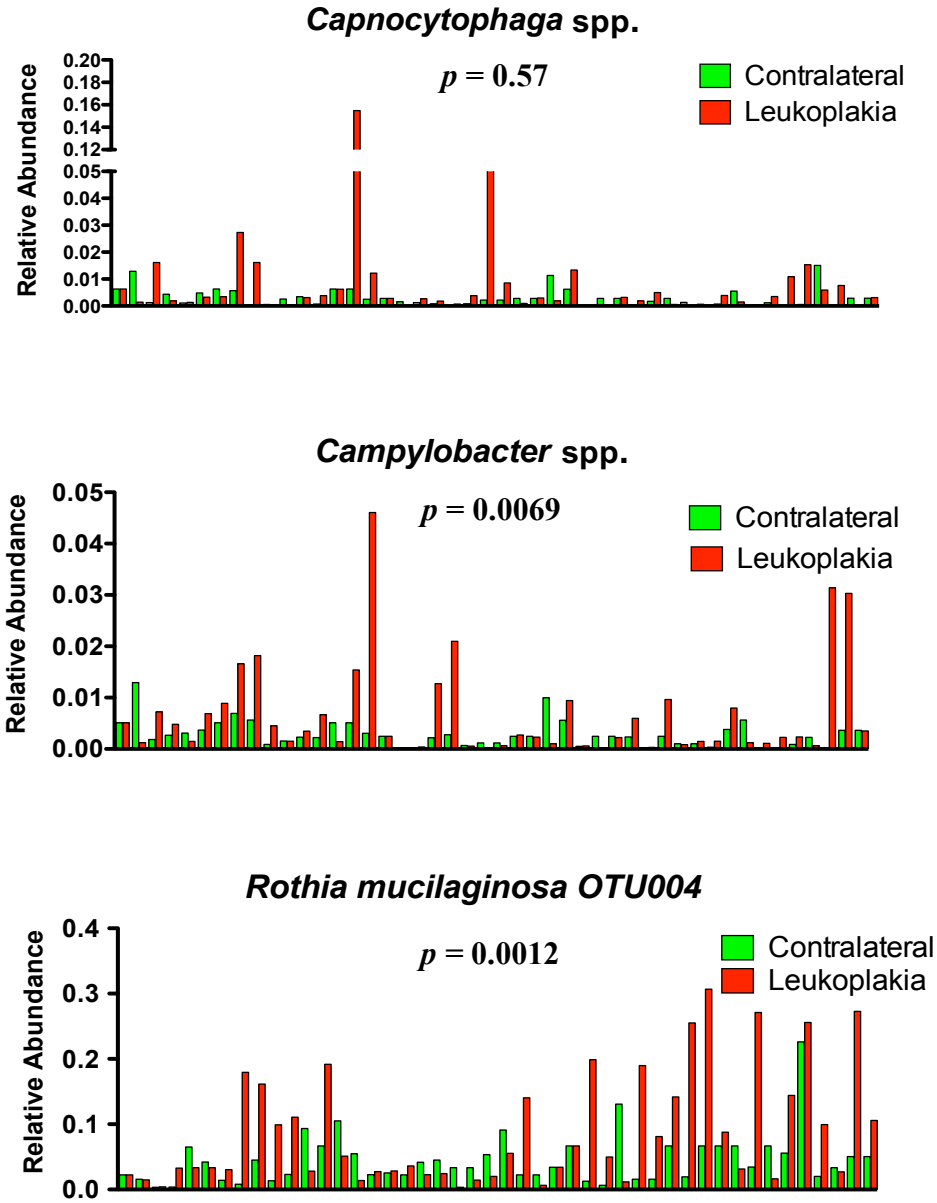



Figure 3.23 Showing the relative proportion of each species in the microbiomes of individual patients. Matched contralateral healthy and leukoplakia samples are shown side by side. The *p* value was calculated using the Wilcoxon matched-pairs test

## 3.4 Discussion

Oral cancer is the sixth most common cancer worldwide (Zini *et al.*, 2010). As with other cancers, survival depends on the stage at which the cancer is discovered, as early stage discovery offers better opportunities for treatment. Delayed diagnosis could reduce survival to less than 30% in the case of large tumors, while this could improve up to 90% in cancers that are diagnosed in the early stages. Oral leukoplakia is a potentially malignant lesion that can turn into oral squamous cell carcinoma (Ganesh *et al*., 2018), which is the most common type of oral cancer. In total, 90% of oral cancers are OSCC. Leukoplakia can transform into OSSC and the rate of transformation of leukoplakia to OSCC can be as high as 10-20%. This percentage can be affected by other risk factors including smoking, alcohol consumption, diet and immune function, all of which may play a role in the malignant transformation of oral leukoplakia. However, the role of the oral microbiome in the development of leukoplakia and its transformation to OSCC is not yet known.

To examine if any specific changes in the oral microbiome were associated with leukoplakia, this study was designed to investigate differences in the bacterial community structure by comparing leukoplakia samples with contralateral normal samples from the same patient and healthy control samples from individuals without leukoplakia. By using contralateral samples from the same patient, this study employed age, gender and oral hygiene matched control samples. It was hypothesised that contralateral samples from the same patient will be exposed to the same variables in oral health (hygiene, periodontitis) as leukoplakia samples, so it can be inferred that any differences in microbiome structure are due to differences in mucosal colonisation and no other factors.

In general, by sequence analysis of the V1-V2 region of the 16S rRNA gene, a total of 12 different bacterial phyla were observed in both buccal and lingual samples. The majority

of the bacteria in both buccal and lingual samples could be identified from the 7 phyla Firmicutes, Proteobacteria, Bacteroidetes, Fusobacteria, Actinobacteria, TM7, and Spirochaetes. Buccal and lingual samples had similar percentages of bacteria from these phyla. To investigate the differences in the structure of communities from leukoplakia and healthy sites, statistical analysis of community richness (alpha diversity) and community structure (beta diversity) was conducted.

Alpha diversity measurements of biodiversity were calculated using the inverse Simpson index in all samples. Lingual healthy control samples showed more biodiversity than the lingual patient samples according to the inverse Simpson index (Fig. 3.6). However, the inverse Simpson index comparison for all buccal samples (leukoplakia, contralateral normal & control healthy) did not reveal significant variations in biodiversity. As a second measure of alpha diversity, species richness were analysed using rarefaction curves to investigate the differences between leukoplakia and normal tissues. Rarefaction curves did not identify any significant variation in all buccal samples (Fig. 3.7a). However, rarefaction curves of lingual samples from patients (leukoplakia samples and contralateral normal samples) compared to the healthy control samples showed a significant decrease in species richness in samples from leukoplakia and contralateral sites (Fig. 3.7b). Beta diversity calculations were made to compare community structure. Using these techniques, we could identify differences in population structure between healthy control samples and leukoplakia samples in both buccal and lingual samples. Using phylogenetic approaches and NMDS, both approaches showed that the populations in buccal leukoplakia and buccal contralateral normal samples were not significantly different. However, buccal healthy control and buccal leukoplakia exhibited differences in community structure, by AMOVA using both the Jaccard index and Bray-Curtis calculators (Table 3.3). Analysis of community structure in lingual samples, showed that

lingual control and lingual leukoplakia samples had significantly different community structures by the Jaccard index (Table 3.4), in addition, the differences between lingual control and both leukoplakia and contralateral normal could be identified using the Jaccard index and the Bray-Curtis metric. This analysis supported the previous alpha diversity results, which show the buccal samples are not significantly different, while lingual samples had more variations in species richness.

Following this, LEfSe analysis was carried out to identify what species accounted for the differences in community structure. The first LEfSe test examined samples from all patients (buccal and lingual) compared to all healthy control samples (buccal and lingual). The samples from healthy controls (buccal and lingual) had greater levels of several common oral residents including *Streptococcus* spp. *Neisseria* spp. and *Lautropia* spp. These data support the observations that healthy samples have more biodiversity compared to the patient samples. Patient samples exhibited greater levels of *Alloprevotella* spp. *N. meningitidis* OTU0050 and *Leptotrichia* spp. OTU0081 (Fig. 3.20). However, these species were not found to be leukoplakia specific, and were enriched in the healthy contralateral mucosa of these patients, at least in lingual samples (Fig. 3.17). Buccal samples from patients compared to healthy controls were found to be enriched for *Campylobacter* spp. *Leptotrichia* spp., and *R. mucilaginosa* OTU0004 (Fig. 3.18).

Comparison of contralateral normal samples and leukoplakia samples from the same patients appeared to give more specific results which identified an enrichment for the family fusobacteriaceae, which includes both *Fusobacterium* spp., and *Leptotrichia* spp. (Fig. 3.19). Buccal leukoplakia samples were also enriched by *Fusobacterium* spp. with flavobacteriaceae and *Campylobacter* spp., relative to contralateral normal tissue (Fig. 3.19). In lingual leukoplakia, LEfSe analysis revealed enrichment for

*Fusobacterium* spp. and *Capnocytophaga* spp., relative to lingual contralateral (Fig. 3.21). Overall, examination of the results show that the species most associated with oral leukoplakia include *Fusobacterium* spp. *Campylobacter* spp., and *R. mucilaginosa.* Fusobacteriaceae was one of the most consistent enrichments in all leukoplakia (both buccal and lingual). These results have some similarity to previously published data on the microbiome of OSCC. *Fusobacterium* spp. were associated with OSCC, however, the same studies also showed differences with the current study. For example, *Rothia* spp., were associated with healthy control samples by Perera and Al-Hebshi, (2018) and by Schmidt *et al.* (2014) whereas in the current study *Rothia* spp. were associated more with disease samples (i.e. leukoplakia samples).

It is difficult to speculate on the reasons why some species are present at higher levels in leukoplakia compared to contralateral normal samples. It may be that some species can adhere to the leukoplakia tissues while the others are unable to adhere to the dysplastic leukoplakia layer. Additionally, the species that colonise leukoplakias may alter the environment of the site and exclude organisms normally found on healthy tissues. Although this study identified enrichment of some species on mucosa exhibiting leukoplakia, it did not directly implicate these species in the progression of leukoplakia to OSCC. However, enrichment for fusobacteria have been identified in colorectal cancers and oesophageal cancers (Kelly and Yang, 2018), and this organism has been hypothesized to be a driver of malignant transformation. Further studies are needed to determine their role, if any, in the malignant transformation of oral leukoplakia. No potential role in malignant transformation has been proposed for any of the other species identified in this study. However, *Campylobacter* spp., and *Leptotrichia* spp., have been found to be associated with *F. nucleatum* in colorectal cancers (Warren and Freeman, 2013). *R. mucilaginosa* is member of the normal oral microbiota and is a rare cause of

opportunistic infection. Also an examination of *R. mucilaginosa* by Moritani and Takeshita (2015) showed high levels of acetaldehyde produced in presence of ethanol.

# Chapter 4

Analysis of the microbiome of oral leukoplakia by sequence analysis of V3-V4 regions of the 16S rRNA gene and comparison with the V1-V2 regions.

## 4.1 Introduction

The sequences of the 16S rRNA gene are commonly used to identify bacterial species including non-cultivated species. Approximately 50% of the oral microbiome remains uncultivated and have only been identified from the 16S rRNA sequence. The 16S rRNA gene is approximately 1,500 bp long and contains nine variable regions, and these regions are separated by conserved regions. The 16S rRNA gene, in conjunction with Illumina sequencing technology can also be used to determine the relative abundance of bacteria in a sample. Thousands of bacterial strains have had their 16S genes sequenced, corresponding to approximately 700 oral species and a large database of these sequences are now available at HOMD. 16S rRNA sequencing using the Illumine MiSeq platform is a quick and cost-effective method for understanding complex communities in the microbiome.

## 4.2 Materials and methods

The same materials in Chapter 3 were used also in Chapter 4, the V3-V4 regions of the 16S rRNA gene were targeted for sequence analysis in order to compare to V1-V2 region sequencing for characterising oral communities. The V3-V4 region has been recommended by Illumina for microbiome studies using their established 16S metagenomics sequencing library preparation protocol. The goal of this experiment was to determine whether using different regions of the 16S rRNA would provide different results of bacterial species identification. It was also hoped to confirm the main findings from the analysis of the V1-V2 region in Chapter 3 using different regions of the 16S rRNA gene.

## 4.3 Results

### 4.3.1 Sequence Assembly

In order to confirm the results received from V1-V2 sequencing analysis descried in Chapter 3, the same samples were analysed following amplification of the V3-V4 regions using primers V3-V4-F and V3-V4 -R (Table 2.4). The total number of subjects analysed was 112 and 5 samples were excluded from the analysis as these samples yielded less than 1000 reads in total. The total number of sequences obtained was 9,730,294 and the number unique sequences was 692,275 following alignment of the forward and reverse reads in Mothur. Mothur was used to remove any sequences less than < 300 bp or > 400 bp. Contaminating 18S or any other eukaryotic sequences were also removed. Chimeric sequences were identified in Mothur and removed with Uchime. Rare sequences (1-2 copies) were also removed from the dataset. Following these procedures, the total number of sequences was reduced to 8,500,271, with 12,922 unique sequences.

### 4.3.2 Phylum level in all samples

Firstly, the sequences of all samples were classified to the phylum level. The pie chart showing in Figure 4.1 the percentage of reads associated with each phylum. Fifteen phyla were represented, with 5 phyla representing the majority of the reads (98%). The Firmicutes presented with almost 48%, Bacteroidetes and Proteobacteria each represented 15% in each, phylum fusobacteria and actinobacteria also presented with the same level (10% of the total reads). The remaining phyla Spirochaetes, TM7, SR1, Tenericutes, Synergistetes, Chloroflexi and Chlamydiae represented approximately 2% of the total reads.

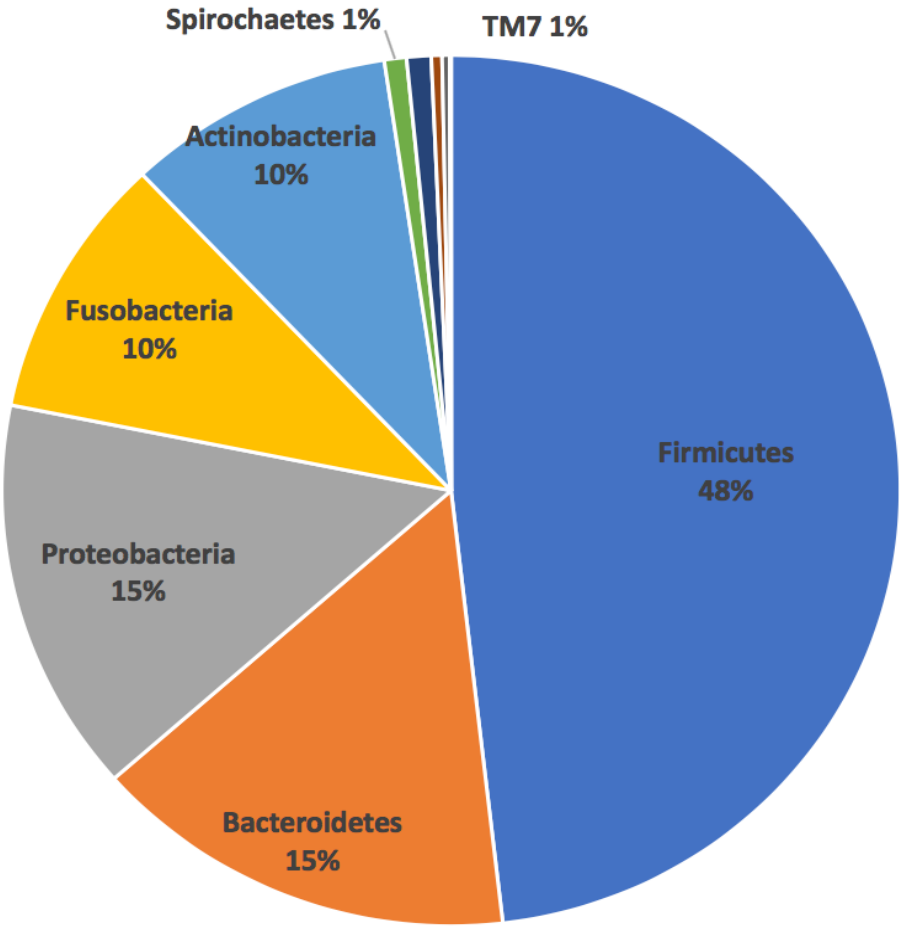



Figure 4.1 Shows the classification of sequences in all samples, buccal and lingual at phylum level.

### 4.3.3 Phylum level analysis of buccal samples

The distribution of the phyla in all buccal sample types was determined. Bacteroidetes, Fusobacteria, Actinobacteria and Spirochaetes were observed at higher levels in patients relative to controls. Phylum TM7 was more abundant in healthy controls (Fig. 4.2).

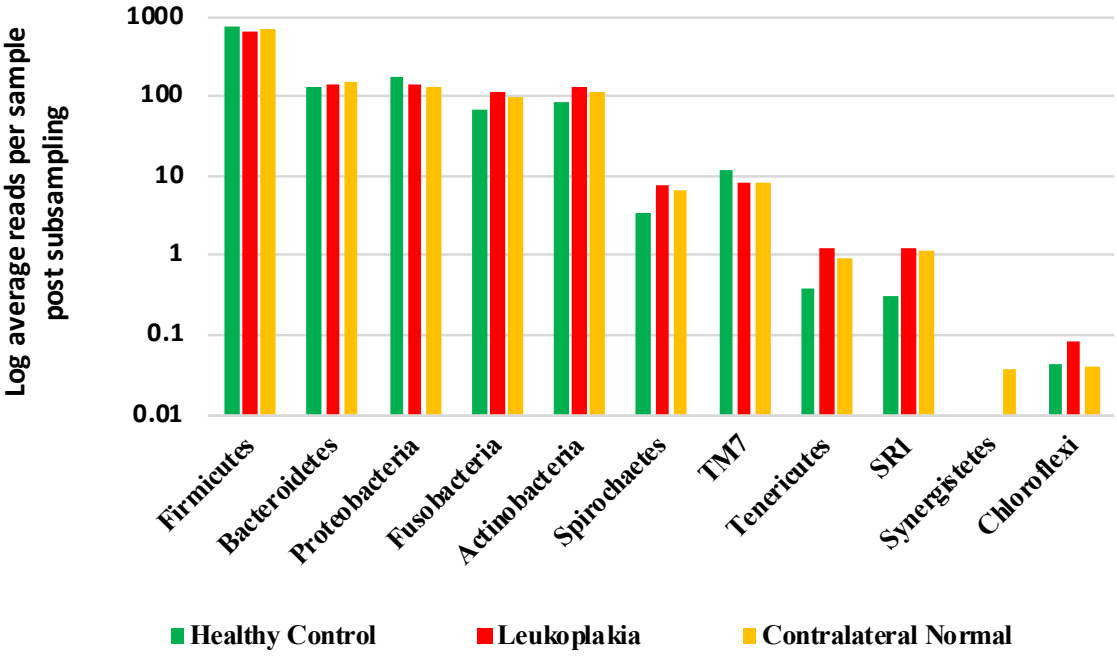



Figure 4.2 Shows the distribution of phyla in all buccal samples

### 4.3.4 Phyla levels in different lingual samples

After buccal samples, phyla distribution were examined in all lingual samples (Fig. 4.3). Overall, the phylum Firmicutes represented the majority of the sequence reads in all samples. Healthy control samples showed higher levels of Proteobacteria and Spirochaetes. Actinobacteria were found at higher levels in leukoplakia samples. Fusobacteria and Bacteroidetes were more abundant in leukoplakia relative to contralateral normal samples, but not significantly greater than healthy controls.

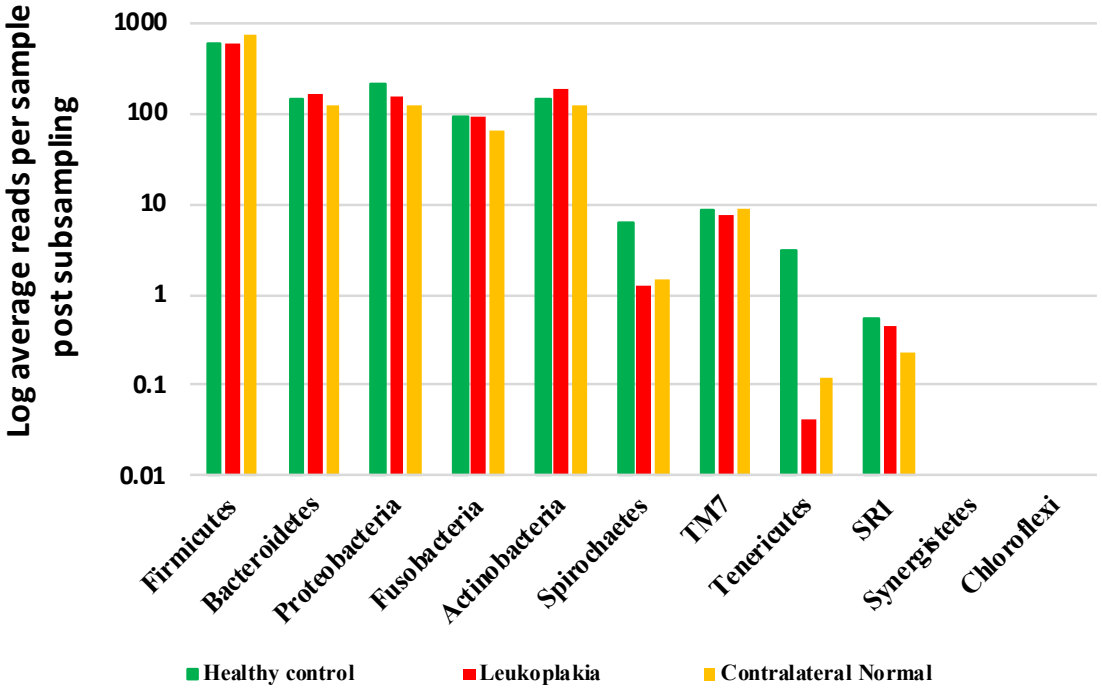



Figure 4.3 Shows the distribution of phyla in all lingual samples.

### 4.3.5 Analysis of Operating Taxonomic Units

Sequences were processed in the Mothur software package was used to classify the reads at level 2% difference in nucleotide sequence. A total of 1,512 OTUs were identified. The lowest number of the OTUs identified were in the lingual contralateral normal samples and the highest was in lingual healthy controls (Table 4.1).

Table 4.1 Table showing the OTUs ranges and average number of OTUs in all samples.

| Sample type | OTU ranges | OTUs average |
|---|---|---|
| Buccal leukoplakia samples | 50-117 | 80 |
| Buccal contralateral normal samples | 45-124 | 75 |
| Buccal healthy control samples | 37-122 | 77 |
| Lingual leukoplakia samples | 31-85 | 54 |
| Lingual contralateral normal samples | 21-117 | 61 |
| Lingual healthy control | 54-134 | 75 |

### 4.3.6 Comparison of OTU classification using V1-V2 and V3-V4 sequence analysis

The number of OTUs identified in the same samples using the V1-V2 sequence analysis described in Chapter 3 was compared to the OTUs identified by V3-V4 analysis. Table 4.2 shows the number of OTUs identified in the major oral phyla using both techniques. A significantly greater number of OTUs were identified in V1-V2 sequences that were assigned to the phyla Firmicutes, Bacteroidetes and Proteobacteria. In fact, a greater number of OTUs could be identified using V1-V2 in all phyla with the exception of Fusobacteria and Saccharibacteria (TM7), both of which yielded a higher number of OTUs using V3-V4 analysis.

**Table 4.2** Shows OTUs numbers identified in the abundant oral phyla using V1-V2 regions compared to V3-V4 regions.

| | Phylum | OTUs in V1-V2 | OTUs in V3-V4 |
|---|---|---|---|
| **1** | Firmicutes | 810 | 503 |
| **2** | Bacteroidetes | 362 | 286 |
| **3** | Proteobacteria | 316 | 254 |
| **4** | Fusobacteria | 165 | 186 |
| **5** | Actinobacteria | 243 | 161 |
| **6** | Synergistetes | 13 | 7 |
| **7** | Spirochaetes | 71 | 49 |
| **8** | Saccharibacteria | 14 | 36 |
| **9** | SR1 | 15 | 2 |

### 4.3.7 Comparison of the most abundant OTUs identified in V1-V2 and V3-V4 analysis

In order to compare the range of taxa identified by V1-V2 and V3-V4 sequence analysis, we selected the 500 most abundant OTUs identified in both datasets for comparison. Following removal of redundant species, we identified 306 unique taxa in the V1-V2 data and 277 taxa in the V3-V4 data. Comparison of these groups showed that 184 taxa were common in both, but the V1-V2 analysis identified 122 unique taxa among the 500 most abundant OTUs whereas only 93 unique taxa were identified among the most abundant OTUs in V3-V4 (Fig. 4.4).

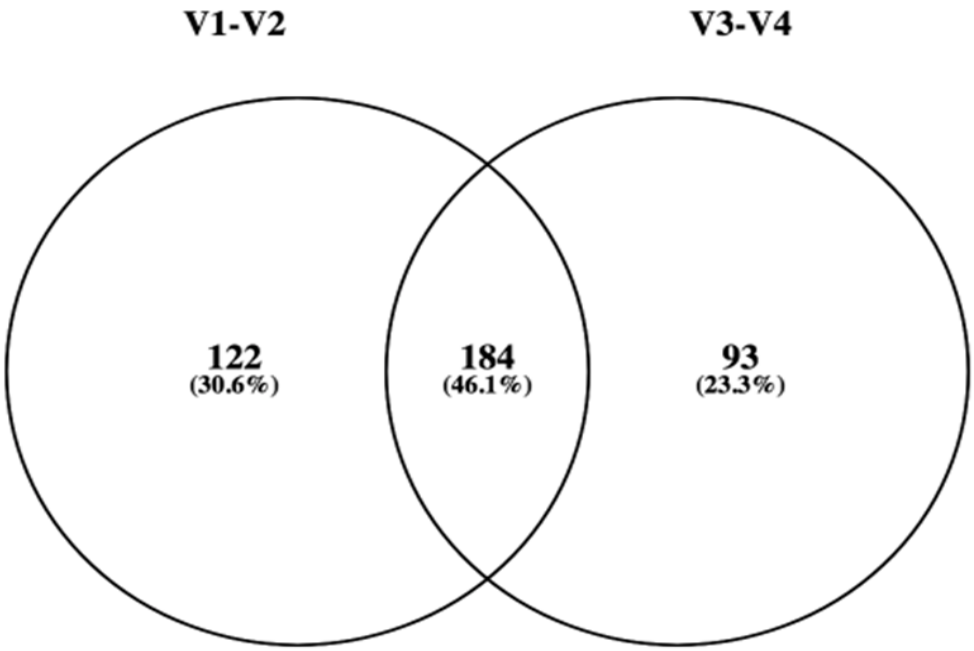



Figure 4.4 Venn diagram showing the number of common and unique taxa identified in the 500 most common OTUs identified by V1-V2 analysis and V3-V4 analysis.

## 4.4 Alpha diversity measurements

### 4.4.1 Inverse Simpson index

The Inverse Simpson index was generated for all sample types. Overall, lingual healthy control samples had the highest levels of biodiversity (Fig. 4.5). Lingual control had higher levels of biodiversity compared to lingual contralateral and leukoplakia. Buccal leukoplakias had higher levels of biodiversity relative to buccal healthy and contralateral from patients.

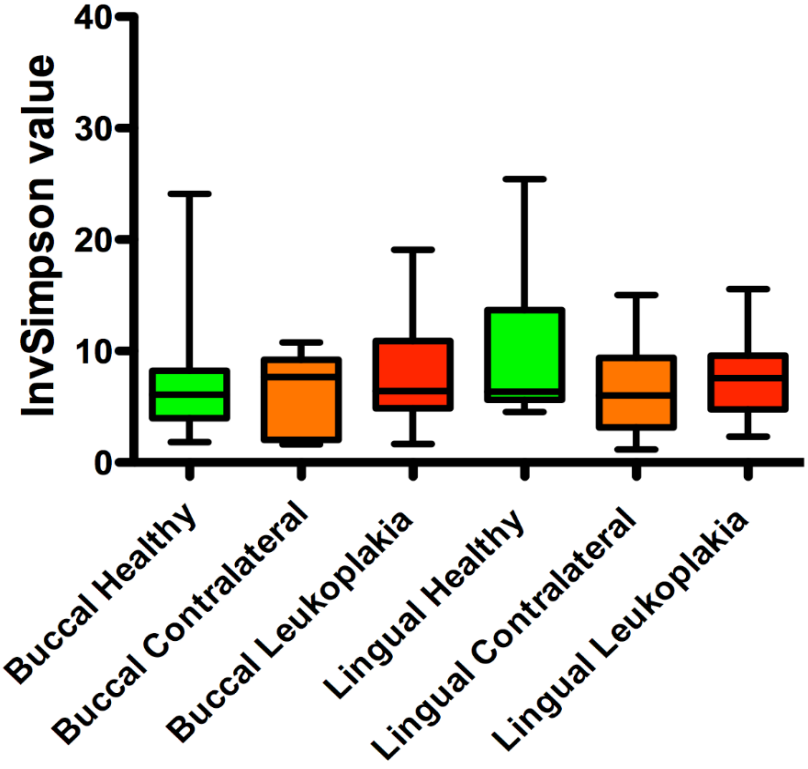



Figure 4.5 Showing the average of the inverse Simpson index in all samples, buccal and lingual.

### 4.4.2 Rarefaction curve

Analysis of the rarefaction curve was carried out in the buccal samples to measure the species richness in each sample type. In general, buccal leukoplakia samples had greater species richness relative to buccal contralateral and healthy control samples (Fig. 4.6). However, with lingual samples, healthy controls had greater numbers of OTUs relative to samples from patients (Fig. 4.7).

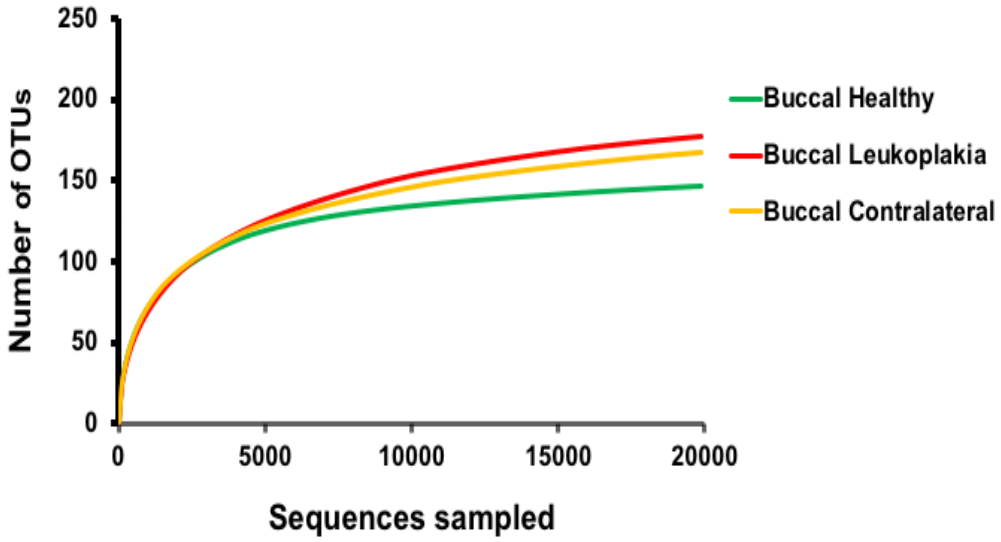



Figure 4.6 Rarefaction curves showing species richness in buccal samples.

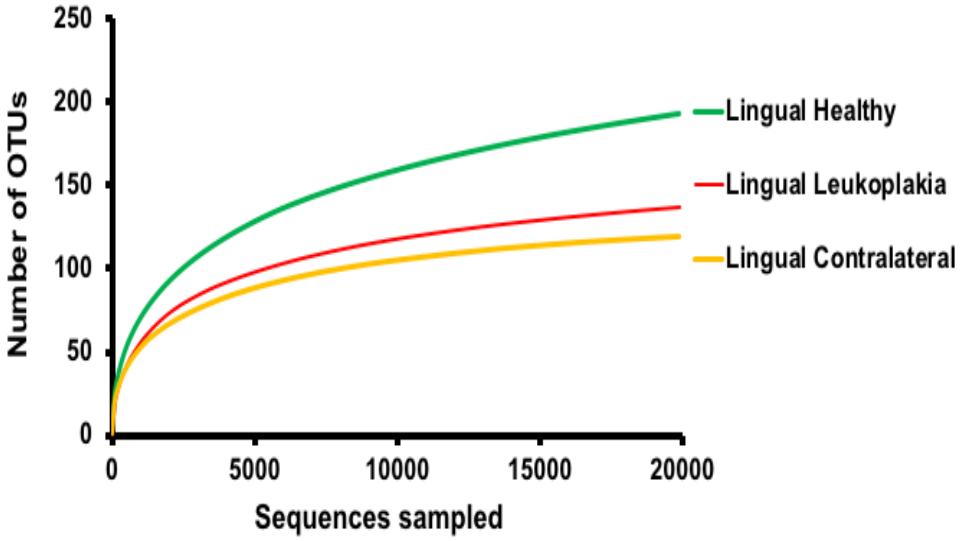



Figure 4.7 Rarefaction curves showing species richness in lingual samples

## 4.5 Beta diversity

Beta diversity also was analysed to measure the difference in microbial community structure in different samples. The Jaccard index and Bray-Curtis dissimilarity index were used to determine the similarity of the communities in each sample.

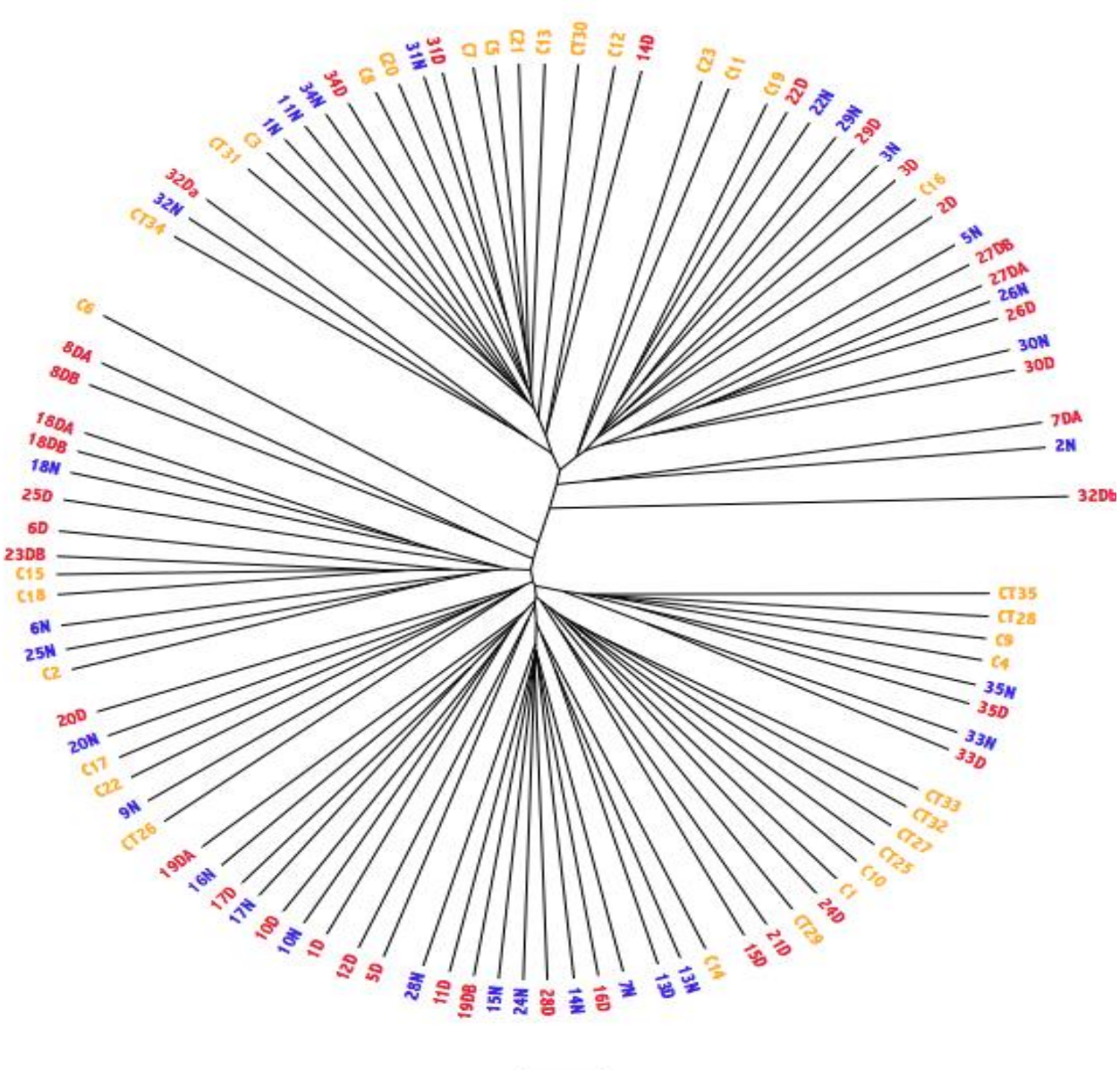
### 4.5.1 Phylogenetic analysis

Phylogenetic trees were generated using distance matrices generated using Jaccard coefficient and Bray-Curtis dissimilarity values. Parsimony showed that buccal samples were not significantly different from each other. In figures 4.8 and 4.9 show samples clustering according to the similarity, Jaccard index coefficient shows lingual healthy control samples exhibited differences in structure compared to samples from patients. Lingual leukoplakia samples were significantly different from healthy control samples ($p = 0.024$) but not contralateral normal samples ($p = 0.07$) (Table 4.3).

Table 4.3 Shows a summary of *p* values from Parsimony analysis of phylogenetic trees generated with Jaccard and Bray-Curtis distance values in all buccal and lingual samples. The indicated sample pairs were compared in each test.

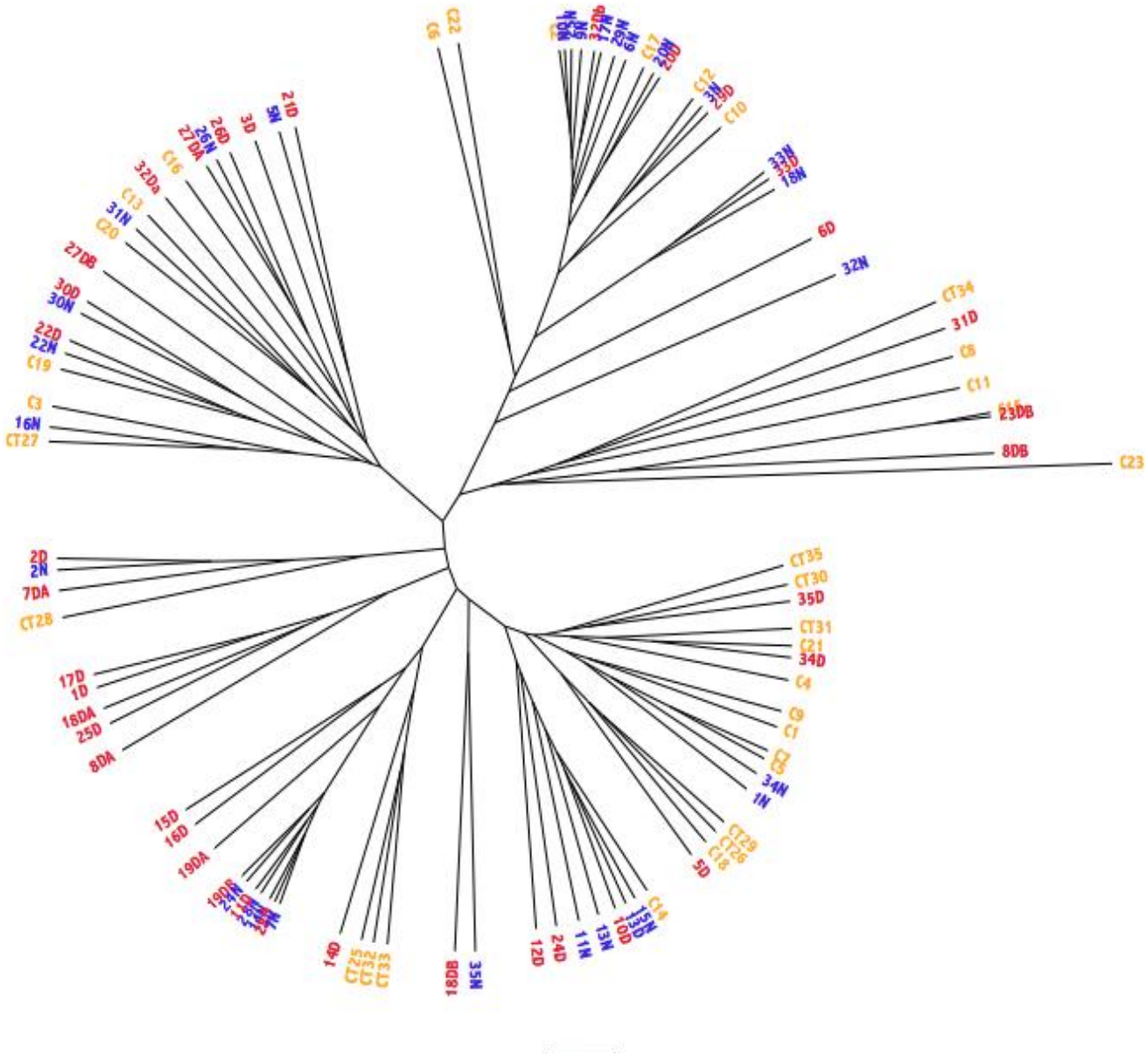
| **Parsimony test: Jaccard index** | ***p* values** | | **Parsimony test: Bray-Curtis index** | ***p* values** |
|---|---|---|---|---|
| BC-BD | $p = 0.238$ | | BC-BD | $p = 0.568$ |
| BC-BN | $P = 0.293$ | | BC-BN | $p = 0.772$ |
| BD-BN | $P = 0.944$ | | BD-BN | $p = 0.76$ |
| LC-LD | $P = 0.024$* | | LC-LD | $p = 0.079$ |
| LC-LN | $P = 0.07$ | | LC-LN | $p = 0.243$ |
| LD-LN | $P = 0.777$ | | LD-LN | $p = 0.393$ |

BC=Buccal controls BD=buccal leukoplakia BN=buccal contralateral normal
LC=Lingual controls LD=lingual leukoplakia LN=lingual contralateral normal
(*) = $p < 0.05$




Leukoplakia disease Contralateral normal Healthy sample control

Figure 4.8 Shows a phylogenetic tree generated using a distance matrix calculated by Jaccard index similarity coefficient ($p = 0.05$).




Leukoplakia disease Contralateral normal Healthy sample control

Figure 4.9 Shows a phylogenetic tree generated using Bray-Curtis dissimilarity index ($p = 0.05$).

### 4.5.2 Non-metric Multidimensional scaling (NMDS)

In order to generate three-dimensional NMDS plots, the Jaccard and Bray-Curtis distance matrices were employed. For this analysis, we carried out a separate analysis of buccal and lingual samples.

### 4.5.3 Buccal samples

Figures 4.10 and 4.11 show NMDS plots generated in Mothur using the Jaccard coefficient and Bray-Curtis dissimilarity distance matrices respectively. AMOVA was used to determine whether the differences in the population structure were significant. Analysis of the Jaccard plot indicated significant differences in population membership between buccal communities from healthy controls and patients (Table 4.4). the Population structure of buccal communities was not considered significantly difference between leukoplakia and contralateral samples.

**Table 4.4** Results of AMOVA analysis of distance matrices generated with the Jaccard coefficient and the Bray-Curtis dissimilarity index from buccal samples. The indicated sample groups were compared in each test.

| **Jaccard index: AMOVA** | ***p* values** | | **Bray-Curtis index: AMOVA** | ***p* values** |
|---|---|---|---|---|
| BC-BD-BN | $p = 0.307$ | | BC-BD-BN | $\boldsymbol{p} = 0.421$ |
| BC-BP | $P = 0.0469$ | | BC-BP | $p = 0.222$ |

BN=Buccal contralateral normal BP= Leukoplakia & contralateral

BC=Buccal healthy control BD= Buccal leukoplakia

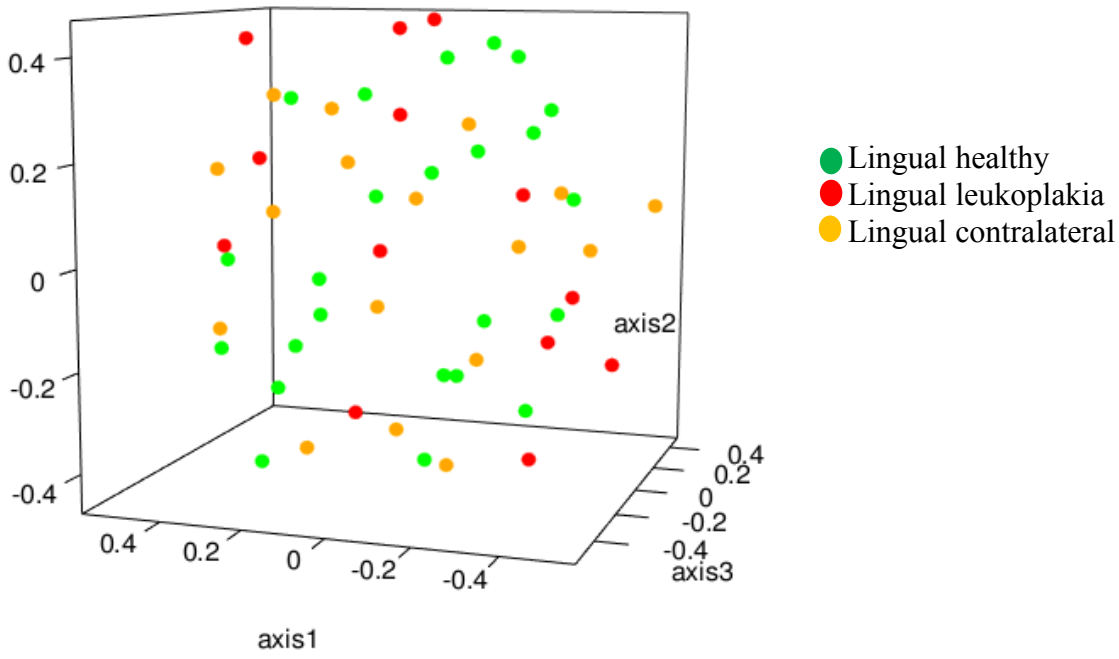



Figure 4.10 Showing NMDS plot of all buccal microbial populations generated using the Jaccard index matrix.

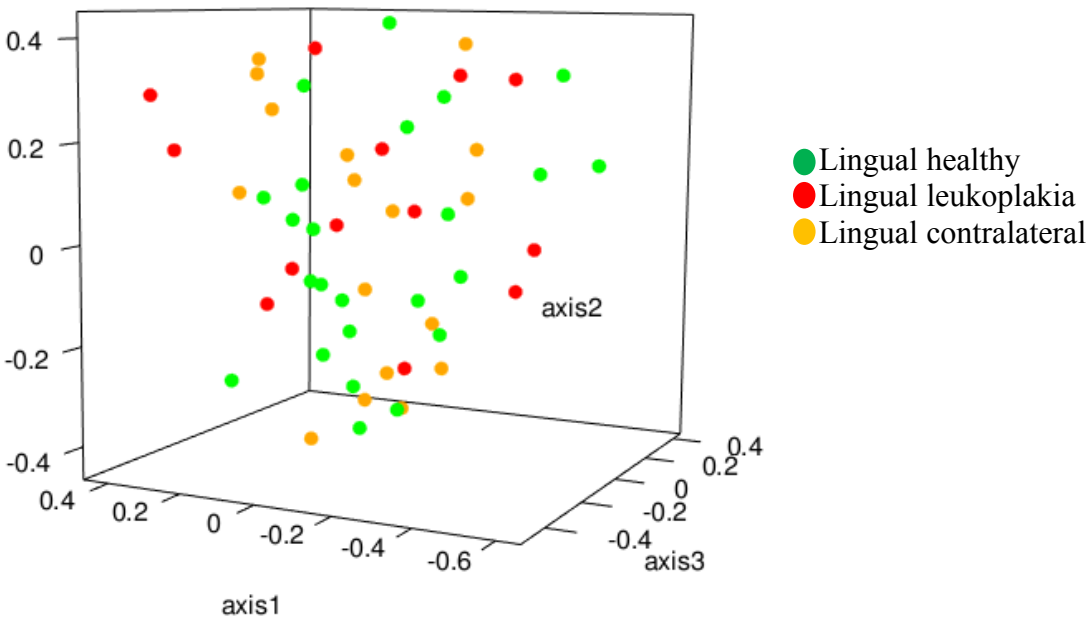



Figure 4.11 Shows NMDS plots of buccal microbial populations generated using the Bray-Curtis dissimilarity index matrix

### 4.5.4 Lingual samples

The Jaccard index and Bray-Curtis distance values were also used to generate NMDS plots of all lingual samples (Figs 4.12 and 4.13). AMOVA was used to determine if communities from healthy controls, leukoplakia and contralateral normal samples could be differentiated.

Analysis of the Jaccard coefficient data showed that healthy control samples were significantly different compared to lingual leukoplakia samples ($p < 0.014$). Also, lingual healthy control samples were significantly different compared to lingual contralateral normal samples ($p < 0.001$). Bray Curtis also showed lingual healthy controls and lingual contralateral normal samples were different ($p = 0.027$).

**Table 4.5** Shows the results of AMOVA analysis of distance matrices generated with Jaccard coefficient and Bray-Curtis dissimilarity index in lingual samples. The indicated sample pairs were compared in each test.

| **Jaccard index: AMOVA** | ***p* values** | | **Bray-Curtis index: AMOVA** | ***p* values** |
|---|---|---|---|---|
| **LC- LD - LN** | $p = 0.011$* | | LC - LD - LN | $p = 0.028$* |
| **LC - LD** | $p < 0.014$* | | LC - LD | $P = 0.242$ |
| **LC - LN** | $p < 0.001$* | | LC - LN | $p = 0.027$* |
| **LD - LN** | $p = 0.592$ | | LD - LN | $p = 0.074$ |

LC= lingual control LD= Lingual leukoplakia LN=lingual contralateral normal

(*) = $p < 0.05$

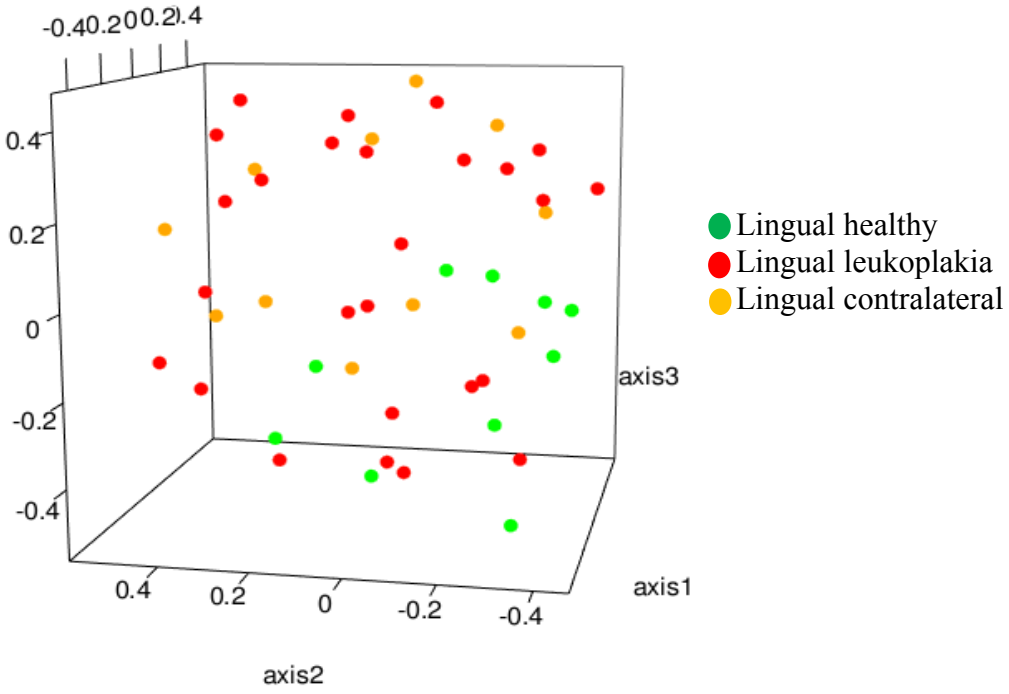



Figure 4.12 Showing NMDS plot of lingual microbial populations generated using Jaccard index matrix.

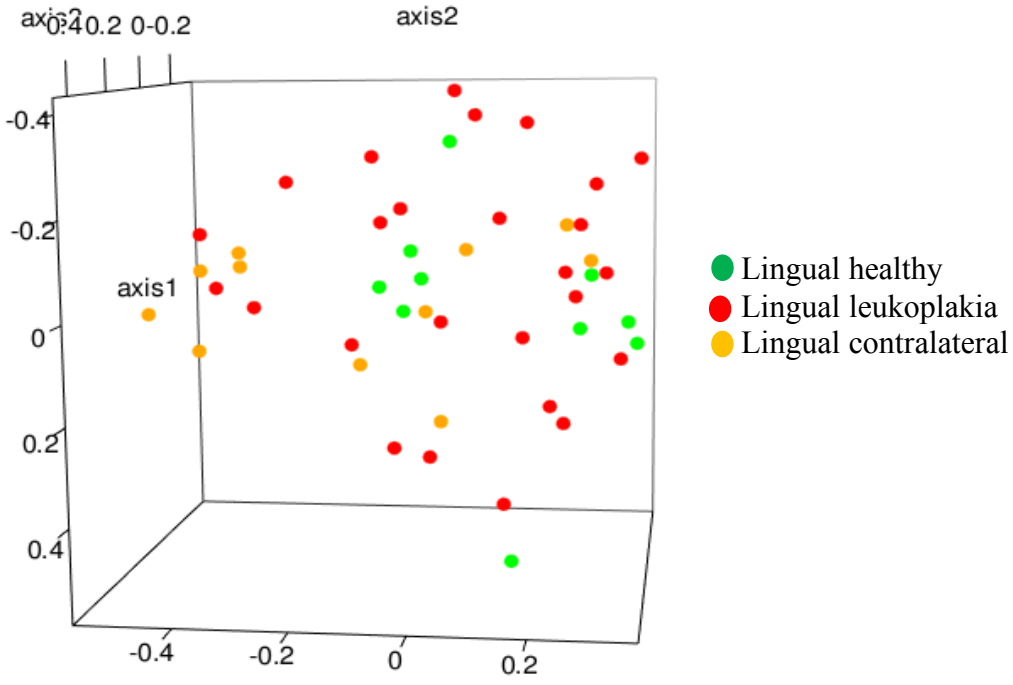



Figure 4.13 Shows NMDS plot of lingual microbial populations generated using the Bray-Curtis dissimilarity index matrix.

## 4.6 Linear discriminant analysis effect size (LEfSe)

LEfSe was applied to analyse the V3-V4 data set to identify any significant enrichments in specific species or OTUs in patients samples (contralateral normal and leukoplakia) compared to healthy control samples. The identity of the OTUs with significant enrichments in the following sections were confirmed by BLAST searches, using the OTU consensus sequence in the HOMD. Analysis of all patient samples (leukoplakia and contralateral) versus all healthy control samples revealed that healthy controls exhibited enrichment for *Neisseria flava* OTU0017, *Veillonella* OT780 (OTU0046), and *Lautropia mirabilis* OTU0049, among others (Fig. 4.14).

A comparison of all buccal and lingual contralateral to all buccal and lingual leukoplakia identified one OTU enrichment in each type of the samples, which was found in contralateral samples, corresponding to streptococcal OTU0160.

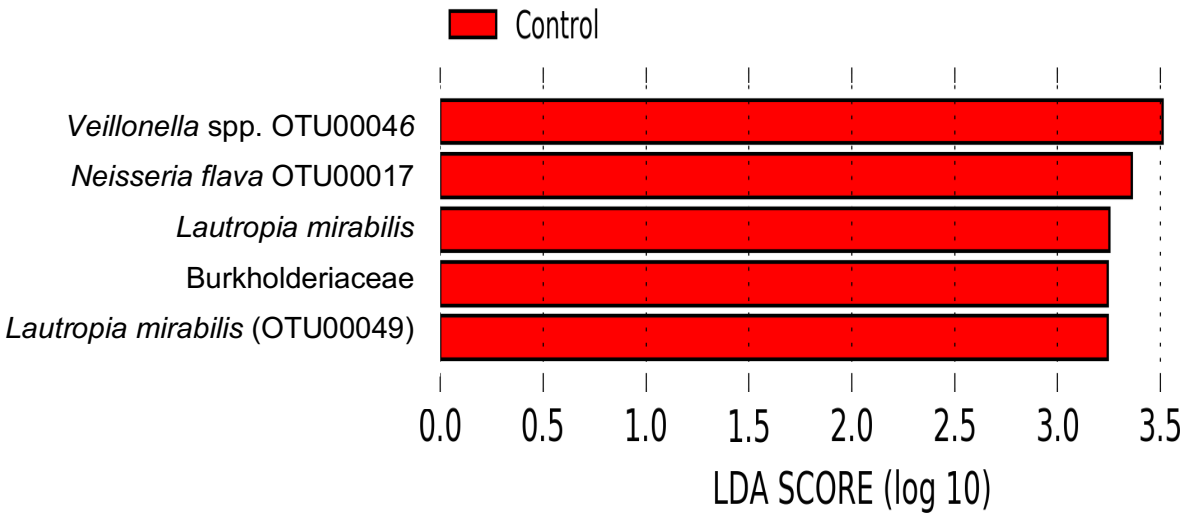



Figure 4.14 Shows LEfSe of V3-V4 sequencing data from all controls compared to all patient samples (leukoplakia and contralateral). All categories shown were significant ($p < 0.05$, LDA > 3.0

### 4.6.1 LEfSe analysis of V3-V4 data from buccal samples

Analysis with LEfSe identified 7 taxonomic groups enriched in the buccal samples from patients (leukoplakia and contralateral) relative to healthy controls, and showed that *R. mucilaginosa* OTU0002, *F. nucleatum* OTU0005, *Campylobacter* spp., and others were overrepresented in patient samples (Fig. 4.15).

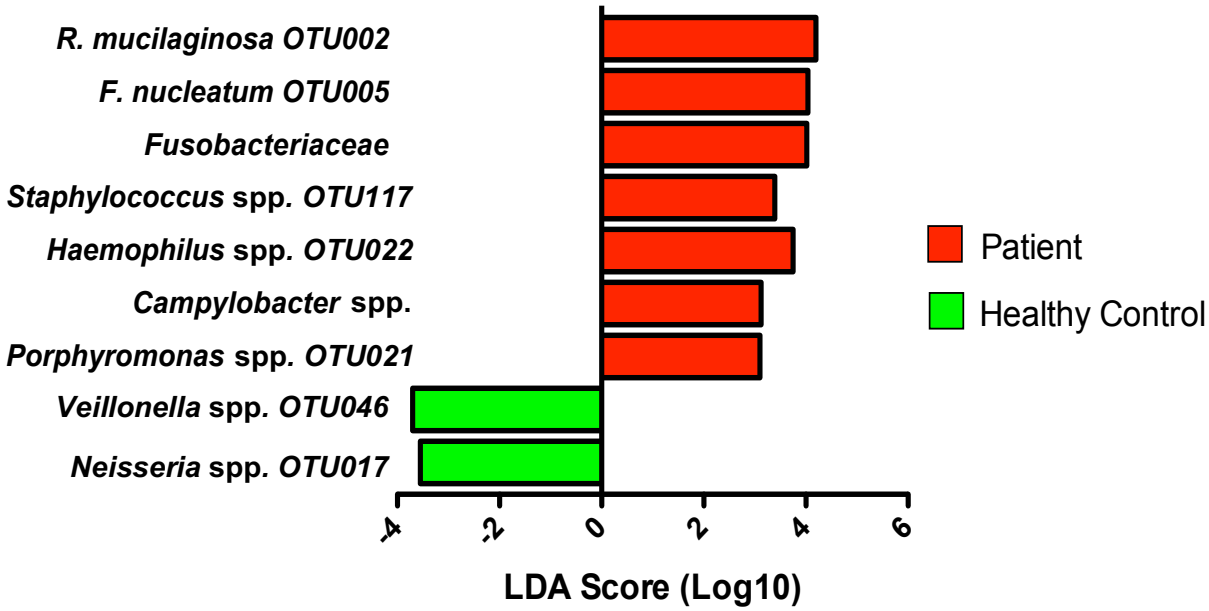



Figure 4.15 Shows LEfSe analysis of V3-V4 sequencing data from all buccal healthy controls compared to all buccal patient samples (leukoplakia and contralateral). All categories shown were significant ($p < 0.05$, LDA > 3.0).

Buccal leukoplakia samples and buccal contralateral normal were they compared. This analysis showed two OTUs enrichments; *Haemophilus parahaemolyticus* OTU0044 in contralateral normal samples and *Veillonella* spp. OT780 OTU0046 in leukoplakia samples (Fig 4.16).

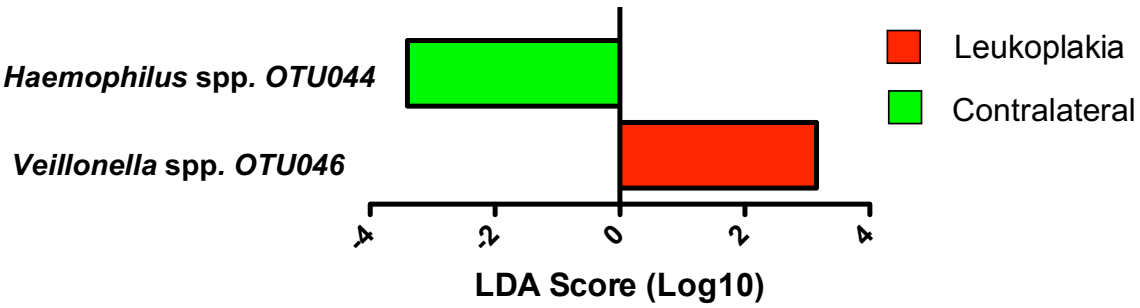



**Figure 4.16** Showing LEfSe analysis of V3-V4 sequencing data from all buccal leukoplakia compared to all buccal contralateral. All categories shown were significant ($p < 0.05$, LDA > 3.0).

## 4.6.2 LEfSe analysis of V3-V4 data from lingual samples

The LEfSe analysis of the lingual patient samples (leukoplakia and contralateral) compared to lingual healthy controls showed that healthy controls were enriched with a large number of species relative to the patient samples and included of Gram-positive and Gram-negative bacteria (Fig. 4.17).

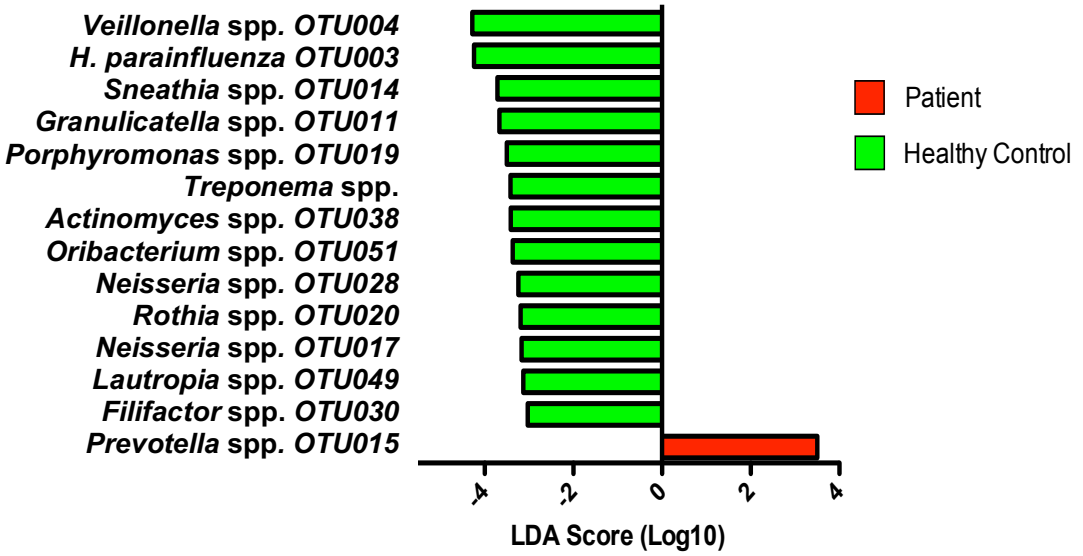



Figure 4.17 Shows LEfSe analysis of V3-V4-16S rRNA sequencing of all healthy control samples compared to all lingual patient samples (leukoplakia and contralateral normal). All categories shown were significant ($p < 0.05$, LDA > 3.0).

### 4.6.3 LEfSe analysis of lingual contralateral and leukoplakia samples

The LEfSe analysis showed lingual leukoplakia samples were enriched with Fusobacteria, *F. nucleatum* OTU0005 accounting for a proportion of this enrichment. The family Porphyromonadaceae were also found at a high level in leukoplakia samples relative to contralateral samples. However, *streptococci* and others others were enriched in contralateral normal samples, (Fig. 4.18).

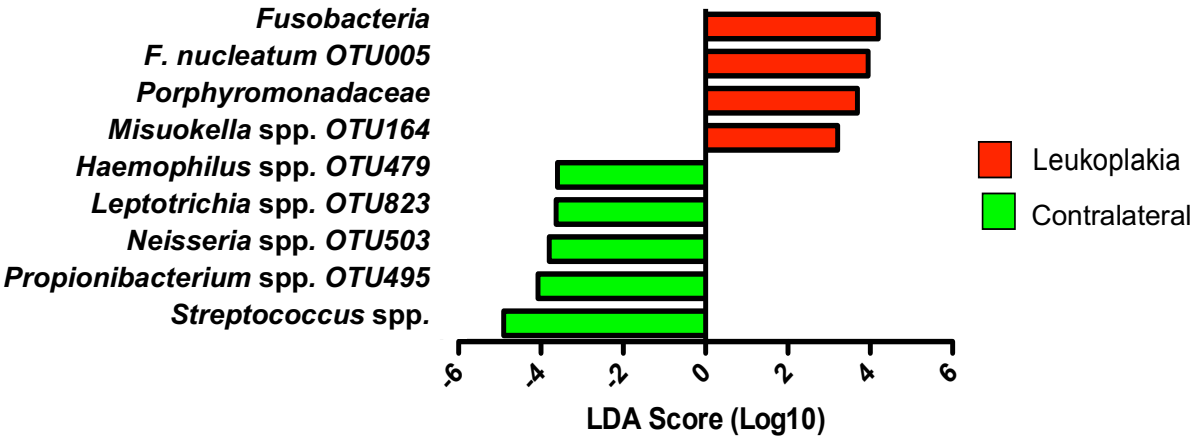



Figure 4.18 LEfSe analysis of V3-V4 sequencing data comparing lingual patient samples (leukoplakia and contralateral). All categories shown were significant ($p < 0.05$, LDA > 3.0).

## 4.7 Discussion

In order to confirm that the analysis was undertaken with the samples described in Chapter 3, we carried out amplifications and sequence analysis of the V3-V4 regions using the Illumina MiSeq. The total number of sequences generated in V1-V2 was approximately 15,000,000 leading to the identification of 96,709 unique sequences following processing using Mothur software. However, the V3-V4 yielded fewer sequences at 8,500,271, and number of unique sequences identified was 12,922. Despite yielding only 50% fewer reads, the V3-V4 analysis generated ~7 times less unique sequences compared to V1-V2 resulting. This indicates that the V3-V4 regions may be more highly conserved than V1-V2 in, resulting in less unique sequences for bacterial identification.

The proportions of the reads classified to each phylum were similar from both V1-V2 and V3-V4. However, V1-V2 analysis showed Firmicutes accounted for 58% of the total reads in contrast to 48% in the V3-V4 reads (Figs. 3.1 and 4.1). Within the Firmicutes, a total number of 810 OTUs were identified in V1-V2, and 503 OTUs in V3-V4 (Table 4.2). Conversely, V3-V4 yielded slightly more OTUs assigned to the phylum Fusobacteria than V1-V2 (186 verses 165). In general, V1V2 generates a greater diversity of sequences and would seems to be superior than V3-V4 for oral microbiome studies.

In terms of the analysis, in buccal samples, Firmicutes exhibited lower levels in leukoplakia disease samples compared to contralateral normal and control samples for both the V1-V2 (Fig. 3.3) and V3-V4 analyse (Fig. 4.2). However, the difference was small and the rarefaction curves (Figs. 3.7 and 4.6) showed that all buccal samples were comparable in terms of species richness using the V1-V2 and V3-V4 regions. In both V1-V2 and V3-V4 analysis the inverse Simpson index also agreed that buccal control samples had similar biodiversity relative to leukoplakia and contralateral samples.

In contrast, analysis of biodiversity using the Inverse Simpson index showed that lingual leukoplakia disease samples had decreased biodiversity compared to lingual healthy controls by both regions V1-V2 (Fig. 3.6) and V3-V4 (Fig. 4.5).

The rarefaction curves also showed reduced species richness in lingual leukoplakia samples in both analyses. Therefore, analysis of both rRNA gene regions concurs that leukoplakia disease samples were less rich and biodiverse relative to healthy controls samples.

Analysis of population structure using the Jaccard coefficient and Bray-Curtis dissimilarity index yielded broadly similar results for both V1-V2 and V3-V4 sequence analysis. NMDS analysis using the Jaccard coefficient showed that buccal populations from healthy control were different to the patient samples in both V1-V2 ($p = 0.014$), and V3-V4 ($p = 0.046$). However, buccal leukoplakia and contralateral samples were not distinguishable. Analysis of lingual samples showed populations from healthy control patients were different to both leukoplakia and contralateral tissue from patients using the Jaccard coefficient (Tables 3.4 and 4.8).

In terms of species enrichment, in V1-V2 and V3-V4 concurred that *Fusobacterium* spp., were one of the most significant enrichment in leukoplakia disease compared to contralateral samples (Figs 3.17 and 4.16). The V1-V2 analysis also showed that *R. mucilaginosa* OTU0004 was abundant in buccal patient samples compared to healthy controls (Fig 3.18, and the same result was observed with *R. mucilaginosa* OTU0002 in V3-V4 (Fig. 4.15).

In conclusion, analysis of the V1-V2 and V3-V4 sequencing yield qualitatively similar results, but important differences in the quantity of the certain taxa exist, in particular those belonging to the Firmicutes. The analysis reveals that the Fusobacteria family is one the most widespread and significantly enriched species in oral leukoplakia.

In conclusion, in terms of recommending V1-V2 or V3-V4 for future studies, the former yielded a greater number of OTUs which may be biologically significant. However, as long-read sequencing technology improves, it is likely that future studies will utilise the full length of the 16S rRNA sequences for bacterial identification, and the debate over which regions to sequence will be become redundant.

# Chapter 5

Analysis of the impact of *Candida* spp., smoking, alcohol consumption and denture wearing on the oral microbiome in oral leukoplakia subjects.

## 5.1 Introduction

*Candida albicans* is a normal commensal organism (Herwald and Kumamoto, 2014), and is the most common fungus found in the healthy oral microbiome. There are many factors which can cause it to become a parasitic organism, depending on the condition of the host immunity. Oral candidiasis (OC) is a common condition which occurs in the oral cavity when the host is immunosuppressed and can occur as a manifestation of malignant disease or systemic illnesses. Infection is common in HIV-infected patients, patients taking broad spectrum antibacterial agents and in elderly patients, especially those with dentures. *Candida albicans* is commonly isolated from oral leukoplakia (Abdulrahim *et al.,* 2013), however, it is still debatable whether it is a true etiological factor or whether it occurs as a secondary infection. *Candida albicans* is one of approximately 200 *Candida* species (Spampinato and Leonardi, 2013), and accounts for the majority of yeasts recovered from the oral cavity. Other *Candida* species are less commonly recovered from the oral cavity and include the closely related species *C. dubliniensis* and *C. tropicalis.*

*Candida albicans* is the most common fungal pathogen in the oral cavity and OC can present in several different ways including oral thrush, erythematous candidiasis and angular chelitis. It is thought that the primary reason for *C. albicans* infection to occur in conjunction with other conditions is because the fungus can only penetrate and damage the host tissues when immunity is weak, for example during HIV-infection or during cancer chemotherapy. This interaction is dependent on the virulence factors of the fungus as well as the host tissue defences. *Candida* spp., can invade the upper layers of oral keratinocytes and secrete degradative enzymes to break down connective tissues (Mohd *et al*. 2010). It has been proposed that the fungus attaches and invades the oral keratinocytes via the E-cadherin pathway and stimulation of this receptor can induce endocytosis by the oral cells.

A study by (Abdulrahim *et al.,* 2013), which analysed 78 samples of oral leukoplakia, showed that 55 out of 78 oral specimens were positive for *Candida* spp.

While *C. albicans* was the most abundant species present in all 55 samples, 7 samples contained an additional *Candida* species including *C. dubliniensis, C. parapsilosis, C. guilliermondii, C. tropicalis* or *C. krusei. Candida albicans* has been shown to interact with various oral bacteria including *S. gordonii* and *F. nucleatum*, however the effects of fungal colonisation on the oral microbiome is relatively unknown. Smoking is the use of tobacco by inhalation and results in the absorption of nicotine and other substances into the bloodstream. The first point of contact of the tobacco is the oral cavity and because of this, the effect on the oral microbiome was investigated in this study. Tobacco contains a variety of substances which interact with the oral mucosa. Agents such as aromatic hydrocarbons, nitrosamines and benzopyrenes are thought to act on the keratinocytes and the oral stem cells. This induces hyperkeratinisation and the formation of harmful reactive oxygen species (ROS) which likely contribute to the development of oral leukoplakia (Stich and Anders, 1989), these changes in the oral mucosa can contribute to variation in the oral microbiome which could lead to further disease. Recent evidence has shown that smoking affects the composition of oral biofilms (Kumar *et al.,* 2011). Bacterial acquisition and colonisation may be affected, resulting in microbiome changes that lead to disease. For example, cigarette smoking is a common risk factor for the development of periodontal disease. Toxicants from the tobacco can disturb the microbiome by depriving cells of oxygen and loss of beneficial bacterial species which leads to colonisation with pathogenic organisms. Previous studies have shown alterations in the abundance of some oral bacteria due to smoking and this forms the basis of investigation in this study. In this study, the level of *Candida* spp., colonisation was investigated in order to identify whether the presence of *Candida* spp., affected the oral microbiome in

oral leukoplakia. In addition, the effect of both smoking, alcohol consumption and denture wearing were also investigated.

## 5.2 Material and methods

### 5.2.1 Quantitative Real-Time PCR (qPCR)

Real Time PCR was carried out to estimate *Candida* levels using the ITSF and ITSR primers (Table 2.4) targeting the *Candida* ITS2 rDNA gene. *Candida* ITS rDNA was amplified using fast Sybr-green master mix (Applied Biosystems) and sequence detection was performed on the ABI 7500 Real-Time PCR (Applied Biosystems). *Candida* levels for all samples (CFU/ml) were estimated using a standard curve as described previously in Chapter 2.

## 5.3 Results

### 5.3.1 Detection of *Candida albicans* using Real-Time PCR

The influence of *Candida* spp., on the oral microbiome was investigated in this study.Real-time PCR was used to detect the presence of *Candida* spp., in oral leukoplakia, contralateral tissue and samples from healthy control. Interpolation of these data using the standard curve showed that a greater proportion of oral leukoplakia samples were positive for carriage of *Candida* spp. (35%) compared to contralateral normal (20%) and healthy control (11%) (Fig. 5.1).

The abundance was not determined, it was estimated of *Candida* spp. (CFU/ml) in all samples was determined buccal sites (Fig. 5.2a) were less positive than the lingual sites (Fig. 5.2b). Overall, *Candida* spp., carriage ranged from ~$3x10^2$ to ~$1x10^5$ CFU/ml. All positive samples has at least ~$3x10^2$ CFU/ml detectable by the PCR assay.

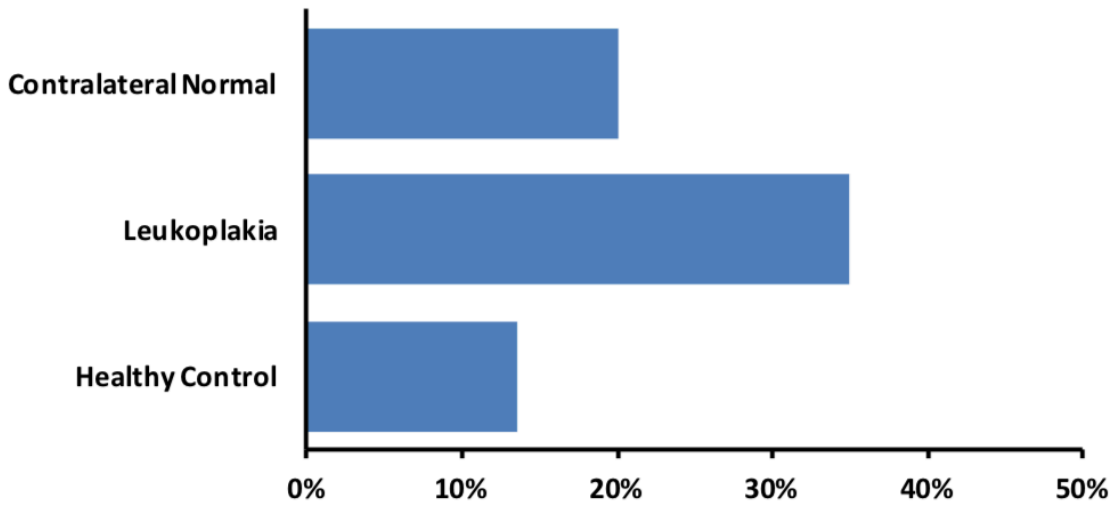
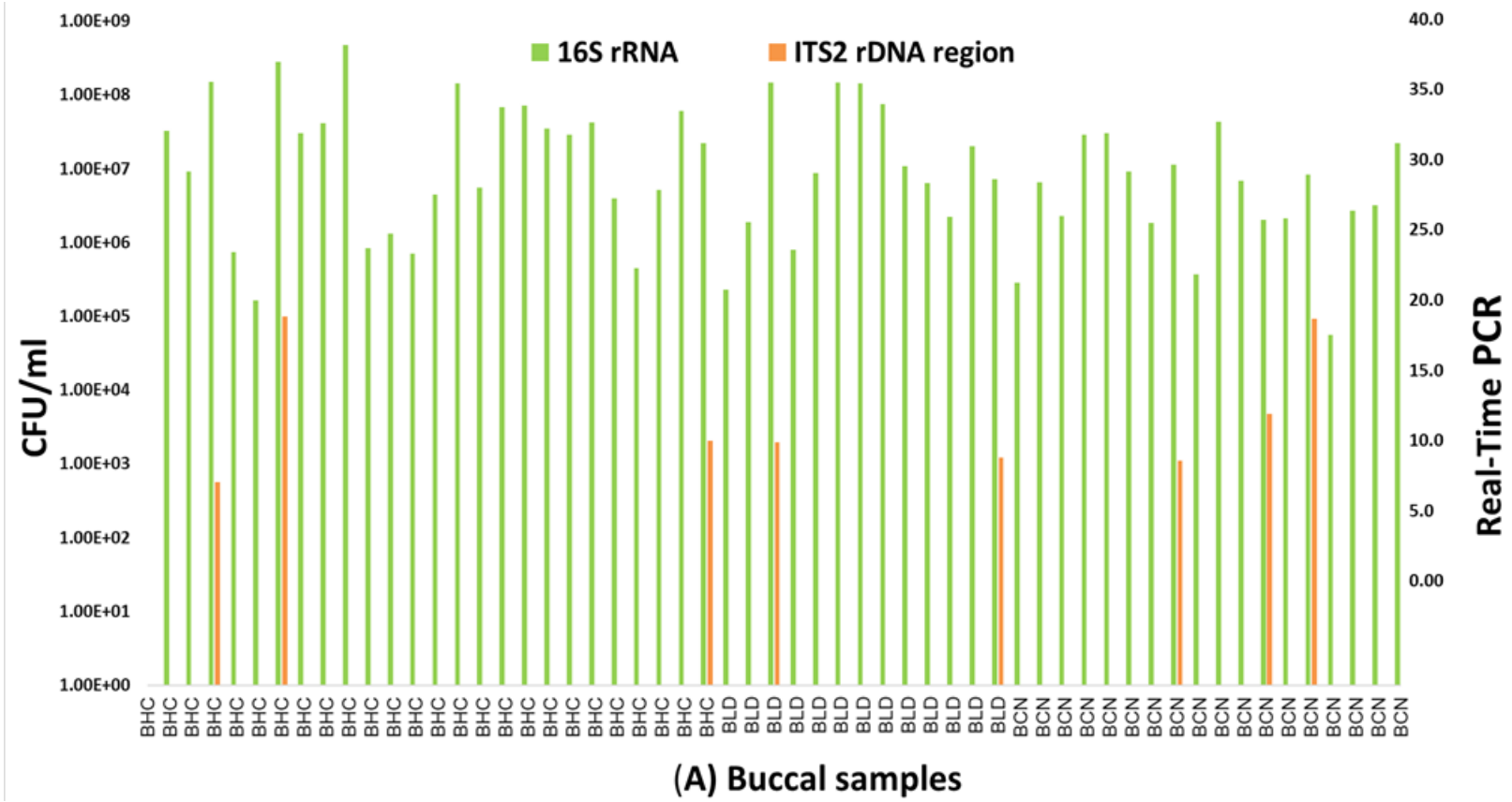
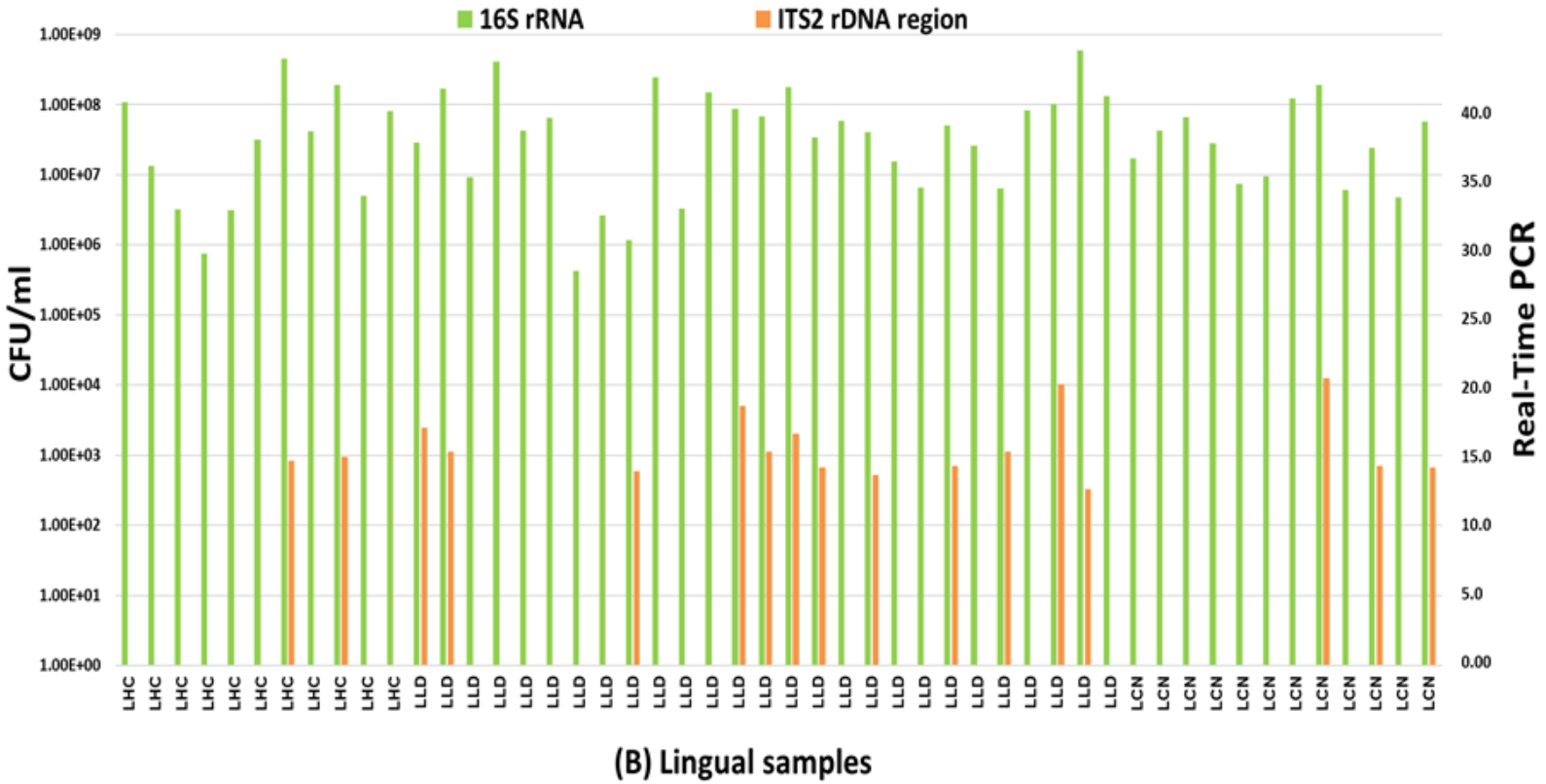



Figure 5.1 *Candida* spp. positive samples recovered from patients (leukoplakia and contralateral normal) and healthy controls, as determined by real-time PCR.




BHC = buccal healthy BLD= buccal leukoplakia BCN= buccal contralateral normal

Figure 5.2a. Shows real-time PCR quantification of bacterial DNA and *Candida* spp. DNA in (A) buccal samples and (B) lingual samples. Levels were quantified using a standard curve of bacterial and *Candida* DNA and expressed as equivalent CFU/ml




LHC = lingual healthy LLD= lingual leukoplakia LCN= lingual contralateral normal

Figure 5.2b. Shows real-time PCR quantification of bacterial DNA and *Candida* spp. DNA in (A) buccal samples and (B) lingual samples. Levels were quantified using a standard curve of bacterial and *Candida* DNA and expressed as equivalent CFU/ml

### 5.3.2 Influence of *Candida* spp. carriage on the microbiome

The impact of *Candida* carriage on the bacterial microbiome was investigated. The differences in bacterial species richness was investigated between *Candida*-positive samples (N = 25) and *Candida*-negative samples (N = 82). A rarefaction curve was used to compare the number of OTUs in each sample and this demonstrated that levels of species richness were higher in samples from *Candida* carriers compared to non-*Candida* carriers (Figure 5.3a). Analysis of biodiversity using the Inverse Simpson index also showed *Candida* carriers had greater biodiversity relative to non-carriers (Fig. 5.3b).

The effect of *Candida* carriage on population stature was then assessed using NMDS analysis of Bray-Curtis dissimilarity values carried out in Chapter 3, it was determined whether *Candida* carriers and non-*Candida* carriers had different microbial population structures (Fig. 5.4a). *Candida* carriers were found to cluster on the left-hand side of axis two. Statistical analysis with AMOVA showed that this clustering was significantly different ($p < 0.001$).

In order to determine which OTUs were responsible for this difference, we carried out LEfSe analysis of the V1-V2 microbiome data comparing *Candida*-carriers and non-*Candida* carriers (Fig. 5.4b). This was conducted showed that *Candida*-carriers had higher levels of some bacterial species including *H. parainfluenza* (OTU0002), *N. flavescens* OTU0010, *F. nucleatum* (OTU0016), *P. pasteri* (OTU0024), and *Leptotrichia* HOT215 (OTU0044). Levels of *Actinomyces* spp., and *Streptococcus* spp. were generally higher in non-*Candida* carriers.

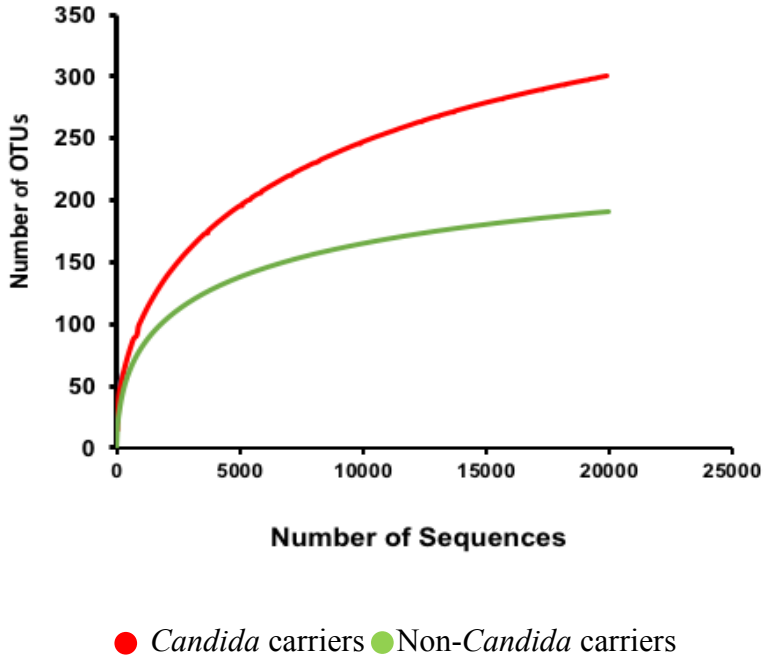



Figure 5.3a Shows a rarefaction curve of species richness in *Candida*

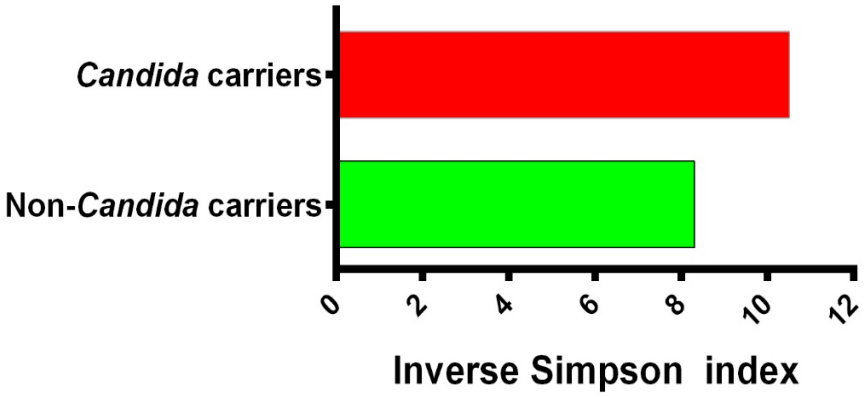



Figure. 5.3b

Figure 5.3b Shows Inverse Simpson index in *Candida* carriers and non-*Candida* carriers

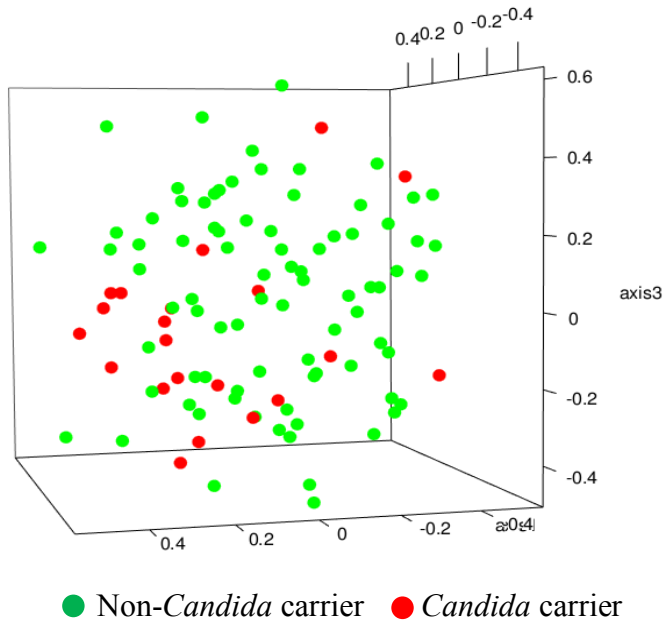
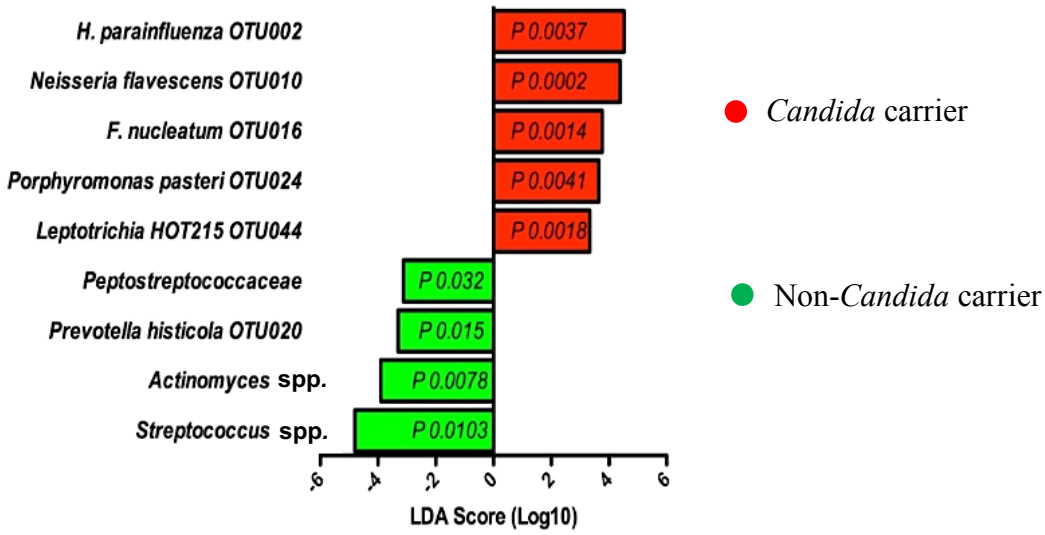



Figure 5.4a Shows NMDS plot generated using the Bray-Curtis index showing population structure in *Candida* carriers and non-*Candida* carriers (AMOVA $p < 0.001$)




(B)

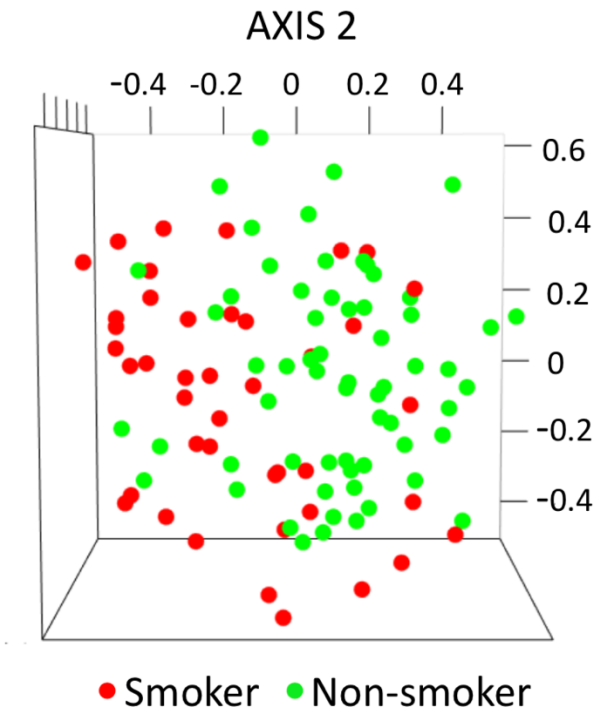
Figure 5.4b Shows LEfSe analysis of the V1-V2 16S rRNA microbiome data in *Candida* carriers and non-*Candida* carriers.

### 5.3.3 The effects of smoking on the oral microbes

Samples were classified into two groups, smokers and non-smokers. NMDS of Bray-Curtis dissimilarity values was performed to compare bacterial population structure in the two groups. This analysis showed that the populations of smoker and non-smokers could be separated along axis two of the graph (Fig. 5.5a). Statistical analysis with AMOVA showed that this separation was highly significant ($p = 0.001$).

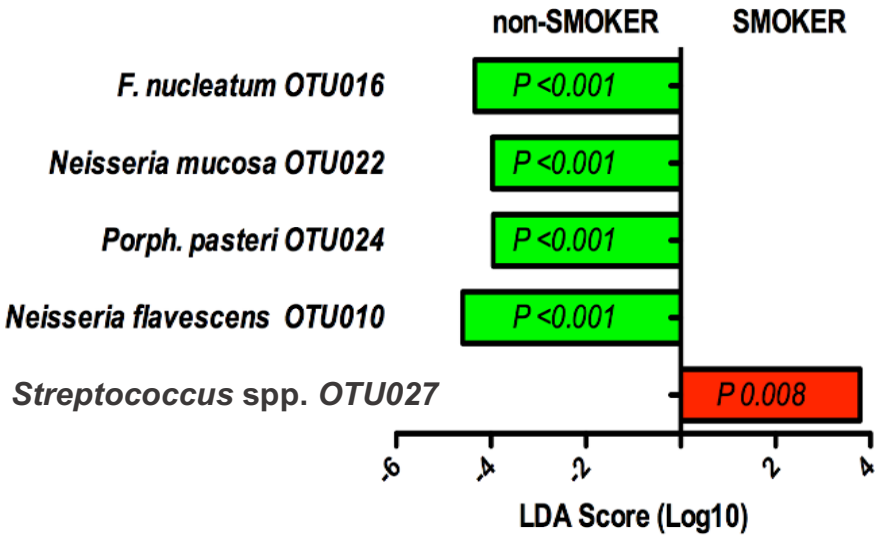
In order to determine which OTUs were responsible for this difference, we carried out LEfSe analysis of the V1-V2 microbiome data comparing smokers and non-smokers. This analysis showed that only one bacterial OTU was enriched in smokers, *Streptococcus* spp. OTU0027. However, LEfSe analysis identified several OTUs with increased abundances in non-smokers, including *F. nucleatum* OTU0016, *N. mucosa* OTU0022, *N. flavescens* OTU0010 and *P. pasteri* OTU0024 (Fig. 5.5b).

In order to confirm these findings, plots showing the abundance of each of these species were generated showing their abundance in smokers and non-smokers. *F. nucleatum* OTU0016 levels were higher in non-smokers relative to smokers (5.6a), and leukoplakia samples from non-smokers exhibited higher levels than smokers. Similar results were shown for both *N. mucosa* OTU0022 and *P. pasteri* OTU0024 (Figs. 5.6 b and c)



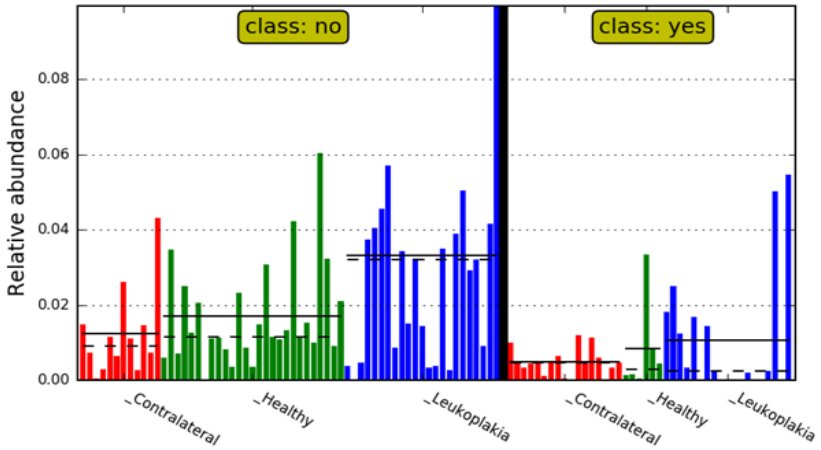

(a)

Figure 5.5a Shows NMDS plot generated using the Bray-Curtis dissimilarity index comparing smokers and non-smokers.

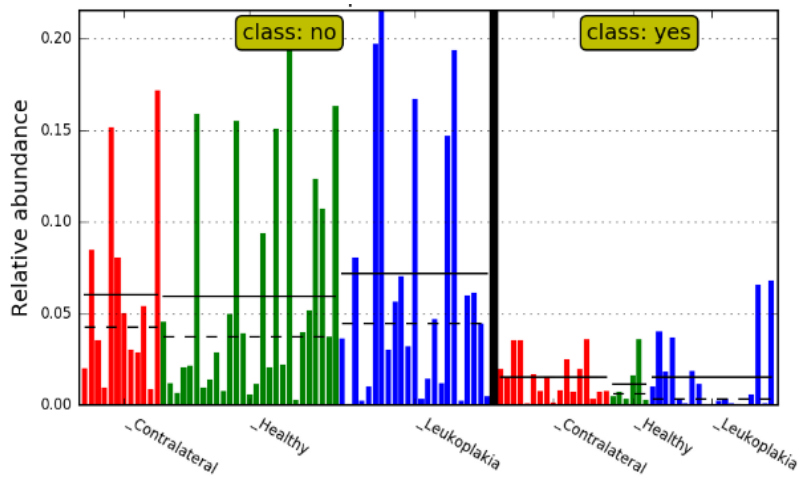



(b)

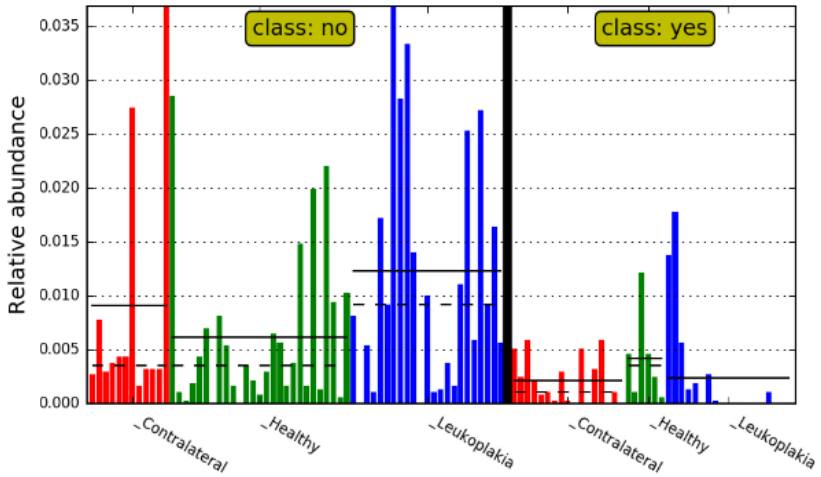
Figure 5.5b Shows LEfSe analysis of smokers and non-smokers showing the abundant species.




(a) *Fusobacterium nucleatum* OTU0016




(b) *Neisseria mucosa* OTU0022




(c) *Porphyromonas pasteri* (OTU0024)

Figure 5.6 Plots showing the relative abundance of (a) *F. nucleatum* OTU0016, (b) *N. mucosa* OTU0022 and (c) *P. pasteri* OTU0024 (class: yes = smokers) - (class: no = non-smokers)

### 5.3.4 The influence of alcohol consumption on the microbiome

Population structure was compared in alcohol drinkers and non-alcohol drinkers. It was found that there was no significant difference when Bray-Curtis dissimilarity values were analysed (AMOVA $p = 0.879$) and the NMDS plot showed that the samples that were not grouped differently (Fig. 5.7). However, the LEfSe test was able to identify three OTUs enriched in patients that consumed alcohol. This including *Campylobacter gracilis* OTU0067, *Kingella denitrificans* OTU023 and *Oribacterium* spp. OTU0075 (Fig. 5.8). *Campylobacter* spp., were previously shown in Chapter 3 (Fig. 3.19) to exhibit enrichment in leukoplakia samples, and this analysis showed that leukoplakia patients who consumed alcohol exhibited the highest levels of *Campylobacter gracilis* OTU0067 (Fig. 5.8).

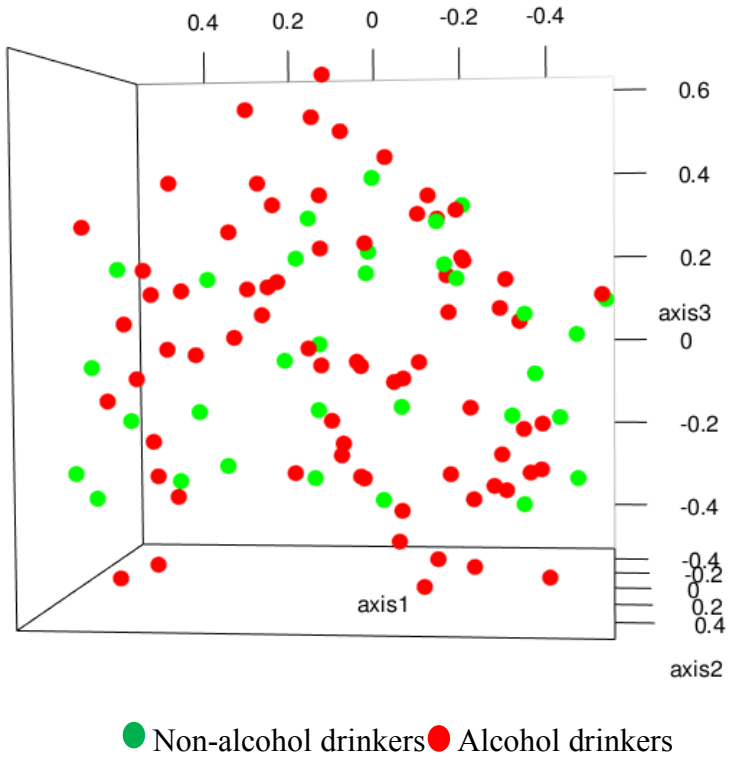



Figure 5.7 Shows NMDS plot generated using the Bray-Curtis dissimilarity index comparing alcohol drinkers and non-drinkers

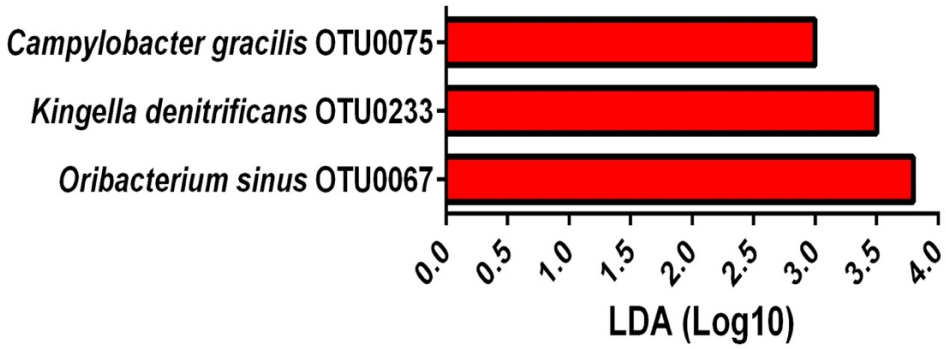
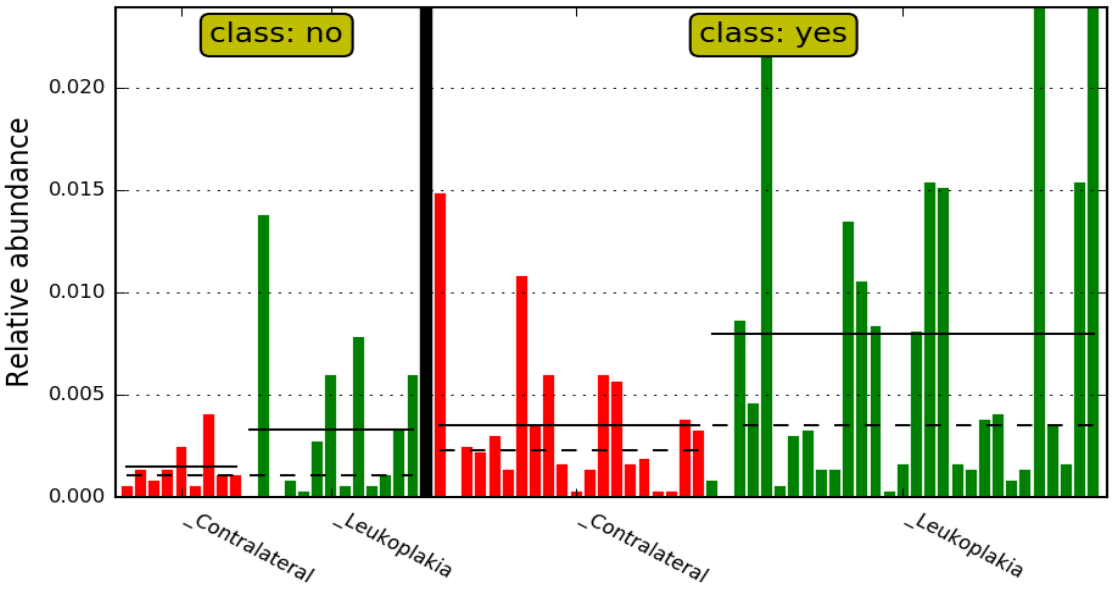



Figure 5.8 LEfSe analysis of alcohol drinkers and non-drinkers showing the abundant species.




class no = non-alcohol drinker class yes = alcohol drinker

Figure 5.9 Plot showing the relative abundance of *Campylobacter gracilis* OTU0067 in leukoplakia and contralateral samples from patients with and without a history of alcohol consumption.

### 5.3.5 The influence of mouthwash uses and denture wearing on the oral microbiome

Samples were also investigated for the effects of mouthwash use. LEfSe analysis determined that non-mouthwash users had increased levels of *S. gordonii* OTU0028, *Kingella oralis* OTU0099 and *Eikenella corrodens* OTU0192 compared with mouthwash users (Fig. 5.10)

Denture wearers were found to harbour higher levels of *Corynebacterium matruchotii* OTU0088 and *Prevotella oulorum* OTU0071. However, no species were seen to be significantly higher with non-denture wearers (Fig.5.11).

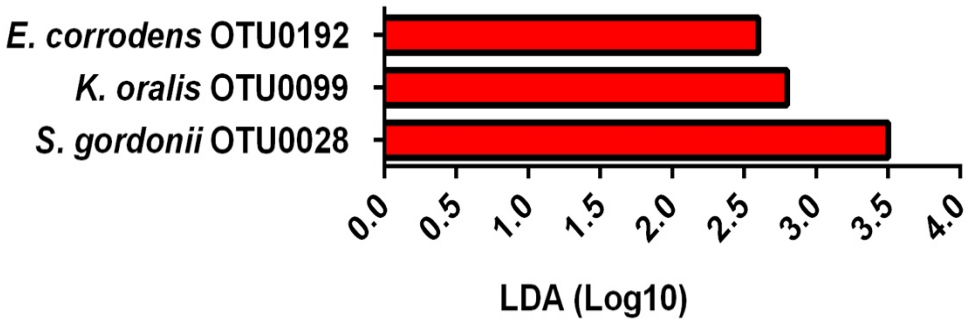



Figure 5.10 LEfSe results showing OTUs enriched in patients who did not use mouth washes

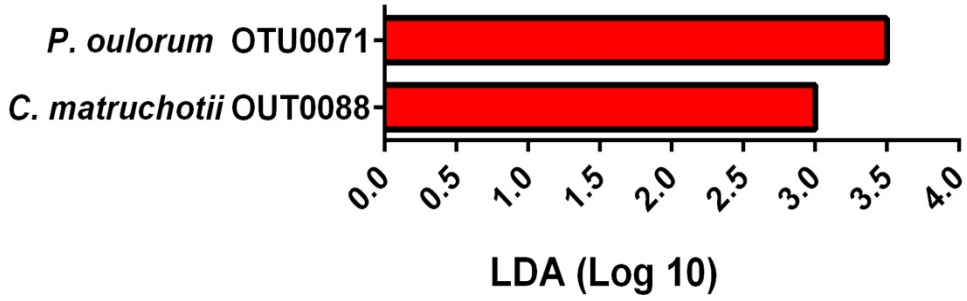



Figure 5.11 LEfSe results showing OTUs enriched in patients with dentures.

### 5.3.6 The influence of the degree of dysplasia in oral leukoplakia on the oral microbiome

In this analysis, the oral microbiome in leukoplakia samples to determine whether the degree of dysplasia (mild, moderate or severe) had any influence on the microbiome. LEfSe identified several OTUs that exhibited increased abundance in leukoplakia from tissues diagnosed with severe dysplasia (Fig. 5.12). This identified *Leptotrichia* spp., in addition to the specific OTU *Leptotrichia* spp., oral taxon 215 OTU0044 in tissue with severe dysplasia. In addition to the genus *Leptotrichia*, this analysis identified *Campylobacter concisus* OTU0053 and *Campylobacter gracilis* OTU0067 in association with severe dysplasia. No OTU enrichments were observed in mild or moderate types.

Analysis of the levels of *Leptotrichia* spp., oral taxon 215 OTU0044 in leukoplakia samples showed a large degree of variation in different sample types (Fig. 5.13). Leukoplakias with severe dysplasia tended to show higher levels of this OUT; however, the number of samples in this category were low (n = 6).

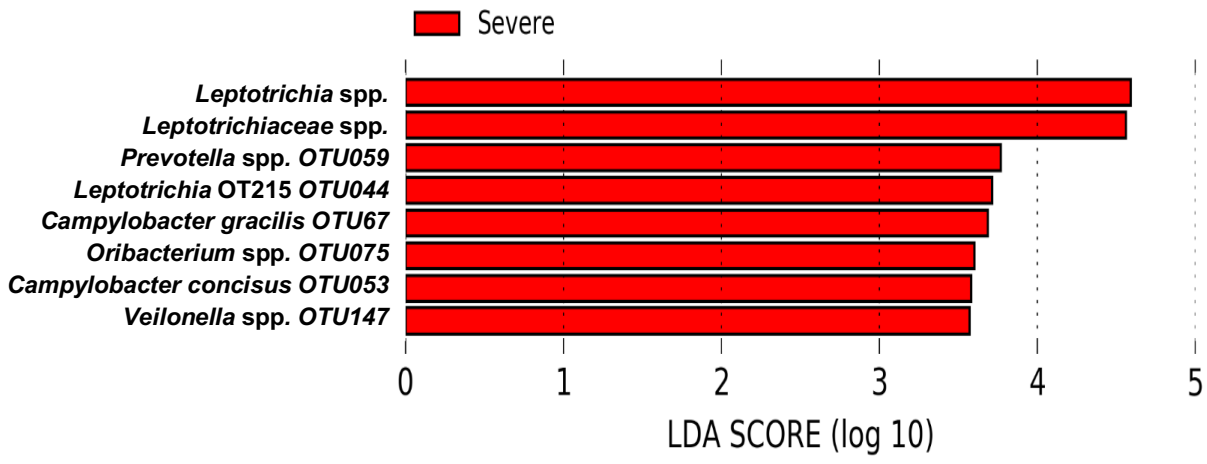



Figure 5.12 Shows LEfSe analysis of microbiome populations in samples from oral leukoplakia comparing mild, moderate and severs dysplasia. Only severe dysplasia exhibited enrichments deemed significant (LDA < 3.0, $p < 0.05$).

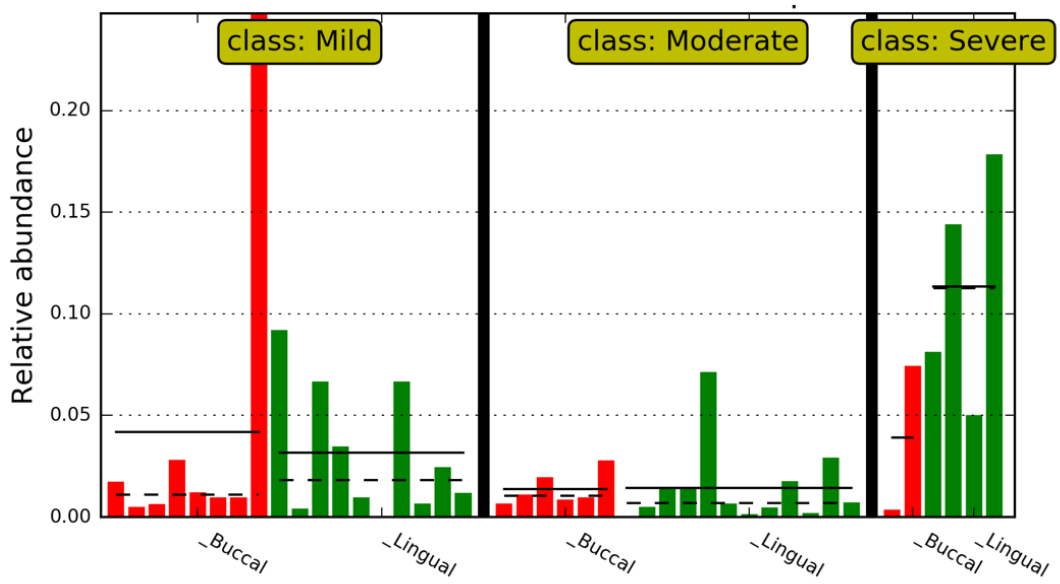



Figure 5.13 Abundance of *Leptotrichia* OT215 OTU0044 in leukoplakia samples (mild, moderate and severe). Analysis was done only for the most abundant taxon only.

### 5.3.7 Analysis of co-location patterns in bacterial populations in patients with oral leukoplakia

In order to determine if location (buccal, lingual), smoking, *Candida* carriage or the degree of dysplasia influenced colonisation with the leukoplakia enriched taxa identified in Chapters 3 and 4, heatmap was generated showing the abundance of these species in each of the oral leukoplakia samples (Fig. 5.14). The profiles for each sample were clustered using the Bray-Curtis dissimilarity value generated for each profile. Phylogenetic analysis was used to examine the relationships between these populations and a dendogram was generated to cluster closely related profiles. Analysis of this dendogram shows 5 groups representing different enrichment patterns in the heatmap.These groups were statistically significantly different (UNIFRAC $p < 0.001$). Cluster 3, 4 and 5 consisted of samples recovered from lingual sites, whereas clusters 1 and 2 contained more buccal samples. When the degree of dysplasia is examined in each

sample, most clusters contain a mixture of mild and moderate presentations, with the exception of cluster 3 which contained 3 cases of severe dysplasia among a group of 7 patients.

In order to determine if any species had a significant tendency to co-locate, a Pearson correlation coefficient was determined for each pair of OTUs using Prism 4.0. These data were imported to Cytoscape 3.2 and samples that exhibited highly significant ($p < 0.01$) Pearson associations were shown to cluster to produce (Fig. 5.15). This analysis confirmed the presence of 5 main clusters of bacteria, as indicated in the heatmap. Clusters 1, 2 and 3 were recovered from buccal surfaces, whereas clusters 4 and 5 were recovered from lingual surfaces. Clusters 2 and 4 were predominantly recovered from surfaces colonised with *Candida* spp.

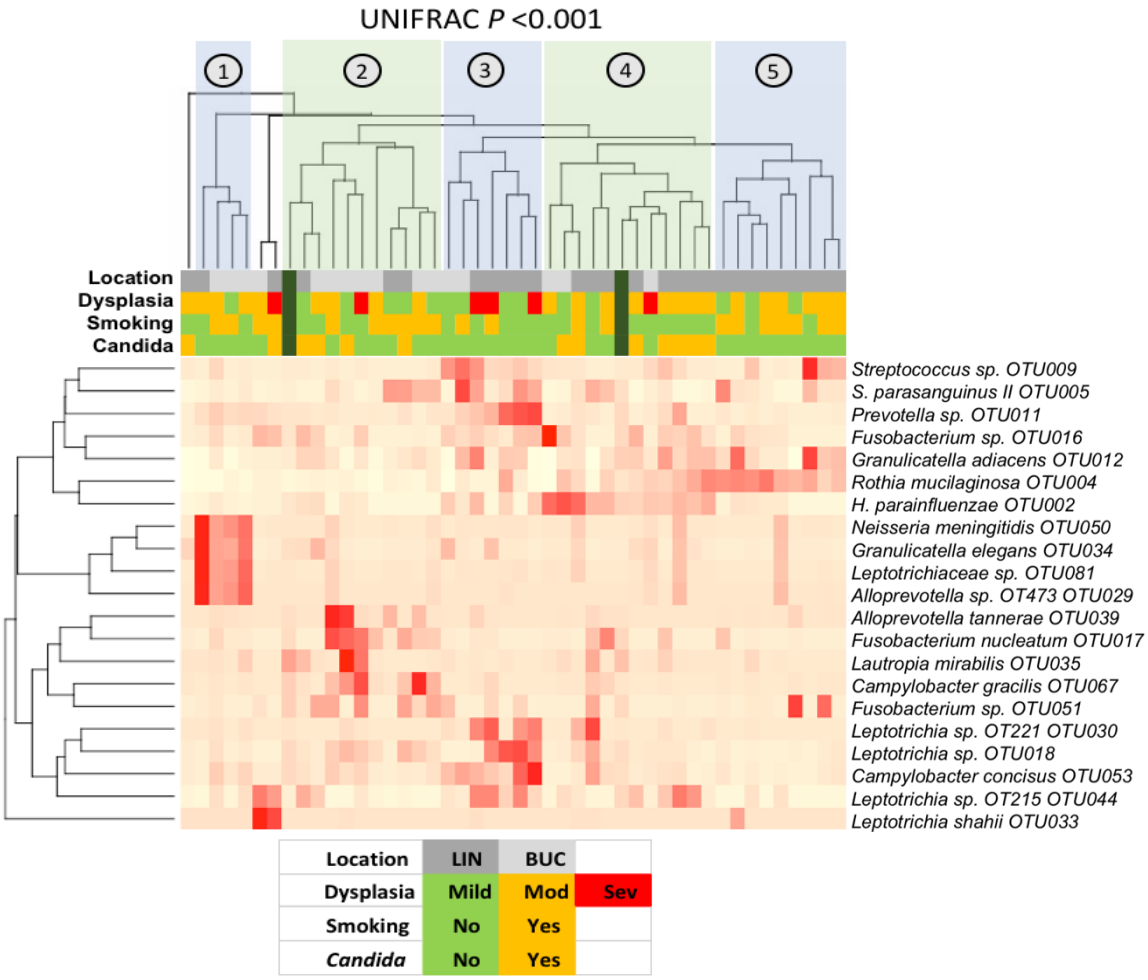



Figure 5.14 Heatmap of selected taxa identified in OLK patients by LEfSe analysis. Heatmap and dendogram were generated using Bray– Curtis dissimilarity values. Patient clusters were separated and labelled from group one to 5 in the dendogram and this clustering was highly significant (unweighted UNIFRAC $p < 0.001$). Samples were coloured based on whether it was the buccal or lingual site, the dysplasia degree (mild, moderate or severe stage), and whether the patient smoked or consumed alcohol.

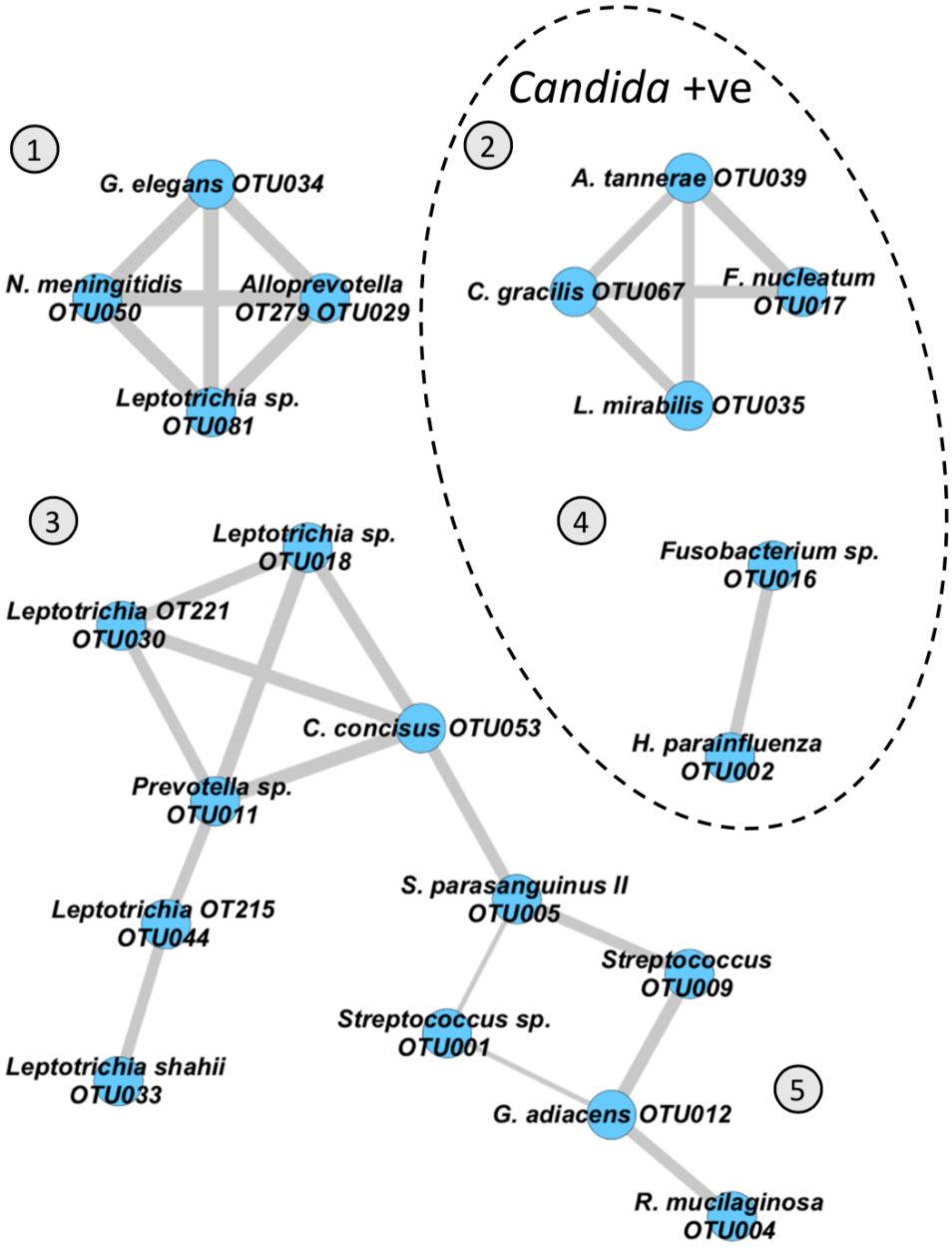



Figure 5.15 Association map generated in Cytoscape 3.2 from Pearson correlation coefficients (r) generated in Prism ($p < 0.01$) showing the 5 groups clustered separately. All of the *Candida* positive samples were located in clusters 2 and 4, and were in associated with *F. nucleatum.*

## 5.4 Discussion

Previous studies have shown that oral leukoplakia is commonly colonised by *Candida* species (Abdulrahim *et al.,* 2013). In the current study, using quantitative PCR, it was shown that 35% of leukoplakia samples were associated with *Candida* spp., whereas *Candida* spp., were detected in approximately 20% of contralateral normal sites and only 11% of healthy control patients. These results clearly support the hypothesis that *Candida* spp*.,* are associated with oral leukoplakia. The level of *Candida* colonisation was estimated using a standard curve generated with DNA extracted from known concentrations of *Candida* cells. The lowest level of colonisation recorded was equivalent to ~300 CFU/ml. Low level colonisation (< 300 CFU/ml) was not detected. This may be because low level colonisation was not present or perhaps the qPCR assay was not sensitive enough to detect low levels of *Candida* spp., in clinical samples. For this reason, it can be surmised that the assay detects 'significant' *Candida* colonisation, (equivalent to > 300 CFU/ml), and the analysis described examines the effect of significant *Candida* colonisation on the oral microbiome.

Previous analysis showed that leukoplakia samples showed a lower level of Firmicutes, which includes common species such as *Streptococcus* spp. (Chapter 3). Also, compared to contralateral normal sites, leukoplakia was associated with higher levels of the phylum Fusobacteria (Chapter 3). In order to investigate whether *Candida* carriage influenced the bacterial microbiome, we analysed alpha and beta diversity in *Candida* carriers and non-carriers. Rarefaction curves (Fig. 5.3a) comparing *Candida* carriers to non-carriers showed that OTU numbers were higher in *Candida* carriers. *Candida* carriers also exhibited increased microbiota biodiversity compared to non-*Candida* carriers (Fig. 5.3b). This difference was also reflected in the population structure analysis, which showed *Candida* carriers and non-carriers had significantly different populations

(AMOVA $p < 0.001$). *Candida* carriers were shown to have reduced levels of Gram-positive species and elevated levels of *H. parainfluenza* OTU0002, *N. flavescens* OTU0010, *F. nucleatum* OTU0016 and *Leptotrichia* HOT215 OTU044. These co-enrichment patterns may be due to the fact that the surface of the leukoplakia is suited to colonisation with this cohort of species. It may also indicate direct co-aggregation of *Candida* species and these bacterial species could promote colonisation of these bacteria. Previous studies have shown co-aggregation of *F. nucleatum* and *C. albicans* and the presence of this yeast could promote adherence and colonisation with *F. nucleatum*. However, in non-carriers, S*treptococcus* spp., *Actinomyces* spp., *Prevotella histicola* and the *Peptostreptococaceae* were more abundant, suggesting that these species could either inhibit candidal growth or thrive in conditions that do not favour *Candida* colonisation.

Smoking was also shown to significantly affect the mucosal microbiome. Non-smokers were found to have a more diverse microbiome including *F. nucleatum* OTU016, *N. flavescens* OTU0010, *N. mucosa* OTU0022, and *P. pasteri* OTU0024. These data are similar to those recently published which showed in a large cohort of 1,204 American adults that smokers had reduced biodiversity and in particular showed reduced levels of *Neisseria* spp., among others (Wu *et al.,* 2016). Smoking has profound effects on the oral cavity and can suppress local immunity as well as having direct toxic effects on the local microbiome. The organisms which show enrichment in oral leukoplakia (e.g. Fusobacteria, *R. mucilaginosa*) do not seem to be related to smoking, suggesting that the enrichments observed in Chapter 3 are not directly related to smoking factor.

We also analysed the samples for enrichments associated with alcohol consumption. Although population structure was not affected by alcohol consumption, 3 species were found at higher levels in alcohol consumers, *Oribacterium* spp. OTU0075, *Kingella denitrificans* OTU0233 and *Campylobacter gracilis* OTU0067. It is difficult to determine

the significance of these enrichments; however, *C. gracilis* OTU0067 generally exhibits higher levels on oral leukoplakia tissues and this enrichment is higher in alcohol consumers. The effects of alcohol consumption could be immunosuppressive or the data could simply reflect better oral hygiene in non-drinkers. A larger cohort study is needed to determine the effects of alcohol consumption on the oral microbiome.

Also, analysis of the samples in relation to the degree of dysplasia revealed that severe dysplasia had specific enrichments relative to mild and moderate types. These species included *Leptotrichia* spp. and *Campylobacter* spp. Two out of six samples exhibiting severe dysplasia were recovered from patients who had a history of smoking and only 1 sample out of 6 exhibited *Candida* carriage, suggesting that these enrichments were not associated with these factors. This enrichment might be due to the nature of the oral epithelial membrane structure, which may allow only certain species to adhere and colonise the diseased lesion.

Further analysis was carried out to determine how the factors described above effected the bacterial enrichments on leukoplakia identified here and in Chapter 3 Species that exhibited an association with OLK or severe dysplasia were included in a Heatmap which when subjected to phylogenetic analysis, showed the presence of 5 different enrichment patterns, which were significant (UNIFRAC $p > 0.001$). Pearson correlation coefficients also supported the existence of specific clusters of bacteria (Fig. 5.15). It was clear from this analysis that leukoplakias from buccal and lingual sites had different enrichment patterns, with lingual sites exhibiting *F. nucleatum subsp polymorphum* OTU0016 and *R. mucilaginosa*, and buccal sites exhibiting higher levels of *F. nucleatum subsp vincentii* OTU0017. *Candida* carriage also separated the clusters, with most *Candida* carriers clustering in group 2 (buccal) and group 4 (lingual). This analysis shows that although the general enrichments associated with oral leukoplakia identified in Chapter 3 and

Chapter 4 are significant, the composition of the microbiome on individual leukoplakias is influenced by many factors, in particular smoking, *Candida* carriage and the anatomical location of the lesion. In order to completely elucidate the influence of these various factors, a larger cohort study is required. This study included a relatively small sample size. For example, only 6 samples exhibiting severe dysplasia were included in our analysis, making it difficult to draw any significant conclusions from the analysis in this chapter.

# Chapter 6

## Analysis of acetaldehyde production by oral isolates of *Rothia mucilaginosa*

## 6.1 Introduction

Analysis of the V1-V2 regions of the 16S rRNA gene in Chapter 3, showed that leukoplakia samples were significantly enriched by some bacteria compared to healthy samples, such as *R. mucilaginosa* OTU0004 (Fig 3.18), while *F. nucleatum* OTU0017 and *Campylobacter* spp., were overrepresented in OLK compared to CN (Fig. 3.19). In this chapter we present a functional analysis of these communities using the metagenome prediction tool PICRUSt. As, this analysis predicts a reduced capacity to metabolise acetaldehyde in OLK communities due to reduced levels of the gene encoding acetaldehyde dehydrogenase. We hypothesise that this may be due to the enrichment for *R. mucilaginosa* in these tissues and carry out an examination of the acetaldehyde generating capacity of this species.

There are 5 species of *Rothia* described in HOMD, namely *Rothia mucilaginosa, Rothia dentocariosa, Rothia aeria, Rothia nasimurium,* and *Rothia amarae*. *Rothia mucilaginosa* was previous known as *Stomatococcus mucilaginosa* (Janda and Abbott, 2007), *R. mucilaginosa* are Gram-positive, aerobic and non-spore forming bacterium classified to the phylum Actinobacteria. It has been reported that *Rothia* spp., can cause opportunistic infections, such as joint infections, meningitis, peritonitis and soft skin infection (Ramanan *et al.,* 2014). *Rothia* spp. may also play an accessory role in dental caries and periodontal disease (Nambu *et al.,* 2013), while other studies suggest that *R. mucilaginosa* is a member of the healthy oral microbial community (Zaura *et al.,* 2009).

Acetaldehyde is the first compound produced by microbes during ethanol metabolism by the enzyme alcohol dehydrogenase (Elamin *et al.,* 2014). Acetaldehyde can be a further converted to acetate by microbial and host acetaldehyde dehydrogenases (Fig. 6.1) (Seitz and Becker, 2007). Acetaldehyde has been classified as a carcinogenic substance according to the International Agency for Research on Cancer (IARC) (Elamin *et al.,*

2014). Patients with oropharyngeal cancer have been found to have increased levels of ACH concentrations in their saliva (Jokelainen *et al.,*1996). Acetaldehyde is a highly reactive and a carcinogenic compound, some oral microbes have been implicated in the production of ACH from ethanol including *Neisseria* spp., and *Candida* spp. (Moritani *et al.,* 2015). Also a study of healthy individuals found high levels of acetaldehyde in their saliva, and associated with high proportion of the *R. mucilaginosa in their salvia* (Yokoyama *et al.,* 2018).

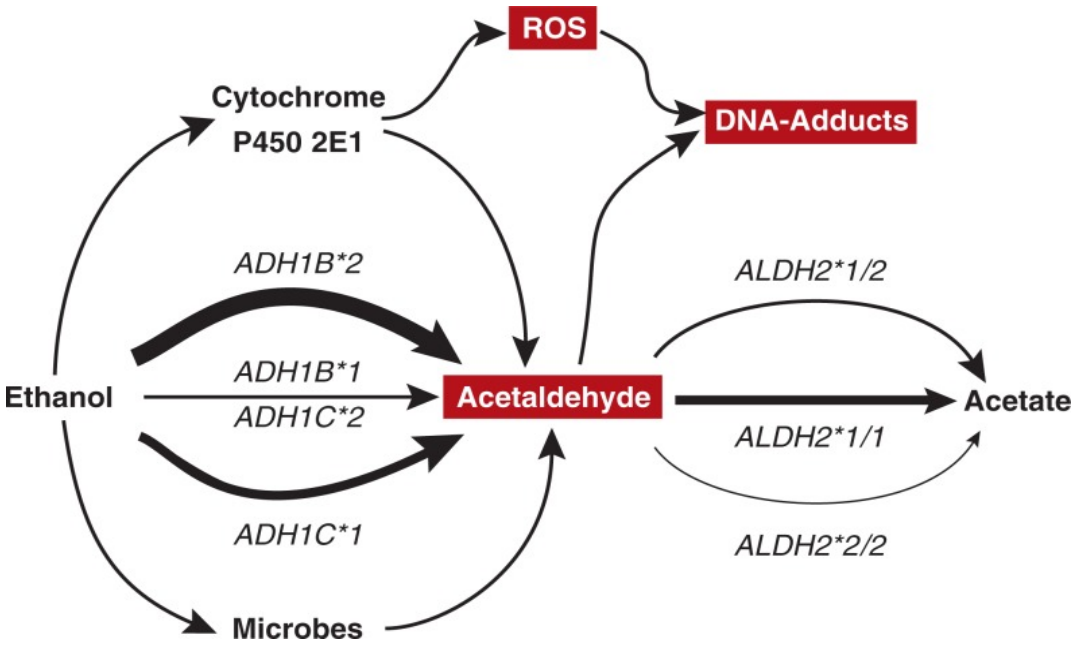



Figure 6.1 Different ways of ethanol metabolism and ACH production and conversion to acetate in the human body.

Figure reproduced with permission from Seitz and Becker (2007).

## 6.2 Materials and Methods

### 6.2.1 Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt)

Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) is a bioinformatics software package that can be used to predict gene family abundance (e.g. the metagenome) in microbial communities for which only marker gene (e.g. 16S rRNA gene) data are available. The package was developed by Langille *et al.* (2013), and is freely available to download (http://picrust.github.io/picrust/).

The relative abundance data for each OTU was used to generate a biome format file for use in PICRUSt. Output files showing the distribution of genes (catalogued by Kegg ontologies) and pathway abundances were generated. These files were analysed using LEfSe to identify categories enriched in OLK or healthy tissues.

### 6.2.2 Isolation and identification of *Rothia mucilaginosa*

Swabs recovered from OLK patients were suspended in 500 µl of 1 X PBS buffer. From this, a 100 µl aliquot was spread on *Rothia* selective Brain Heart Infusion (BHI) agar plates, with the remainder frozen at -80˚C for later DNA extraction. Selective BHI agar media consisted of standard BHI agar supplemented with sodium selenite ($Na_2SeO_3$) at a concentration of 50 mg/l and colistin at concentration 10 mg/l (Kobayashi *et al.*, 2012).

Although this medium did not suppress the growth of all other bacteria, *R. mucilaginosa* produced larger colonies (Fig. 6.2). Typical colonies were sub-cultured on selective BHI to ensure purity. DNA extraction was carried out using phenol: chloroform; isoamyl alcohol (25:24:1) as described (Chapter 2). PCR amplification was carried out using the 27F and 1492R primers to amplify the 16S rRNA gene. This was purified using AMPure XP magnetic purification beads (Beckman Coulter, Brea, CA, USA). PCR products were

then diluted to 10 ng/µl and sent to Source Biosciences for Sanger DNA sequencing with the 27F and 1492R primers.

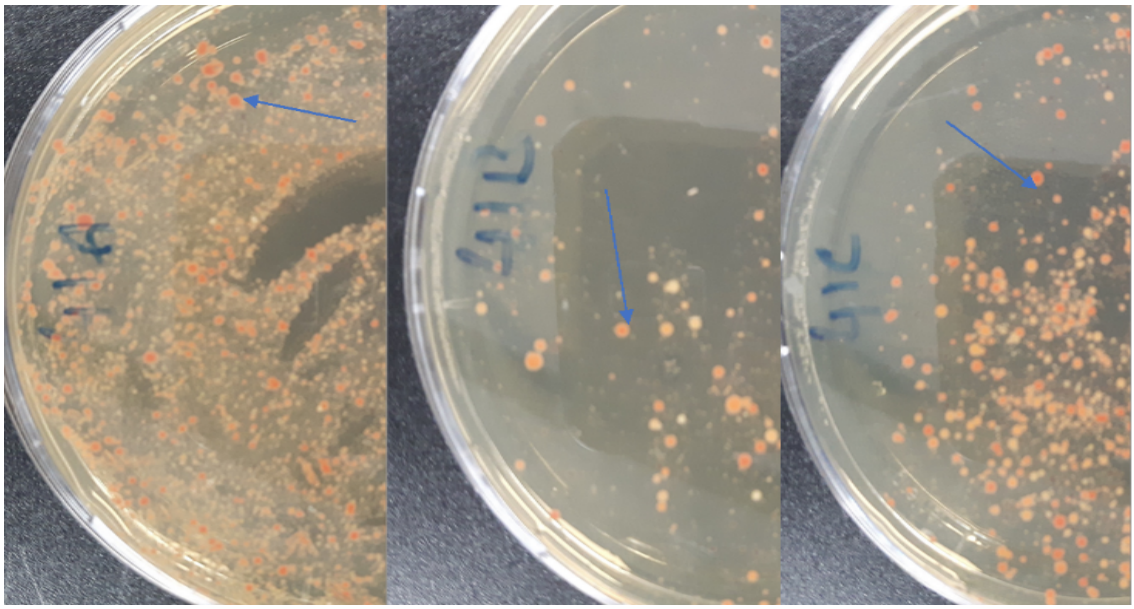


Figure 6.2 Shows BHI media were used to recover *Rothia* spp. The blue arrows indicate putative colonies were similar to *Rothia* spp.

### 6.2.3 Acetaldehyde detection

Colorimetric aldehyde detection assay kits were used to detect aldehydes produced by oral microbes at different final concentrations of ethanol. Several common oral microbes were included in this study, including *S. mitis* NCTC12261, *S. gordonii* DL1, *Neisseria mucosa* DSM-17611 and *C. albicans* 132A (Gallagher *et al.*, 1992).

Bacteria were grown overnight in BHI broth at 37˚C for the indicated times before harvesting and washing with PBS buffer. A suspension measuring 0.5 at OD 600 nm was prepared. A 450 µl volume was transferred to screw capped tube (Starstedt) and 50 µl of ethanol, was added to yield a final concentration of 10, 50 or 100 mM. After incubation for 2 h the reaction was stopped by frozen at -80˚C. Total aldehydes were measured using a colorimetric aldehyde detection assay kit according to the manufacturer's instructions.

Samples were incubated with the detection reagent for 1 h and the coloured reaction product was measured by determining the absorbance at 540 nm in a Genios 96-well plate reader (Tecan, Männedorf, Switzerland). Acetaldehyde concentration in each sample was estimated from a standard curve using the positive control values from a serial dilution of a glutaraldehyde standard.

### 6.2.4 Measurement of ACH induced ROS production

To estimate the levels of oxidative stress in TR146 cells, reactive oxygen species (ROS) production was assessed using 2',7'Dichlorodihydrofluorescein diacetate (DCF-DA) which is non-fluorescent unless oxidized by intracellular ROS. TR146 cells were incubated with 100 μM DCF-DA in 96-well plates for 1 h at 37˚C. Measurement of intracellular ROS was performed by incubating TR146 cells with either serum-free medium only as negative controls or acetaldehyde (25, 50, 75, 100 μM) for 3 h with or without 1 h pretreatment with 100 μM of the antioxidant N-acetylcysteine (NAC). $H_2O_2$ (30 mM) was used as a positive control. Cells were then washed two times in Hank's balanced salt solution buffer (HBSS) and the fluorescence (485 nm excitation/530 nm emission) was measured using a Tecan plate-reader (Genios, Tecan). The effect of ACH at these levels was also examined using the fluorescence microscope DCF-DA fluorescence was visualised with a GFP filter set (485 nm excitation/530 nm emission) and Hoechst stained nuclei visualised with a Zeiss AX10 Epifluorescence microscope (Zeiss Instruments, Oberkochen, Germany).

## 6.3 Results

### 6.3.1 Analysis of gene and pathway enrichment

Using the PICRUSt tool, the relative abundance of genes in the metagenome of each sample was predicted along with the abundance of metabolic pathways. Using LEfSe, the abundance of these categories was compared between OLK and normal contralateral tissues and samples from healthy. Examination of the pathway level data showed that OLK samples exhibited enrichment for genes in the citric acid cycle and amino acid related pathways compared to both samples from contralateral sites and healthy patients. In addition, OLK was shown to be enriched for genes involved in lipopolysaccharide biosynthesis compared to the contralateral samples (Fig. 6.3). Carbohydrate metabolism was generally more prevalent in contralateral tissues and healthy controls (Figs. 6.3 and 6.4).

At the gene level, the metagenomes of OLK samples were enriched in RNA polymerase $\sigma^{70}$ factors compared to both contralateral and healthy control samples (Figs. 6.5 and 6.6). A number of additional categories were differentially enriched between OLK and contralateral normal sites including several genes involved in lipid metabolism (OLK) and in contralateral tissues several sugar transporter systems were more abundant. We also observed an increased level of genes encoding acetaldehyde dehydrogenase (K04072) in normal contralateral tissue compared to OLK samples.

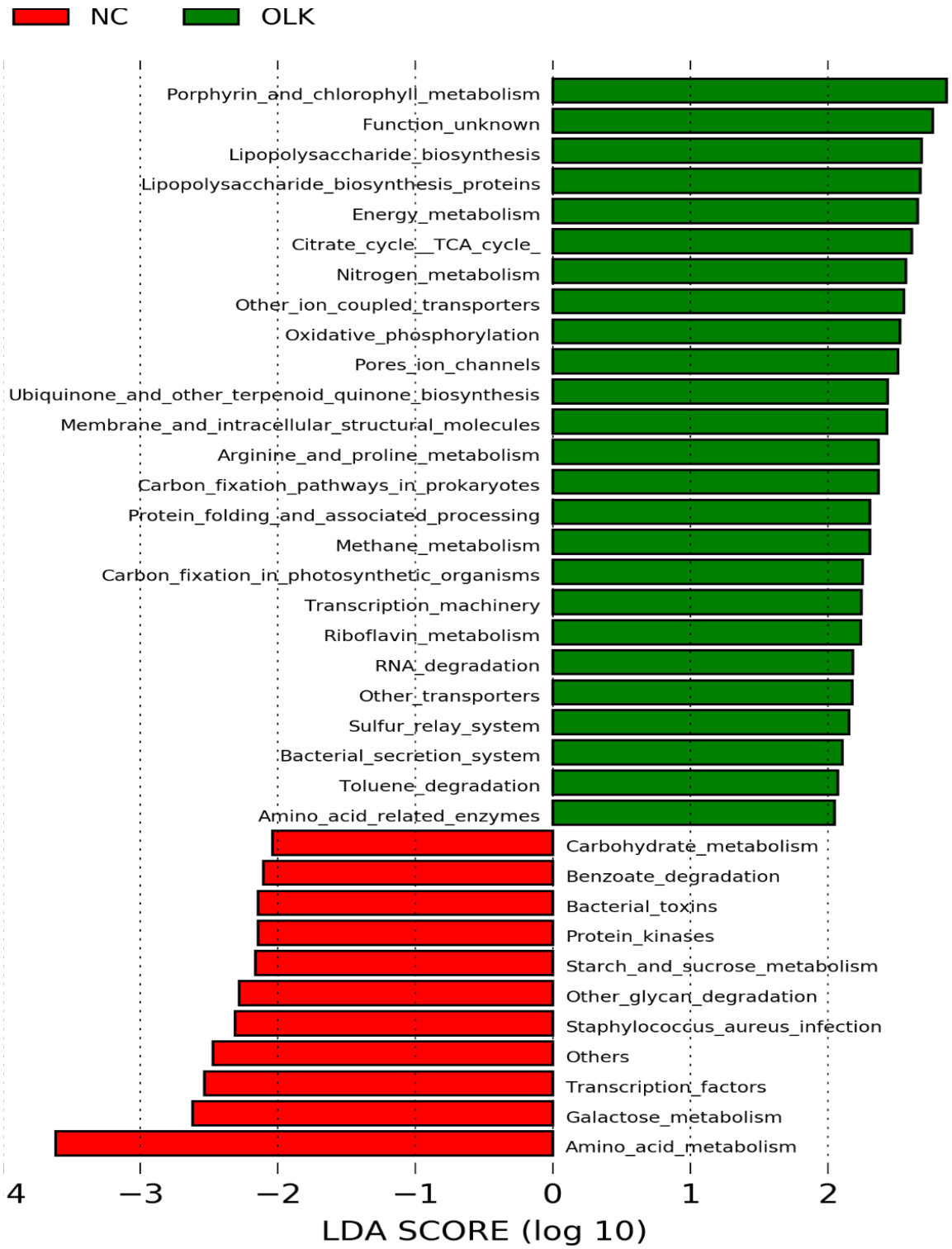



Figure 6.3 Pathways that were significantly differentially enriched between the leukoplakia samples (OLK) and normal contralateral samples (NC) identified by LEfSe.

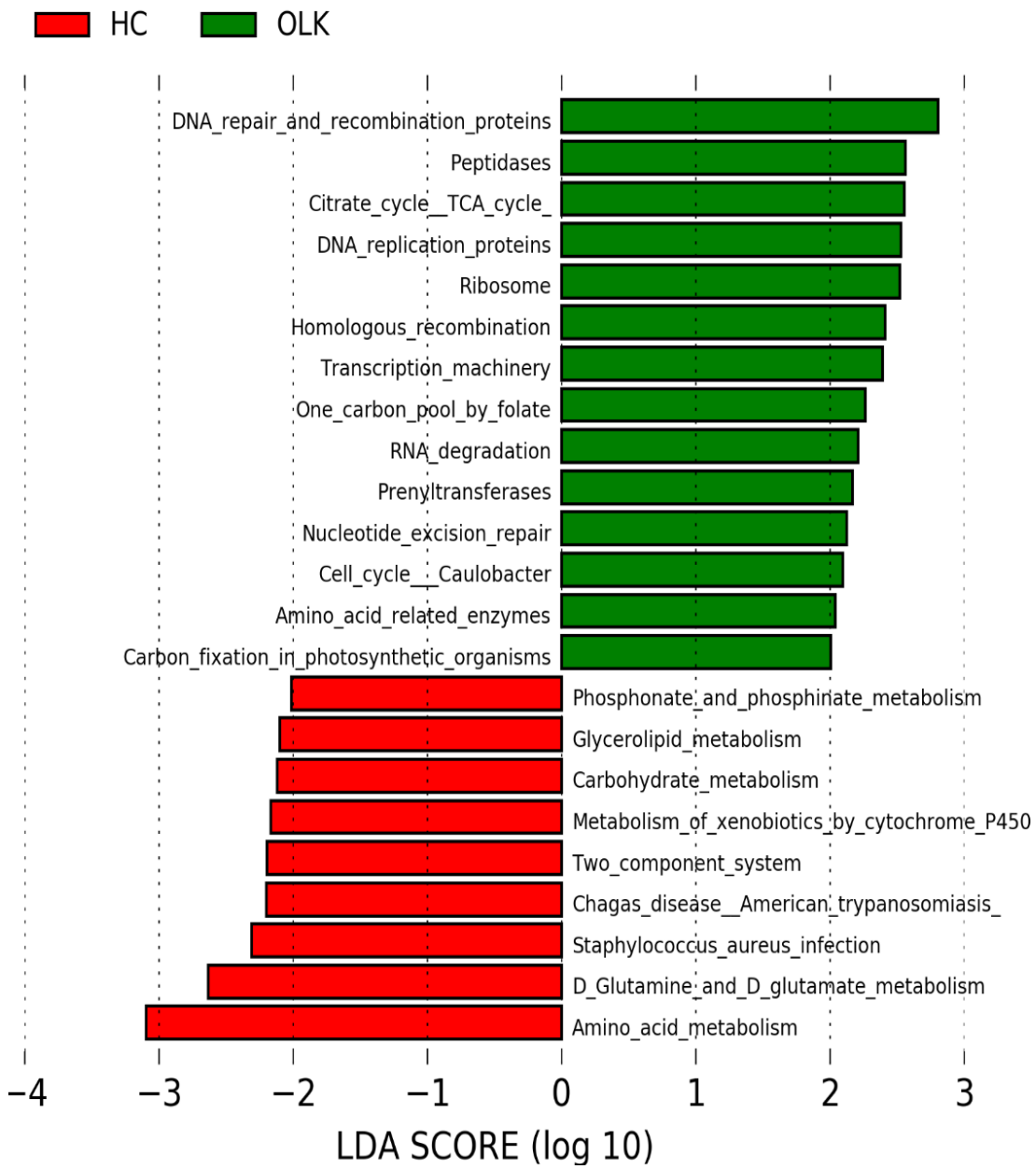



Figure 6.4 Showing different enriched of the pathways between leukoplakia samples (OLK) and healthy controls (HC) identified by LEfSe

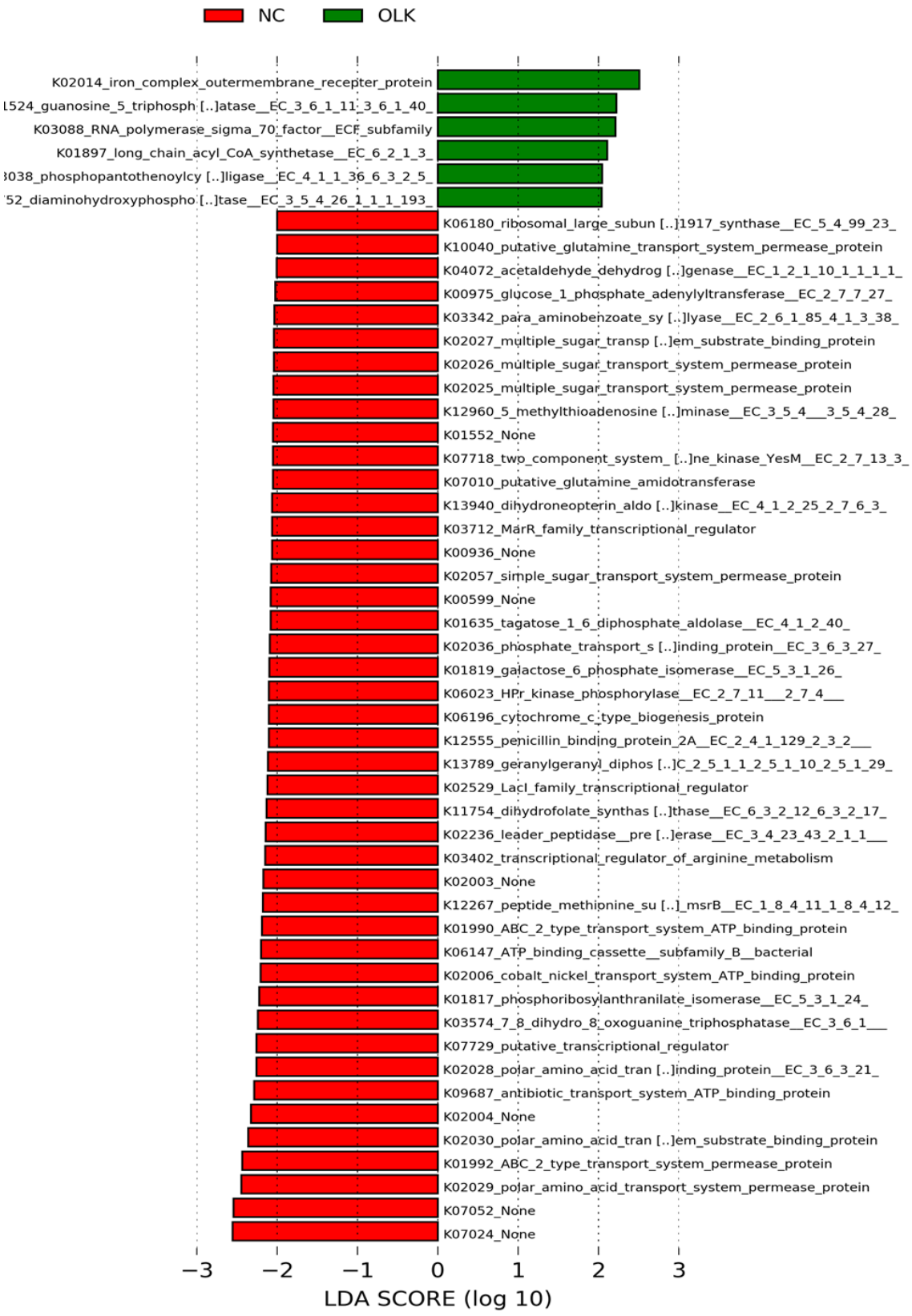



Figure 6.5 Differentially enriched of genes in the predicted metagenomes of oral leukoplakia (OLK) and contralateral normal (NC) samples identified by LEfSe

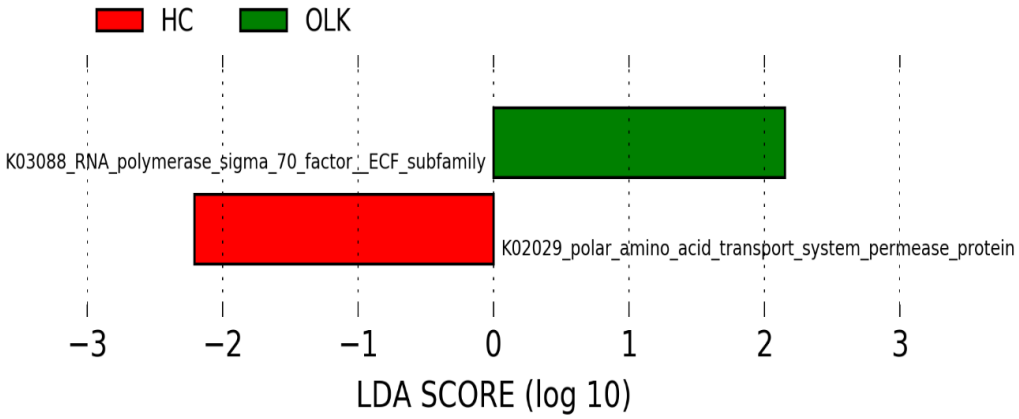



Figure 6.6 Differentially enriched of genes in the predicted metagenomes of oral leukoplakia (OLK) and healthy control (HC) samples identified by LEfSe.

### 6.3.2 Distribution of acetaldehyde dehydrogenase genes in oral bacteria

Due to the identification of acetaldehyde dehydrogenases (K04072) as being less abundant in OLK tissues, an examination of the distribution of ACH dehydrogenase genes (ACDH) in common oral bacterial species was performed using the PATRIC database (https://patricbrc.org). A total of 652 fully sequenced bacterial genome sequences representing the main genera of oral bacteria present in the oral microbiome were included (Table 6.1). In total, 260 streptococcal genomes were analysed and these were found to contain 479 genes. Some anaerobic species such as *Porphyromonas* spp., and *Prevotella* spp. did not contain any ACDH genes. Of the non-obligate anaerobic species, the Actinobacteria *Actinomyces* spp. and *Rothia* spp. had the lowest levels of ACDH genes (8.57 and 5.7% respectively). Due to the enrichment of *R. mucilaginosa* observed in OLK samples, we hypothesised that this enrichment may account for the reduced ACDH gene complement in these samples and that the absence of this gene could render *R. mucilaginosa* unable to metabolise the toxic by-products of alcohol metabolism.

**Table 6.1** Shows ACDH genes present in different oral bacteria.

| Bacterial genus | Number of genomes | Number of ACDH genes according to PATRIC database | Percentage of genomes with ACDH |
|---|---|---|---|
| *Actinomyces* spp | 105 | 9 genes | 8.57% |
| *Fusobacterium* spp | 15 | 18 genes | 100% |
| *Gemella* spp | 4 | 5 genes | 100% |
| *Granulicatella* spp | 6 | 8 genes | 100% |
| *Haemophilus* spp | 13 | 5 genes | 38.5% |
| *Leptotrichia* spp | 15 | 20 genes | 100% |
| *Neisseria* spp | 13 | 15 genes | 100% |
| *Porphyromonas* spp | 71 | 0 genes | 0.0% |
| *Prevotella* spp | 66 | 0 genes | 0.0% |
| *Rothia* spp | 70 | 4 genes | 5.7% |
| *Sneathia* spp | 1 | 2 genes | 100% |
| *Streptococcus* spp | 260 | 479 genes | 100% |
| *Tannerella* spp | 2 | 0 genes | 0.0% |
| *Veillonella* spp | 11 | 4 genes | 36.4% |

### 6.3.3 Isolation of *R. mucilaginosa* from OLK and healthy samples

A total of 11 isolates of *R. mucilaginosa* were recovered including from 6 OLKs, 2 from NC sites and 3 from healthy control patients (Table 6.2). Strains were isolated on selective BHI agar and were identified by DNA sequence analysis of the 16S rRNA gene (Table 6.3). The reference strain *R. mucilaginosa* DY-18 (DSM-20746) was used as a control.

**Table 6.2** Details of the 11 *Rothia* strains identified according to HOMD

| Patient ID | Primers | Strain ID | % 16S rRNA identity to *R. mucilaginosa* DY-18 |
|---|---|---|---|
| DY-18 (DSM20746) | N/A | *Rothia mucilaginosa* DSM-20746 | NA |
| | N/A | *Rothia mucilaginosa* DSM-20746 | NA |
| p 40A2 | 27F | *Rothia mucilaginosa* ID (40A2) | 99.7 |
| | 1249R | *Rothia mucilaginosa* ID (40A2) | 99.6 |
| p 41A2 | 27F | *Rothia mucilaginosa* ID (412) | 100 |
| | 1249R | *Rothia mucilaginosa* ID (41A2) | 100 |
| p 41C2.2 | 27F | *Rothia mucilaginosa* ID (41C2.2) | 99.8 |
| | 1249R | *Rothia mucilaginosa* ID (41C2.2) | 99.7 |
| p 43B1 | 27F | *Rothia mucilaginosa* ID (43B1) | 94.8 |
| | 1249R | *Rothia mucilaginosa* ID (43B1) | 91.9 |
| p 44A1 | 27F | *Rothia mucilaginosa* ID (44a1) | 99.5 |
| | 1249R | *Rothia mucilaginosa* ID (44A1) | 99.4 |
| p 44A2 | 27F | *Rothia mucilaginosa* ID (44A2) | 97.9 |
| | 1249R | *Rothia mucilaginosa* ID (44A2) | 97.9 |
| p107 | 27F | *Rothia mucilaginosa* ID (107) | 99.6 |
| | 1249R | *Rothia mucilaginosa* ID (107) | 99.6 |
| p 66A9 | 27F | *Rothia mucilaginosa* ID (66A9) | 99.7 |
| | 1249R | *Rothia mucilaginosa* ID (66A9) | 99.9 |
| p H2.2 | 27F | *Rothia mucilaginosa* ID (H2.2) | 99.1 |
| | 1249R | *Rothia mucilaginosa* ID (H2.2) | 97.8 |
| p H4.1 | 27F | *Rothia mucilaginosa* ID (H4.1) | 99.4 |
| | 1249R | *Rothia mucilaginosa* ID (H4.1) | 98.5 |
| p H4.2 | 27F | *Rothia mucilaginosa* ID (H4.2) | 97.7 |
| | 1249R | *Rothia mucilaginosa* ID (H4.2) | 97.7 |

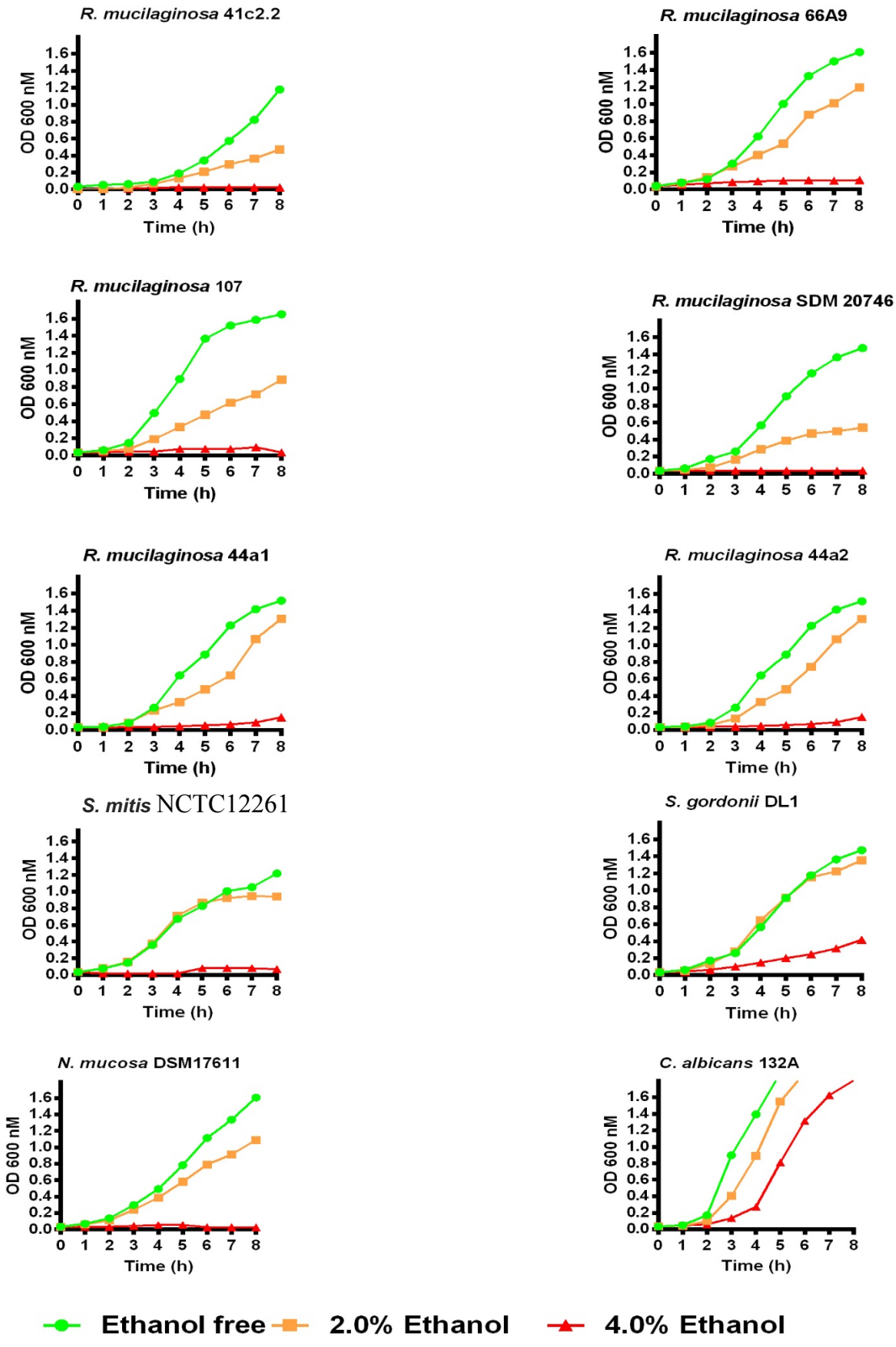
**Table 6.3** Characteristics of the patients from which *R. mucilaginosa* isolates were recovered.

| Sample Num. | Sample ID | Patient | Smoker | Alcohol consumption |
|---|---|---|---|---|
| 1 | *R. mucilaginosa* DY-18 DSM-20746 | AP | N/A | N/A |
| 2 | *R. mucilaginosa* ID (40A2) | OLK | yes | yes |
| 3 | *R. mucilaginosa* ID (41A2) | OLK | yes | yes |
| 4 | *R. mucilaginosa* ID (41C2.2) | NC | yes | yes |
| 5 | *R. mucilaginosa* ID (43B1) | NC | yes | yes |
| 6 | *R. mucilaginosa* ID (44A1) | OLK | no | yes |
| 7 | *R. mucilaginosa* ID (44A2) | OLK | no | yes |
| 8 | *R. mucilaginosa* ID (107) | OLK | yes | yes |
| 9 | *R. mucilaginosa* ID (66A9) | OLK | yes | yes |
| 10 | *R. mucilaginosa* ID (H2.2) | HC | no | yes |
| 11 | *R. mucilaginosa* ID (H4.1) | HC | no | no |
| 12 | *R. mucilaginosa* ID (H4.2) | HC | no | no |

AP: Apical periodontitis; OLK: oral leukoplakia; NC: contralateral normal; HC: healthy control.

### 6.3.4 Determination of growth curves of *Rothia* spp. and other oral microbes in ethanol

In order to determine the effect of ethanol on the growth of *Rothia* strains compared to other species, all strains were incubated in BHI broth with ethanol at a final concentration (v/v) of 0.0%, 2.0% and 4.0%. In general, all bacteria were severely inhibited by 4% ethanol, whereas *C. albicans* 132A was still capable of growth following a short lag period (Fig. 6.7). At 2% ethanol, all strains of *R. mucilaginosa* and *Neisseria mucosa* exhibited reduced growth whereas *S. mitis* NCTC12261 and *S. gordonii* DL1 were largely unaffected by growth in 2% ethanol.




Figures 6.7 Shows oral microbes were in BHI broth and different ethanol concentration.

### 6.3.5 Acetaldehyde production by *R. mucilaginosa* strains in the presence of ethanol

An assay was carried out using an exponential culture to determine the level of acetaldehyde produced by oral bacteria *in vitro* in the presence of ethanol. Cultures were grown in BHI media for 6 h at 37℃ until the OD 600 nm reached 0.8 - 1.0 (late exponential). Samples were then incubated in 0.0 mM, 10 mM, 50 mM, and 100 mM ethanol. All bacterial strains and *C. albicans* 132A produced different levels of aldehyde. *R. mucilaginosa* strains 41C2.2 and isolate P 107 produced over 100 μM of acetaldehyde. The known ACH producer *N. mucosa* also produced high levels of acetaldehyde (>100 μM) as did *C. albicans* 132A; however, *R. mucilaginosa* DY-18 was a weaker ACH producer, as was *S. mitis* NCTC12261 strain analysed here (Fig. 6.8). Acetaldehyde production was also measured in cells from stationary overnight cultures which were incubated for 24 h at 37℃. The results showed that *R. mucilaginosa* 41C2.2 was still capable of producing high levels of ACH (~70μM), while *R. mucilaginosa* DY-18 did not yield any ACH (Fig. 6.9).

Following these dose response experiments, a larger panel of *R. mucilaginosa* strains (n = 8), *S. mitis* NCTC12261, *S. gordonii* DL1, *Neisseria mucosa* DSM-17611 and *C. albicans* 132A were examined for acetaldehyde production. Isolates were incubated for 2 h with and without ethanol (100 mM). *R. mucilaginosa* 41C2.2 again showed the greatest amount of acetaldehyde produced and *R. mucilaginosa* strains 43B1, 66A9 and H2.2 were also capable of high level acetaldehyde production at approximately 50 μM or greater. The other *Rothia* isolate produced between 5 μM to 45 μM acetaldehyde. While *R. mucilaginosa* H4.2 did not produce acetaldehyde.

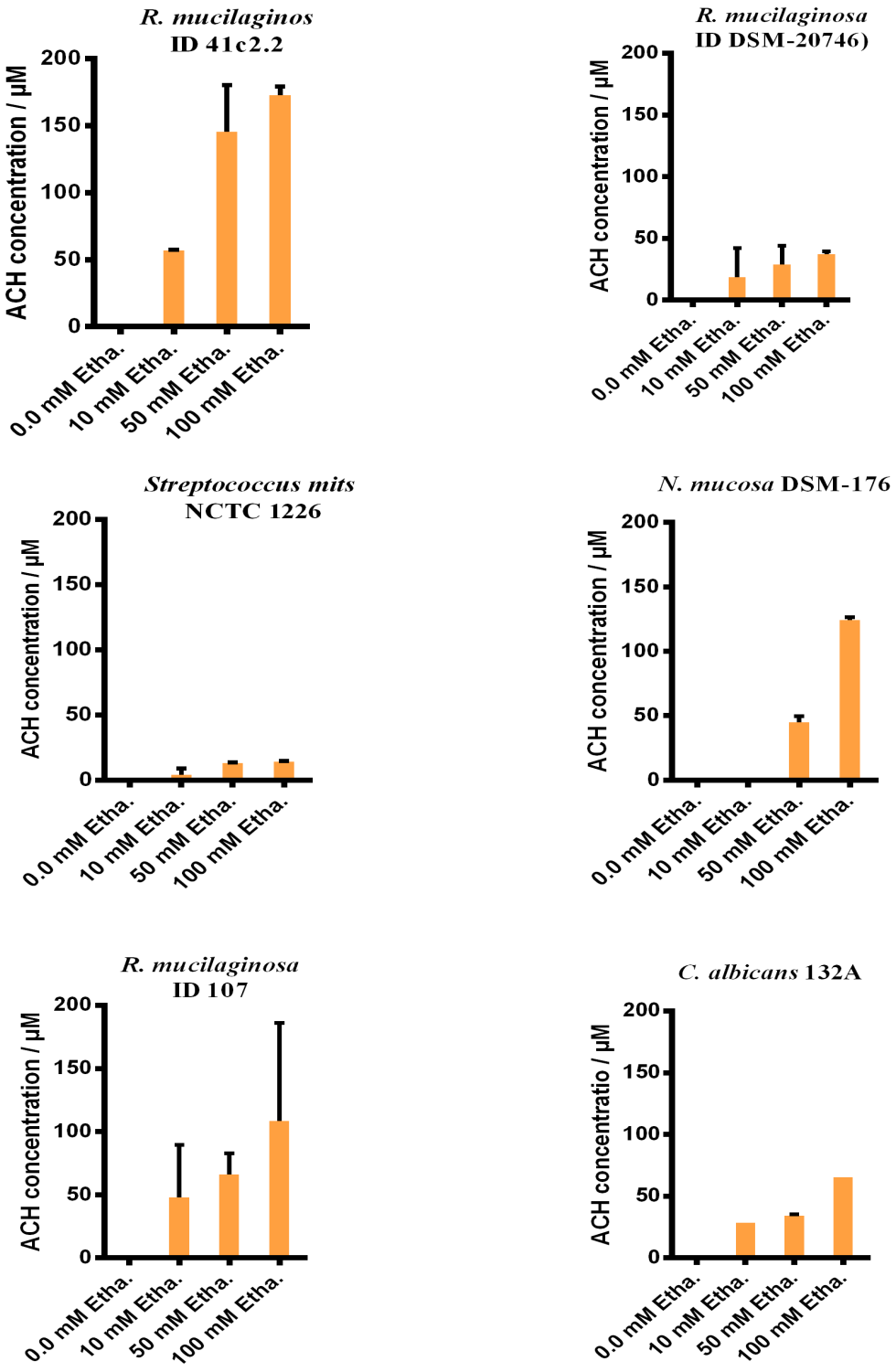



Figure 6.8 Levels of the acetaldehyde produced by *R. mucilaginosa* strains and other bacterial species and *C. albicans* from exponential cultures incubated in ethanol. Etha. = Ethanol

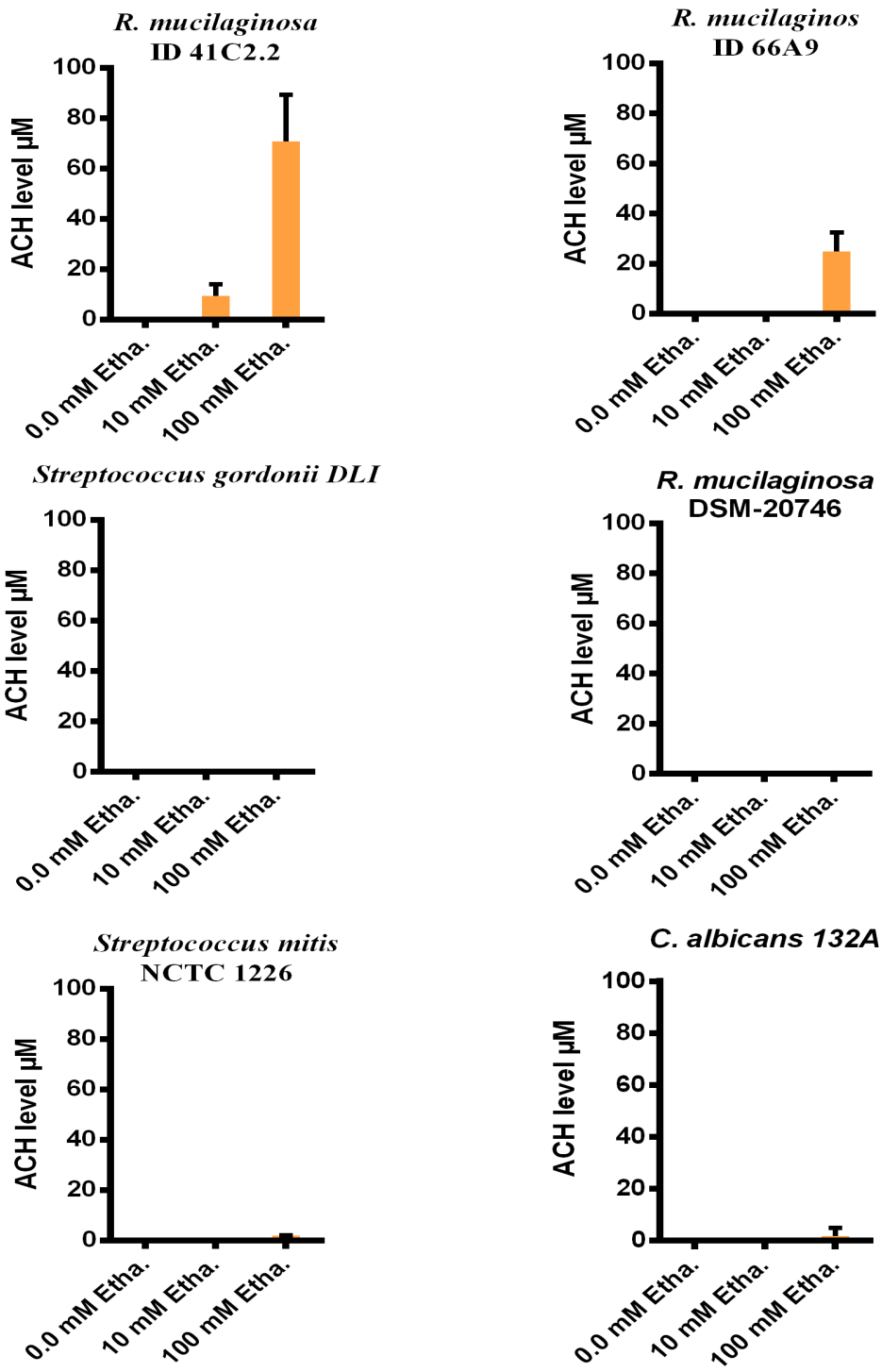
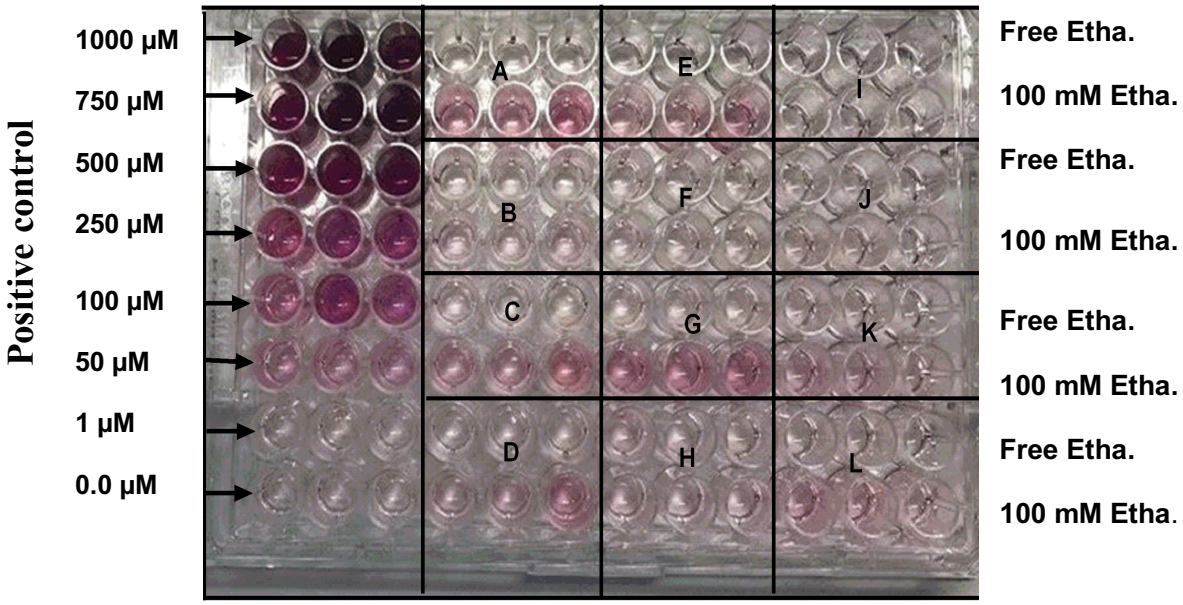



Figure 6.9 Showing the level of the acetaldehyde (ACH) produced by *R. mucilaginosa* strains and other bacterial species and *C. albicans* from stationary cultures incubated in ethanol.

Etha. = Ethanol




A = *R. mucilaginosa* 41C2.2 B = *R. mucilaginosa* 43B1 C = *R. mucilaginosa* 107 D = *R. mucilaginosa* DSC 20746

E = *R. mucilaginosa* H2.2 F = *R. mucilaginosa* H4.2 G = *R. mucilaginosa* 66A H = *R. mucilaginosa* 44A1

I = *S. mitis* NCTC 12261 G = *S. gordonii* DL1 K = *N. mucosa* DSM 17611 L = *C. albicans* 132A

**Etha. = Ethanol**

**Figure 6.10**a Shows the colorimetric assay plate for aldehyde detection

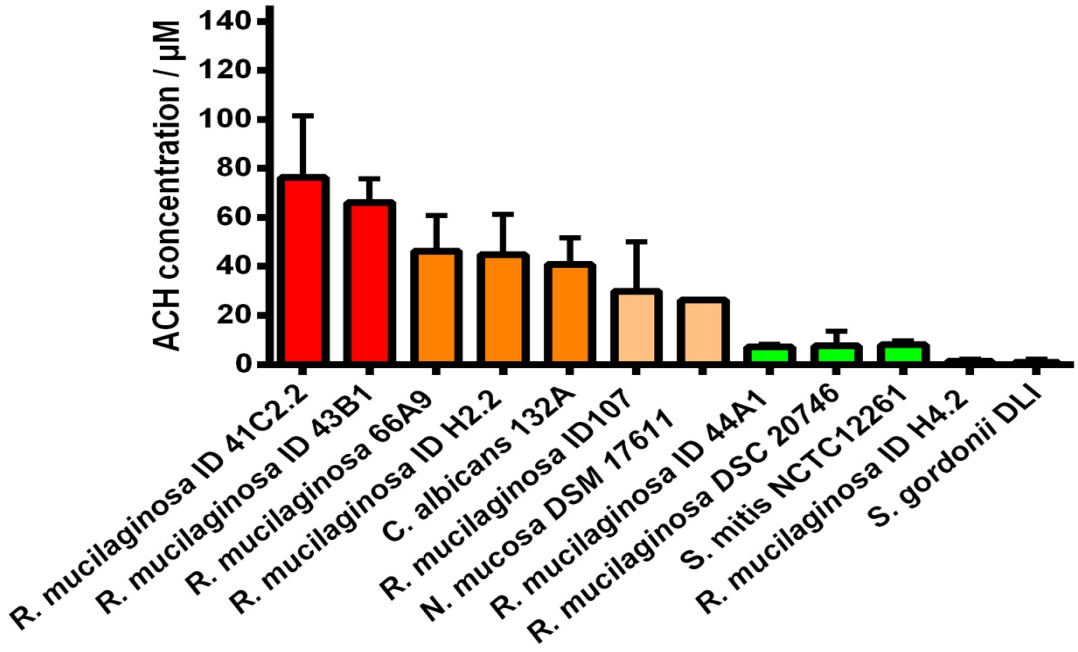



**Figure 6.10**b Shows the calculated amount of aldehyde produced by the indicated oral bacteria and *C. albicans*132A in 100 mM ethanol final concentration.

### 6.3.6 Gas chromatography-mass spectrometry

Gas chromatography-mass spectrometry (GC-MS) was carried out in a specialist laboratory at Dublin institute of technology (DIT) in order to precisely measure ACH rather than general acetaldehyde

levels. These experiments included *R. mucilaginosa* 41C2.2, *R. mucilaginosa* 66A9, *R. mucilaginosa* DY-18, *Neisseria mucosa* and *S. mitis* NCTC12261. The results showed the amount of the ACH was produced by *R. mucilaginosa* ID 41C2.2, and *Rothia mucilaginosa* 66A9 are significantly high (166 µM and 80 µM) respectively and are similar to the levels detected using the colorimetric kit. *R. mucilaginosa* DY-18 produced only 4 µM.

**Table 6.4** Showing the level of acetaldehyde produced following 2 h incubation in 100 mM ethanol estimation by measured by gas chromatography-mass spectrometry.

| Sample ID | Amount of ACH produced |
|---|---|
| *R. mucilaginosa* 41C2.2 | 166 µM |
| *R. mucilaginosa* 66A9 | 80 µM |
| *R. mucilaginosa* DY18 | 4 µM |
| *S. mitis* NCTC12261 | 1 µM |
| *Neisseria mucosa* DSM-17611 | 8 µM |

### 6.3.7 Effect of ACH on oxidative stress in oral keratinocytes

After measuring the level of ACH produced by *R. mucilaginosa* strains, it was then examined whether these levels were capable of inducing oxidative stress in the oral keratinocyte. I examined the effects on the oral keratinocyte cell line TR146. TR146 cells

were incubated with ACH at different concentrations, representative of the levels observed in previous experiments in 96 well plates. A 30 mM concentration of the $H_2O_2$ was used as positive control. Measurement of reactive oxygen species (ROS) production was assessed using 2',7'dichlorodihydrofluorescein diacetate (DCF-DA). N-acetylcysteine was used as an antioxidant protective agent to determine if fluorescence was ROS associated. Increasing levels of ACH resulted in increased DCF-DA fluorescence (Fig. 6.11). The addition of NAC reduced the level of ROS induced fluorescence. The highest concentration of ACH did not yield significantly higher levels of fluorescence, which may be due to the cell toxicity or death.

The effect of ACH at these levels was also examined using the fluorescence microscope. ACH was used to induce oxidative stress at 50 μM and 100 μM, and 30 mM of $H_2O_2$ was used as a positive control. N-acetylcysteine was again used as an antioxidant. Increasing ACH concentration resulted in increasing fluorescence, which could be reduced by the addition of NAC (Fig. 6.12).

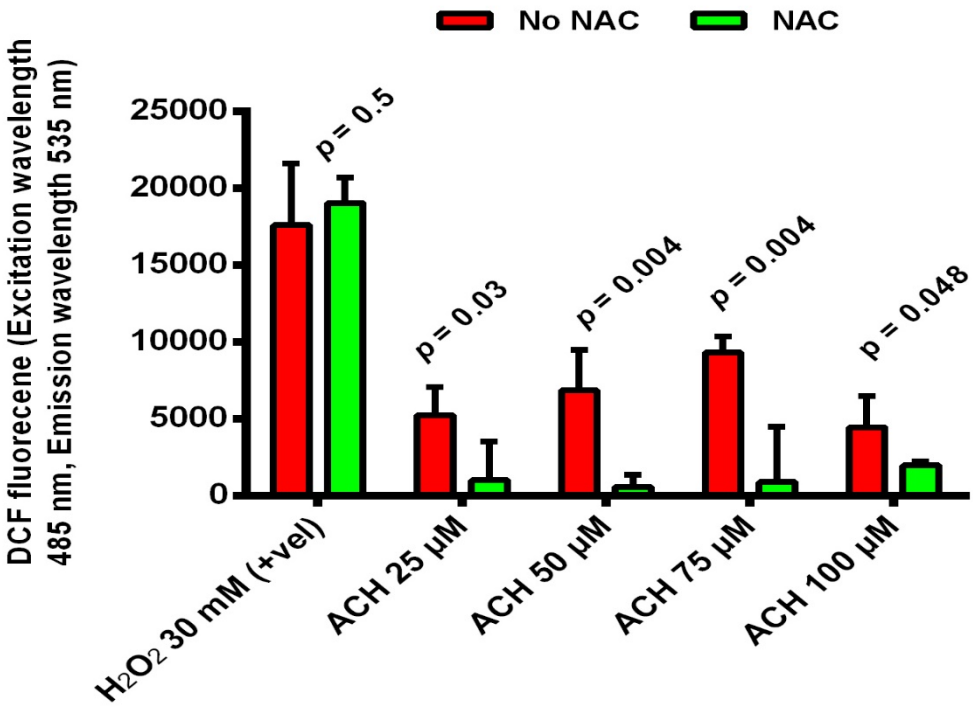



Figure 6.11 Acetaldehyde induced oxidative stress in TR146 oral keratinocytes. Cells were incubated with ACH in different concentrations. A concentration of 30 mM $H_2O_2$ was used as positive control. Serum free medium was used as a negative control. Reactive oxygen species (ROS) was induced using 2',7'Dichlorodihydrofluorescein diacetate (DCF-DA), and N-Acetylcysteine as an antioxidant

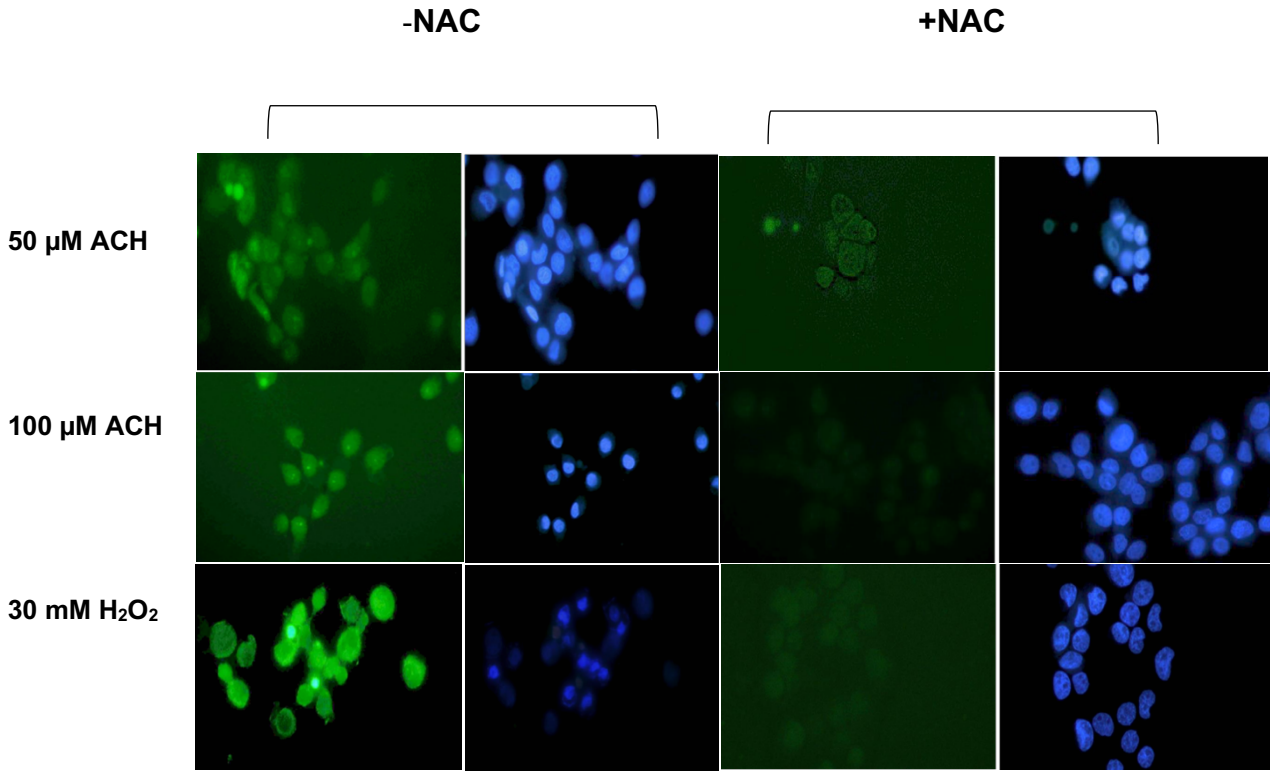



Figure 6.12 Shows TR146 cells strained with DCF-DA to detect ROS (green) and Hoechst (blue fluorescent dye to stain nuclei). Cells were in preincubated in 100 µM of DCF-DA with or without NAC. Cells were then exposed to stress (50 or 100 µM ACH or 30 mM $H_2O_2$) and visualised using a Zeiss epifluorescence microscope.

## 6.4 Discussion

Recently, analysis of microbiomes has begun to move away from analysis of simply what isolates are present to focus on the functions of the metagenome. In this chapter data on the structure of microbial communities based on 16S rRNA sequencing were used to predict the genes found in each type of samples. Using these data, PICRUSt allowed metagenome prediction in OLK and healthy mucosa. Analysis of OLK communities showed an enrichment for genes involved in lipopolysaccharide (LPS) biosynthesis. lipopolysaccharide is associated with Gram-negative bacteria and is recognised as a key factor in septic shock (septicemia). This result is perhaps not surprising due to the increased levels of Fusobacteria observed in these samples. It indicates that these isolates may have greater inflammatory potential which may play a role in malignant progression. Increased inflammatory potential has also been shown in biopsy samples recovered from OSCC (Perera *et al.,* 2018).

Acetaldehyde dehydrogenase genes (ACDH) were also overrepresented in contralateral normal tissue compared to OLK (Fig. 6.5). Analysis of common oral bacteria for ACDH genes showed that these genes were underrepresented in obligate anaerobes and the Actinobacteria. In *Rothia* spp., only 4 genes were present in 70 genomes according to the PATRIC database (Table 6.1) while in *Streptococcus* spp., a total of 479 ACDH encoding genes were found in 260 streptococcal genomes. We hypothesied that expression of these genes could convert the ACH produced by alcohol dehydrogenases to acetone/acetate. Reduced levels of these genes in the community could result in a community with less capacity to metabolise toxic levels of ACH. As *R. mucilaginosa* levels are greatly increased in certain OLK communities, this could result in a risk of ACH toxicity on OLK mucosa, perhaps contributing to mutation and malignant transformation.

Next, *R mucilaginosa* isolated from OLK patients and healthy controls were examined in order to determine their capacity to generate ACH from ethanol. In this study a total number of 11 *R. mucilaginosa* isolates were isolated from OLK patients and healthy subjects using a selective medium consisting of BHI supplemented with sodium Selenite and colistin. Presumptive strains were identified by 16S rRNA sequencing. However, many other species still have the ability to grow on this medium, resulting in the isolation of non-*Rothia* isolates. Further experiments may be required to develop a more selective media with the ability to isolate *R. mucilaginosa* strains more accurately.

A comparison of the growth curves was carried out using 10 microbial strains, including six *R. mucilaginosa* isolates, two streptococci (*S. mitis* NCTC12261 and *S. gordonii* DL1), *Neisseria mucosa* DSM 17611 and *C. albicans* 132A. All bacteria were inhibited at 4% ethanol final concentration with only *C. albicans* showing the ability to tolerate high ethanol concentrations. The streptococci were not greatly affected by the presence of 2% ethanol whereas the *R. mucilaginosa* strains and the *Neisseria mucosa* strain showed reduced growth. This may be due to the enhanced capacity of the streptococci to metabolise alcohol or its toxic oxidation products ACH to produce acetate.

Acetaldehyde is a highly reactive compound and classed as a carcinogenic substance (International Agency for Research on Cancer, IARC). Acetaldehyde is a mutagenic compound that can damage DNA structure at concentrations in the range 50-100 μM (Jokelainen *et al.,* 1996). These concentrations can be found in the oral cavity after alcohol consumption. Analysis of the *R. mucilaginosa* strains isolated here shows that most isolates were capable for producing acetaldehyde in the range of 50-100 μM. Similar levels were detected in microbes previously shown to generate ACH including *N. mucosa* and *C. albicans*. *R. mucilaginosa* 41C2.2, recovered from a patient with OLK produced the highest amount ACH detected at ~166 μM by GC-MS. Many of these *R. mucilaginosa*

isolates were recovered from patients with a history of alcohol consumption placing them potentially at risk of ACH exposure.

*R. mucilaginosa* is a part of the normal oral flora and can be found as part of the tongue microbiome. Recently Yokoyama *et al.* (2018), showed that individuals with the highest capacity to generate ACH in saliva has increased levels of *R. mucilaginosa* in their salivary microbiome. Our microbiome studies show that OLK tissues can harbor higher levels of *R. mucilaginosa* relative to contralateral normal tissue from the same patient. The metagenome analysis here shows that this shift in abundance reduces the potential capacity of the community to metabolise ACH due to reduced levels of ACDH coding capacity, which could lead to local build-up of ACH at the site of the OLK. If this hypothesis is true, it may contribute to the malignant transformation of these lesions. The data shown here indicates that these levels of ACH (50-100 μM) are capable of inducing oxidative stress in oral keratinocytes, indicating that this ACH production is biologically significant.

In conclusion, more studies are required on *Rothia* species and the biological significance of its ability to produce carcinogenic levels of the ACH and how these levels could affect the development of oral cancer. If this can be confirmed, strategies to control *Rothia* spp. could be useful in preventing transformation of OLK to OSCC.

In this study results showed that some common oral species are overrepresented on OLK including *Fusobacterium* spp. and *R. mucilaginosa,* amongst others. However, further work is required to support the hypothesis oral bacteria could play roles with other common risk factors such as tobacco and alcohol consumption in the transformation of OLK to OSCC.

Future studies could include: A longitudinal analysis of the microbiome of OLK to determine if certain microbial enrichments increase the risk of malignant transformation.

For example, this could determine if those patients with elevated levels of *F. nucleatum, Leptotrichia* and *Campylobacter* spp. are more likely to progress to OSCC than those who do not.

Metagenomic analysis of these communities could also reveal if specific metabolic functions are associated with the development of OSCC.

Culture based studies are required to verify the enrichments and characterise the bacteria implicated in these studies. In this study I have shown that *R. mucilaginosa* can generate acetaldehyde *in vitro*. Significant variation in the ability of the small number of isolates examined here was detected. Further analysis, perhaps using whole genome sequencing may identify the basis of these difference and determine if certain strains are more associated with acetaldehyde generation.

The effects of *F. nucleatum* on oral keratinocytes are relatively unknown. Studies are required to determine if this species can activate cellular proliferation pathways in oral cells or induce morphological changes associated with malignant transformation. The effects of mixed communities of organisms including *Leptotrichia* spp, *Campylobacter* spp. and *C. albicans* together is also unexplored the effect of these mixed communities on oral cells, perhaps using RNAseq analysis could be explored.

The long-term benefits of these studies could impact upon the development of OSCC from OLK. At present there are no treatments for OLK. If these species can be implicated in the malignant transformation process, it may be possible to cure or control OLK with topical antibiotic therapy. It may also be possible to identify the patients at high risk of malignant transformation based on the microbiome and the presence of certain species.

# Chapter 7

General discussion

## 7.1 General discussion

### 7.1.1 Summary of the results

The present study was designed to investigate if there are differences in oral bacterial community structure between oral leukoplakia and healthy mucosal tissues from contralateral normal sites and healthy controls. Influence of *Candida* carriage and common risk factors such as smoking and alcohol consumption were also investigated.

In the results discussed previously in Chapters 3 and 4, bacterial populations from the buccal cavities of healthy patients were significantly different to OLK patients (Table 3.3, AMOVA $p < 0.025$). In the case of lingual OLK populations, these exhibited differences in population structure to contralateral samples from the same patient and healthy controls (Table 3.4, AMOVA $p < 0.025$). A complex pattern of enrichments was identified on OLK tissues and the main taxa identified in Chapters 3 and 4 as being OLK associated are summarised in Table 7.1. The most consistent enrichments identified in the analysis of buccal and lingual samples using both V1-V2 and V3-V4 sequences were *R. mucilaginosa*, *F. nucleatum* and *Campylobacter* spp. These finding have some similarity to previous studies of OSCC which showed that *F. nucleatum* and *Campylobacte*r spp. are enriched in biopsies and swabs of OSCC. Conversely, *Rothia* spp. are generally less abundant in OSCC compared to healthy mucosa.

In the following sections, the possible involvement of these organisms in the malignant progression of OLK and future work that needs to be carried out in order to validate these hypotheses will be discussed.

**Table 7.1** Showing all bacteria species overrepresented in patient leukoplakia samples and contralateral normal identified by LEfSe analysis from V1-V2 regoins compared to V3-V4 regions . Taxa in red appear in both datasets.

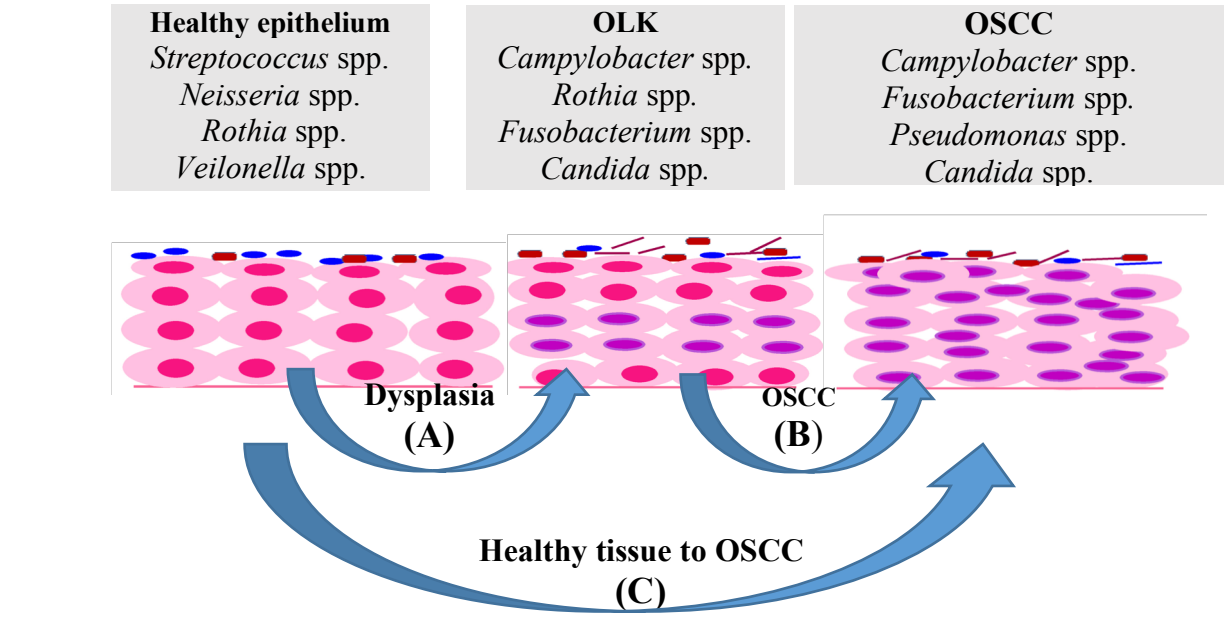
| **Site** | **Samples** | **V1-V2** | **V3-V4** |
|---|---|---|---|
| **BUCCAL** | All patient (OLK, CN) | *Rothia mucilaginosa* | *Rothia mucilaginosa* |
| | | *Alloprevotella* spp. | *F. nucleatum* |
| | | Bacteroidetes | *Staphylococcus* spp. |
| | | *Neisseria meningitidis* | *Campylobacter* spp. |
| | | *Leptotrichia* spp. | *Haemophilus* spp. |
| | OLK | *Fusobacterium* spp. | *Veillonella* spp. |
| | | Flavobacteriaceae | |
| | | *Campylobacter* spp. | |
| **LINGUAL** | All patient (OLK, CN) | *Alloprevotella* HOT473 | *Prevotella* spp. |
| | | *Neisseria meningitis* | |
| | | *Leptotrichia* spp. OTU081 | |
| | OLK | *Fusobacterium* spp. | *F. nucleatum* |
| | | *Streptococcu*s spp. | Fusobacteria |
| | | *Capnocytophaga* spp. | Porphyromonodaceae |

### 7.1.2 Common bacteria associated with OLK

Oral mucosal surface is comprised by a complex bacterial community which exists in homeostasis with the host. Keratinocyte shedding and local immunity keeps this population under control. The early stages of dysplasia may change the mucosal surface, altering the capacity of this normal microbiome to adhere. These early changes may be induced by smoking or by alcohol consumption (Fig. 7.1). Acetaldehyde production by oral microbes may have a role to play at this early stage as healthy individuals with high levels of *R. mucilaginosa* have a high capacity to produce acetaldehyde in saliva (Yokoyama *et al.,* 2018). During the development of dysplasia, certain species that can colonise the abnormal tissue increase in abundance, and therefore increasing the level of their products. In Chapter 3 (Fig. 3.18) and Chapter 4 (Fig. 4.15) *R. mucilaginosa* was highly associated with OLK, suggesting the hypothesis that these bacteria could increase the level of dysplasia by local production of acetaldehyde. Results of this study showed that *R. mucilaginosa* can produce high levels of ACH, more than 100 μM *in vitro*., while analysised of the *R. mucilaginosa* genome showed this organism does not encode for acetaldehyde dehydrogenase, (Table 6.1) leading to accumulation of toxic levels of acetaldehyde *in vitro*.

The development of dysplasia results in changes to the mucosa that may impact on the types of bacteria that can adhere to the surface of the tissue. Our data indicate that several Gram negative species increase in abundance at this stage (Fig. 7.1). This included *F. nucleatum*, *Leptotrichia* spp. and *Campylobacter* spp. Dysplasia is associated with abnormal features in the epithelium, such as loss of cellular cohesion. The normal oral epithelial cells are linked together by tight cell-cell adhesion mediated by cadherin (Calcium-Dependent Adhesion molecules), which play an important role in cell-cell interactions and during cell growth and division. Recent studies have shown that Gram-

negative bacteria including *F. nucleatum* can decrease expression of E-cadherin in oral keratinocytes (Abdulkareem *et al.,* 2018).




**(A)** Smoking and alcohol induced dysplasia. This may be driven by toxins in cigarette smoke, microbial acetaldehyde production and or tissue invasion by *Candida* spp. Poor diet may increase susceptibility.

**(B)** Increased levels of Gram-negative colonisation on OLK induces inflammation. colonisation may be influenced by poor oral hygiene, smoking and Acetaldehyde production. Drives malignant transformation of ~10% of OLK.

**(C)** Some OSCC arises *de novo*, this may involve genetic predisposition in tandem with smoking, alcohol and microbial factors.

Figure 7.1 Possible mechanisms of transformation for healthy mucosa to dysplasia and OSCC.

This property was associated with the presence of gene associated with LPS biosynthesis were significantly increased in communities from OLK (Fig. 6.3). Invasion of oral keratinocytes by oral bacteria may also impact on cellular proliferation. As described in Chapter 1, *F. nucleatum* can activate β-catenin signalling which may increase cell growth and proliferation in colonic epithelial cells. Similar processes may occur in oral keratinocytes and could drive the progression of OLK to OSCC.

The level of Gram-negative bacteria may be influenced by lifestyle factors such as oral hygiene, smoking and alcohol consumption. For example, in our data, non-smoking patients had higher levels of *F. nucleatum* OTU016 relative to smokers (Fig. 5.5b). Interestingly, two studies show that non-smoking OLK patients are more likely to progress to OSCC, and we hypothesis that this may be due to the higher levels of *F. nucleatum* in these patients (Silverman *et al.,* 1984; Schepman *et al.,* 1998). Alcohol consumption was also shown to influence the level of another Gram-negative organism, *Campylobacter gracilis* OUT 0075 Samples from patients who consume alcohol were found to have increased levels of *C. gracilis* (Fig.5.8).

The presence of *Candida* species has been associated with OLK in many previous studies. In this present study, 35% of OLK samples were positive for *Candida* spp., by PCR compared to 20% of contralateral samples and 11% of samples from healthy controls. *Candida* spp. may also produce acetaldehyde and other toxic compounds such as nitrosamines. Interestingly, the presence of *Candida* may also influence the bacterial microbiome. *Candida* spp. positive patients were also found to have higher levels of *F. nucleatum*, *Porphyromonas pasteri* and *Leptotrichia* HOT215. These communities could have additive effects on oral cells to drive malignant transformation. *Candida albicans* and *C. dubliniensis* have been shown to co-aggregate with *F. nucleatum* and the presence

of one of these organisms may promote colonisation with the other through co-aggregation (Jabra *et al.,* 1999).

The generated in this current study support a role for host environment and antimicrobial interactions in shaping the microbiome of OLK. Analysis of the samples with severe dysplasia in (Fig. 5.12), showed that a community including *Leptotrichia* spp. and *Campylobacter* spp. was most strongly associated with severe dysplasia. Studies of colonic cancers and OSCC have shown that these organisms are also present in malignant tissue (Perera and Al-Hebshi, 2018). Schmidt *et al.* (2014) showed that levels of *R. mucilaginosa* are reduced in OSCC relative to healthy tissue, indicating that this species enrichment is lost in the transformation from OLK to OSCC. However, in general, our data identify species enrichments on OLK which have also been identified in OSCC, indicating that these species may play a role in the transformation of OLK to OSCC. Although the data in this study showed an interesting association of these species with OLK, further work will be required in order to demonstrate if these organisms are associated with malignant transformation.

## 7.2 Future work

To our knowledge, this is the first study to compare the oral microbiome from OLK sites and contralateral normal tissue in the same patient. such studies are just the beginning of investigations into the role of the microbiome in the development of OSCC from OLK. Our study shows that some common oral species are overrepresented on OLK including *Fusobacterium* spp. and *R. mucilaginosa*. However, further work is required to support the hypothesis that oral bacteria could play roles with other common risk factors such as smoking and alcohol consumption in the transformation of OLK to OSCC.

Future studies could include:

A longitudinal analysis of the microbiome of OLK to determine if certain microbial enrichments increase the risk of malignant transformation. For example, this could determine if those patients with elevated levels of *F. nucleatum, Leptotrichia* and *Campylobacter* spp. are more likely to progress to OSCC. Metagenomic analysis of these communities could also reveal if specific metabolic functions are associated with the development of OSCC. Culture-based studies are required to verify the enrichments and characterise the bacteria implicated in these studies. This study has that *R. mucilaginosa* can generate acetaldehyde *in vitro*. Significant variation in the ability of the small number of isolates examined here was detected. Further analysis, perhaps using whole genome sequencing may identify the basis of these differences and determine if certain strains are more associated with acetaldehyde generation.

The effects of *F. nucleatum* on oral keratinocytes are relatively unknown. Studies are required to determine if this species can activate cellular proliferation pathways in oral cells or induce morphological changes associated with malignant transformation. The effects of mixed communities of organisms including *Leptotrichia* spp, *Campylobacter*

spp. and *C. albicans* together is also unexplored and the effect of these mixed communities on oral cells, perhaps using RNAseq analysis could be explored.

The long-term benefits of these studies could impact upon the development of OSCC from OLK. At present there are no treatments for OLK. If these species can be implicated in the malignant transformation process, it may be possible to cure or control OLK with topical antibiotic therapy. It may also be possible to identify patients at high risk of malignant transformation based on the microbiome and the presence of certain species.

# References


1. Aas, J. A., B. J. Paster, L. N. Stokes, I. Olsen and F. E. Dewhirst (2005). "Defining the normal bacterial flora of the oral cavity." J Clin Microbiol 43(11): 5721-5732.
2. Abdulrahim, M. H., B. A. McManus, S. R. Flint and D. C. Coleman (2013). "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(9): e73738.
3. Abidullah M., Kiran G., Gaddikeri K., Raghoji S. & Ravishankar T. S. (2014) "Leukoplakia-Review of potentially malignant disorder." J Clin Diagn Res 8: ZE01–ZE04
4. Al-Hebshi, N. N., A. T. Nasher, M. Y. Maryoud, H. E. Homeida, T. Chen, A. M. Idris and N. W. Johnson (2017). "Inflammatory bacteriome featuring *Fusobacterium nucleatum* and *Pseudomonas aeruginosa* identified in association with oral squamous cell carcinoma." Sci Rep 7(1): 1834.
5. Al-Hebshi, N. N., W. S. Borgnakke and N. W. Johnson (2019). "The microbiome of oral squamous cell carcinomas: a functional perspective." Curr oral Health Rep 6(2): 145-160.
6. Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." J Mol Biol 215(3): 403-410.
7. Blaut, M. and T. Clavel (2007). "Metabolic diversity of the intestinal microbiota: implications for health and disease." J Nutr 137(3 Suppl 2): 751s-755s.
8. Bouda, M., V. G. Gorgoulis, N. G. Kastrinakis, A. Giannoudis, E. Tsoli, D. Danassi-Afentaki, P. Foukas, A. Kyroudi, G. Laskaris, C. S. Herrington and C. Kittas (2000). ""High risk" HPV types are frequently detected in potentially

malignant and malignant oral lesions, but not in normal oral mucosa." Mod Pathol 13(6): 644-653.

9. Brown, L. M., D. P. Check and S. S. Devesa (2012). "Oral cavity and pharynx cancer incidence trends by subsite in the United States: changing gender patterns." J Oncol 2012: 649498.
10. Bullman, S., C. S. Pedamallu, E. Sicinska, T. E. Clancy, X. Zhang, D. Cai, D. Neuberg, K. Huang, F. Guevara, T. Nelson, O. Chipashvili, T. Hagan, M. Walker, A. Ramachandran, B. Diosdado, G. Serna, N. Mulet, S. Landolfi, Y. C. S. Ramon, R. Fasani, A. J. Aguirre, K. Ng, E. Elez, S. Ogino, J. Tabernero, C. S. Fuchs, W. C. Hahn, P. Nuciforo and M. Meyerson (2017). "Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer." Science 358(6369): 1443-1448.
11. Canabarro, A., C. Valle, M. R. Farias, F. B. Santos, M. Lazera and B. Wanke (2013). "Association of subgingival colonisation of *Candida albicans* and other yeasts with severity of chronic periodontitis." J Periodontal Res 48(4): 428-432.
12. Castellarin, M., R. L. Warren, J. D. Freeman, L. Dreolini, M. Krzywinski, J. Strauss, R. Barnes, P. Watson, E. Allen-Vercoe, R. A. Moore and R. A. Holt (2012). "*Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma." Genome Res 22(2): 299-306.
13. Curry, J. M., J. Johnson, M. Mollaee, P. Tassone, D. Amin, A. Knops, D. Whitaker-Menezes, M. G. Mahoney, A. South, U. Rodeck, T. Zhan, L. Harshyne, N. Philp, A. Luginbuhl, D. Cognetti, M. Tuluc and U. Martinez-Outschoorn (2018). "Metformin Clinical Trial in HPV+ and HPV- Head and neck squamous cell carcinoma: Impact on cancer cell apoptosis and immune infiltrate." Front Oncol 8: 436.

14. Donders, G. G., E. Bosmans, A. Dekeersmaecker, A. Vereecken, B. Van Bulck and B. Spitz (2000). "Pathogenesis of abnormal vaginal bacterial flora." Am J Obstet Gynecol 182(4): 872-878.
15. Dong, Y., Q. Zhao, X. Ma, G. Ma, C. Liu, Z. Chen, L. Yu, X. Liu, Y. Zhang, S. Shao, J. Xiao, J. Li, W. Zhang, M. Fu, L. Dong, X. Yang, X. Guo, L. Xue, F. Fang, Q. Zhan and L. Zhang (2015). "Establishment of a new OSCC cell line derived from OLK and identification of malignant transformation-related proteins by differential proteomics approach." Sci Rep 5: 12668.
16. Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince and R. Knight (2011). "UCHIME improves sensitivity and speed of chimera detection." Bioinformatics 27(16): 2194-2200.
17. Elamin, E., A. Masclee, F. Troost, J. Dekker and D. Jonkers (2014). "Cytotoxicity and metabolic stress induced by acetaldehyde in human intestinal LS174T goblet-like cells." Am j physiol gastrointest liver physiol 307(3): G286-294.
18. Fabian, T. K., P. Hermann, A. Beck, P. Fejerdy and G. Fabian (2012). "Salivary defense proteins: their network and role in innate and acquired oral immunity." Int J Mol Sci 13(4): 4295-4320.
19. Feller, L. and J. Lemmer (2012). "Oral leukoplakia as it relates to HPV infection: a review." Int J dent 2012: 540561.
20. Ferlay, J., E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J. W. Coebergh, H. Comber, D. Forman and F. Bray (2013). "Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012." Eur J Cancer 49(6): 1374-1403.
21. Findley, K. and E. A. Grice (2014). "The skin microbiome: a focus on pathogens and their association with skin disease." PLoS Pathog 10(10): e1004436.

22. Frank, J. A., C. I. Reich, S. Sharma, J. S. Weisbaum, B. A. Wilson and G. J. Olsen (2008). "Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes." Appl Environ Microbiol 74(8): 2461-2470.

23. Gallagher, P. J., D. E. Bennett, M. C. Henman, R. J. Russell, S. R. Flint, D. B. Shanley and D. C. Coleman (1992). "Reduced azole susceptibility of oral isolates of *Candida albicans* from HIV-positive patients and a derivative exhibiting colony morphology variation." J Gen Microbiol 138(9): 1901-1911.

24. Gao, L., T. Xu, G. Huang, S. Jiang, Y. Gu and F. Chen (2018). "Oral microbiomes: more and more importance in oral cavity and whole body." Protein Cell 9(5): 488-500.

25. Ganesh, D., P. Sreenivasan, J. Ohman, M. Wallstrom, P. H. Braz-Silva, D. Giglio, G. Kjeller and B. Hasseus (2018). "Potentially Malignant Oral Disorders and Cancer Transformation." Anticancer Res 38(6): 3223-3229.

26. Ghannoum, M. A., R. J. Jurevic, P. K. Mukherjee, F. Cui, M. Sikaroodi, A. Naqvi and P. M. Gillevet (2010). "Characterisation of the oral fungal microbiome (mycobiome) in healthy individuals." PLoS Pathog 6(1): e1000713.

27. Gillum, A. M., E. Y. Tsay and D. R. Kirsch (1984). "Isolation of the *Candida albicans* gene for orotidine-5'-phosphate decarboxylase by complementation of S. cerevisiae ura3 and E. coli pyrF mutations." Mol Gen Genet 198(2): 179-182

28. Graver, M. A. and J. J. Wade (2011). "The role of acidification in the inhibition of *Neisseria gonorrhoeae* by vaginal lactobacilli during anaerobic growth." Ann Clin Microbiol Antimicrob 10: 8.

29. Gupta, N., R. Gupta, A. K. Acharya, B. Patthi, V. Goud, S. Reddy, A. Garg and A. Singla (2016). "Changing trends in oral cancer - a global scenario." Nepal J epidemiol 6(4): 613-619.

30. Ha, E., F. Kim, J. Blanchard and H. S. Juon (2019). "Prevalence of chronic hepatitis B and C infection in mongolian immigrants in the Washington, District of Columbia, Metropolitan area, 2016-2017." Prev Chronic Dis 16: E08.
31. He, X. S. and W. Y. Shi (2009). "Oral microbiology: past, present and future." Int J Oral Sci 1(2): 47-58.
32. Herrera-Goepfert, R., M. Lizano, S. Akiba, A. Carrillo-Garcia and M. Becker-D'Acosta (2009). "Human papilloma virus and esophageal carcinoma in a Latin-American region." World J Gastroenterol 15(25): 3142-3147.
33. Herwald, S. E. and C. A. Kumamoto (2014). "*Candida albicans* Niche Specialization: Features That Distinguish Biofilm Cells from Commensal Cells." Curr Fungal Infect Rep 8(2): 179-184.
34. Ikebe, M., Y. Kitaura, M. Nakamura, H. Tanaka, A. Yamasaki, S. Nagai, J. Wada, K. Yanai, K. Koga, N. Sato, M. Kubo, M. Tanaka, H. Onishi and M. Katano (2009). "Lipopolysaccharide (LPS) increases the invasive ability of pancreatic cancer cells through the TLR4/MyD88 signaling pathway." J Surg Oncol 100(8): 725-731.
35. Inoue, A., A. Sakamoto, M. Tsuzuku, N. Kasai, A. Noguchi, S. Kamata, M. Uchida, M. Hirata and S. Harashima (1985). "Early invasive squamous cell carcinoma of the tongue--report of a case." Gan No Rinsho 31(1): 83-86.
36. Jabra-Rizk, M. A., W. A. Falkler, Jr., W. G. Merz, J. I. Kelley, A. A. Baqui and T. F. Meiller (1999). "Coaggregation of *Candida dubliniensis* with *Fusobacterium nucleatum*." J Clin microbiol 37(5): 1464-1468.
37. Janda, J. M. and S. L. Abbott (2007). "16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls." J Clin Microbiol 45(9): 2761-2764.

38. Jokelainen, K., T. Matysiak-Budnik, H. Makisalo, K. Hockerstedt and M. Salaspuro (1996). "High intracolonic acetaldehyde values produced by a bacteriocolonic pathway for ethanol oxidation in piglets." Gut 39(1): 100-104.
39. Jones, R. R., P. J. Weyer, C. T. DellaValle, M. Inoue-Choi, K. E. Anderson, K. P. Cantor, S. Krasner, K. Robien, L. E. Freeman, D. T. Silverman and M. H. Ward (2016). "Nitrate from Drinking Water and Diet and Bladder Cancer Among Postmenopausal Women in Iowa." Environ Health Perspect 124(11): 1751-1758.
40. Khalil Boutaga (2003). "Comparison of Real-Time PCR and culture for detection of *Porphyromonas gingivalis* in subgingival plaque sample."J Clin Microbiol 41(11):4950-4954.
41. Kelly, D., L. Yang and Z. Pei (2018). "Gut Microbiota, Fusobacteria, and Colorectal Cancer." Diseases 6(4).
42. Kobayashi, T., S. Uchibori, O. Tsuzukibashi, H. Goto and M. Aida (2012). selective medium for *Rothia mucilaginosa* and its distribution in oral cavities." J Microbiol Methods 91(3):364-365.
43. Kolenbrander, P. E., R. N. Andersen, D. S. Blehert, P. G. Egland, J. S. Foster and R. J. Palmer, Jr. (2002). "Communication among oral bacteria." Microbiol Mol Biol Rev 66(3): 486-505.
44. Kostic, A. D., D. Gevers, C. S. Pedamallu, M. Michaud, F. Duke, A. M. Earl, A. I. Ojesina, J. Jung, A. J. Bass, J. Tabernero, J. Baselga, C. Liu, R. A. Shivdasani, S. Ogino, B. W. Birren, C. Huttenhower, W. S. Garrett and M. Meyerson (2012). "Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma." Genome Res 22(2): 292-298.

45. Kraneveld, E. A., M. J. Buijs, M. J. Bonder, M. Visser, B. J. Keijser, W. Crielaard and E. Zaura (2012). "The relation between oral *Candida* load and bacterial microbiome profiles in Dutch older adults." PLoS One 7(8): e42770.

46. Kruse, A. L., M. Bredell and K. W. Gratz (2010). "Oral squamous cell carcinoma in non-smoking and non-drinking patients." Head neck oncol 2: 24.

47. Kuczynski, J., C. L. Lauber, W. A. Walters, L. W. Parfrey, J. C. Clemente, D. Gevers and R. Knight (2011). "Experimental and analytical tools for studying the human microbiome." Nat Rev Genet 13(1): 47-58.

48. Kumar, P. S., C. R. Matthews, V. Joshi, M. de Jager and M. Aspiras (2011). "Tobacco smoking affects bacterial acquisition and colonisation in oral biofilms." Infect Immun 79(11): 4730-4738.

49. Lamont, R. J., A. El-Sabaeny, Y. Park, G. S. Cook, J. W. Costerton and D. R. Demuth (2002). "Role of the *Streptococcus gordonii* SspB protein in the development of *Porphyromonas gingivalis* biofilms on streptococcal substrates." Microbiology 148 ( 6): 1627-1636.

50. Langille, M. G., J. Zaneveld, J. G. Caporaso, D. McDonald, D. Knights, J. A. Reyes, J. C. Clemente, D. E. Burkepile, R. L. Vega Thurber, R. Knight, R. G. Beiko and C. Huttenhower (2013). "Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences." Nat Biotechnol 31(9): 814-821.

51. Ma, B., L. J. Forney and J. Ravel (2012). "Vaginal microbiome: rethinking health and disease." Annu Rev Microbiol 66: 371-389.

52. Markopoulos A. K. (2012). "Current aspects on oral squamous cell carcinoma." Open Dent J 6:126–130.

53. Mager, D, (2005). "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 Translat Med 3:27.

54. Marttila E., Uittamo J., Rusanen P., Lindqvist C., Salaspuro M., Rautemaa R. (2013). "Acetaldehyde production and microbial colonization in oral squamous cell carcinoma and oral lichenoid disease." Oral Surg Oral Med Oral Pathol Oral Radiol 116:61–68.

55. McCullough, M., M. Jaber, A. W. Barrett, L. Bain, P. M. Speight and S. R. Porter (2002). "Oral yeast carriage correlates with presence of oral epithelial dysplasia." Oral Oncol 38(4): 391-393.

56. Mehta, R. S., R. Nishihara, Y. Cao, M. Song, K. Mima, Z. R. Qian, J. A. Nowak, K. Kosumi, T. Hamada, Y. Masugi, S. Bullman, D. A. Drew, A. D. Kostic, T. T. Fung, W. S. Garrett, C. Huttenhower, K. Wu, J. A. Meyerhardt, X. Zhang, W. C. Willett, E. L. Giovannucci, C. S. Fuchs, A. T. Chan and S. Ogino (2017). "Association of dietary Patterns with risk of colorectal cancer Subtypes classified by *Fusobacterium nucleatum* in tumor tissue." JAMA Oncol 3(7): 921-927.

57. Meurman, J. H. (2010). "Oral microbiota and cancer." J Oral Microbiol 2:5195.

58. Mills A. L., Alexander M. (1976). "N-Nitrosamine formation by cultures of several microorganisms." Appl Environ Microbiol 31 892–895.

59. Mobeen, N. (2014). "Oral Candidiasis - a short review." Int J Cur Res Rev 6(9): 89-92.

60. Mohd Bakri, M., H. Mohd Hussaini, A. Rachel Holmes, R. David Cannon and A. Mary Rich (2010). "Revisiting the association between candidal infection and carcinoma, particularly oral squamous cell carcinoma." J Oral Microbiol 2:5780.

61. Moritani, K., T. Takeshita, Y. Shibata, T. Ninomiya, Y. Kiyohara and Y. Yamashita (2015). "Acetaldehyde production by major oral microbes." Oral Dis 21(6): 748-754.

62. Mork, J., A. K. Lie, E. Glattre, G. Hallmans, E. Jellum, P. Koskela, B. Moller, E. Pukkala, J. T. Schiller, L. Youngman, M. Lehtinen and J. Dillner (2001). "Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck." N Engl J Med 344(15): 1125-1131.

63. Mortazavi, H., M. Baharvand and M. Mehdipour (2014). "Oral potentially malignant disorders: an overview of more than 20 entities." J Dent Res Dent Clin Dent Prospects 8(1): 6-14.

64. Mortazavi, H., M. Baharvand and M. Mehdipour (2014). "Oral potentially malignant disorders: an overview of more than 20 entities." J Dent Res Dent Clin Dent Prospects 8(1): 6-14.

65. Muscat, J. E., J. P. Richie, Jr., S. Thompson and E. L. Wynder (1996). "Gender differences in smoking and risk for oral cancer." Cancer Res 56(22): 5192-5197.

66. Myburgh, N. G., M. H. Hobdell and R. Lalloo (2004). "African countries propose a regional oral health strategy: The dakar r§eport from 1998." Oral Dis 10(3): 129-137.

67. Nagy, K. N., I. Sonkodi, I. Szoke, E. Nagy and H. N. Newman (1998). "The microflora associated with human oral carcinomas." Oral Oncol 34(4): 304-308.

68. Nakatsuji, T., T. H. Chen, S. Narala, K. A. Chun, A. M. Two, T. Yun, F. Shafiq, P. F. Kotol, A. Bouslimani, A. V. Melnik, H. Latif, J. N. Kim, A. Lockhart, K. Artis, G. David, P. Taylor, J. Streib, P. C. Dorrestein, A. Grier, S. R. Gill, K. Zengler, T. R. Hata, D. Y. Leung and R. L. Gallo (2017). "Antimicrobials from

human skin commensal bacteria protect against *Staphylococcus aureus* and are deficient in atopic dermatitis." Sci Transl Med 9(378).

69. Nambu, T., K. Yamane, T. Yamanaka, C. Mashimo, H. Maruyama, M. Yoshida, H. Hayashi, K. P. Leung and H. Fukushima (2013). "Identification of disulphide stress-responsive extracytoplasmic function sigma factors in *Rothia mucilaginosa*." Arch Oral Biol 58(6): 681-689.
70. Neville, B. W. and T. A. Day (2002). "Oral cancer and precancerous lesions." CA Cancer J Clin 52(4): 195-215.
71. Nie, S., B. Tian, X. Wang, D. H. Pincus, M. Welker, K. Gilhuley, X. Lu, Y. W. Han and Y. W. Tang (2015). "*Fusobacterium nucleatum* subspecies identification by matrix-assisted laser desorption ionization-time of flight mass spectrometry." 53(4): 1399-1402.
72. Paster, B. J., S. K. Boches, J. L. Galvin, R. E. Ericson, C. N. Lau, V. A. Levanos, A. Sahasrabudhe and F. E. Dewhirst (2001). "Bacterial diversity in human subgingival plaque." J Bacteriol 183(12): 3770-3783.
73. Perera, M., N. N. Al-Hebshi, I. Perera, D. Ipe, G. C. Ulett, D. J. Speicher, T. Chen and N. W. Johnson (2018). "Inflammatory Bacteriome and oral squamous cell carcinoma." J Dent Res 97(6): 725-732.
74. Peters, B. A., J. Wu, R. B. Hayes and J. Ahn (2017). "The oral fungal mycobiome: characteristics and relation to periodontitis in a pilot study." BMC Microbiol 17(1): 157.
75. Pindborg, J. J., J. Reibel, B. Roed-Peterson and F. S. Mehta (1980). "Tobacco-induced changes in oral leukoplakic epithelium." Cancer 45(9): 2330-2336.

76. Pushalkar, S., S. P. Mane, X. Ji, Y. Li, C. Evans, O. R. Crasta, D. Morse, R. Meagher, A. Singh and D. Saxena (2011). "Microbial diversity in saliva of oral squamous cell carcinoma." FEMS Immunol Med Microbiol 61(3): 269-277.

77. Rajilic-Stojanovic, M. and W. M. de Vos (2014). "The first 1000 cultured species of the human gastrointestinal microbiota." FEMS Microbiol Rev 38(5): 996-1047.

78. Ramanan, P., J. N. Barreto, D. R. Osmon and P. K. Tosh (2014). "Rothia bacteremia: a 10-year experience at Mayo Clinic, Rochester, Minnesota." J Clin Microbiol 52(9): 3184-3189.

79. Saigal, S., A. Bhargava, S. K. Mehra and F. Dakwala (2011). "Identification of *Candida albicans* by using different culture medias and its association in potentially malignant and malignant lesions." Contemp Clin Dent 2(3): 188-193.

80. Sanders, H. L. (1968). "Marine Benthic Diversity: A Comparative Study." The American Naturalist 102(925): 243-282.

81. Sankari, SL. Gayathri K, Balachander N, Malathi L. (2015). "*Candida* in potentially malignant oral disorders." J Pharm Bioallied Sci 7(Suppl 1):S162-4.

82. Sarkar, R. and G. P. Rathod (2014). "Clinicopathologic assessment of *Candida* colonisation of oral leukoplakia." Indian J dermatol venereol Leprol 80(5): 413-418.

83. Schepman, K. P., E. H. van der Meij, L. E. Smeele and I. van der Waal (1998). "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(4): 270-275.

84. Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn and C. F. Weber (2009). "Introducing mothur:

open-source, platform-independent, community-supported software for describing and comparing microbial communities." Appl Environ Microbiol 75(23): 7537-7541.

85. Schmidt, B. L., J. Kuczynski, A. Bhattacharya, B. Huey, P. M. Corby, E. L. Queiroz, K. Nightingale, A. R. Kerr, M. D. DeLacure, R. Veeramachaneni, A. B. Olshen and D. G. Albertson (2014). "Changes in abundance of oral microbiota associated with oral cancer." PLoS One 9(6): e98741.

86. Segata, N., J. Izard, L. Waldron, D. Gevers, L. Miropolsky, W. S. Garrett and C. Huttenhower (2011). "Metagenomic biomarker discovery and explanation." Genome Biol 12(6): R60.

87. Seitz, Helmut K, and Peter Becker. "Alcohol metabolism and cancer risk." Alcohol research & health: the journal of the National Institute on Alcohol Abuse and Alcoholism vol. 30,1 (2007): 38-41, 44-7.

88. Shanbhag, V. K. L. (2017). "New definition proposed for oral leukoplakia." Dent Res J (Isfahan) 14(4): 297-298.

89. Shulman, S. T., H. C. Friedmann and R. H. Sims (2007). "Theodor Escherich: the first pediatric infectious diseases physician?" Clin Infect Dis 45(8): 1025-1029.

90. Silverman, S., Jr., M. Gorsky and F. Lozada (1984). "Oral leukoplakia and malignant transformation. A follow-up study of 257 patients." Cancer 53(3): 563-568.

91. Sinha, P., T. Hackman, B. Nussenbaum, N. Wu, J. S. Lewis, Jr. and B. H. Haughey (2014). "Transoral laser microsurgery for oral squamous cell carcinoma: oncologic outcomes and prognostic factors." Head Neck 36(3): 340-351.

92. Socransky, S. S. and A. D. Haffajee (1992). "The bacterial etiology of destructive periodontal disease: current concepts." J Periodontol 63(4 Suppl): 322-331.

93. Spampinato, C. and D. Leonardi (2013). "*Candida* infections, causes, targets, and resistance mechanisms: traditional and alternative antifungal agents." Biomed Res Int 2013: 204237.

94. Srinivasprasad, V., J. Dineshshankar, J. Sathiyajeeva, M. Karthikeyan, J. Sunitha and R. Ragunathan (2015). "Liaison between micro-organisms and oral cancer." J Pharm Bioallied Sci 7(Suppl 2): S354-360.

95. Starzynska, A., A. Pawlowska, D. Renkielska, I. Michajlowski, M. Sobjanek, I. Blazewicz and A. Wlodarkiewicz (2015). "Estimation of oral leukoplakia treatment records in the research of the Department of Maxillofacial and Oral Surgery, Medical University of Gdansk." Postepy dermatol alergol 32(2): 114-122.

96. Tandon, P., A. Dadhich, H. Saluja, S. Bawane and S. Sachdeva (2017). "The prevalence of squamous cell carcinoma in different sites of oral cavity at our Rural health care centre in Loni, Maharashtra - a retrospective 10-year study." Contemp Oncol (Pozn) 21(2): 178-183.

97. Tang, J., I. D. Iliev, J. Brown, D. M. Underhill and V. A. Funari (2015). "Mycobiome: Approaches to analysis of intestinal fungi." J Immunol Methods 421: 112-121.

98. Taylor, P. R., C. C. Abnet and S. M. Dawsey (2013). "Squamous dysplasia--the precursor lesion for esophageal squamous cell carcinoma." Cancer epidemiol biomarkers prev 22(4): 540-552.

99. Tobin, A. O. and I. O. Ajayi (2017). "Common oral conditions and correlates: an oral health survey in Kwara State Nigeria." BMC Res Notes 10(1): 568.

100. Vaidya, S., C. Kapoor and N, Ohri (2013). Early invasive squamous cell carcinoma of the tongue. " SRM Journal of research in dental science 4(1): 35-38.

101. Vuckovic N, B.-B. M., Vuckovic D, Picuric I. (2004). "Presence of *Candida albicans* in potentially malignant oral mucosal lesions." Arch Oncol 12(1):51-54.

102. Warnakulasuriya, S. (2009). "Global epidemiology of oral and oropharyngeal cancer." Oral Oncol 45(4-5): 309-316.

103. Warren, R. L., D. J. Freeman, S. Pleasance, P. Watson, R. A. Moore, K. Cochrane, E. Allen-Vercoe and R. A. Holt (2013). "Co-occurrence of anaerobic bacteria in colorectal carcinomas." Microbiome 1(1): 16.

104. Watts, G. S., K. Youens-Clark, M. J. Slepian, D. M. Wolk, M. M. Oshiro, G. S. Metzger, D. Dhingra, L. D. Cranmer and B. L. Hurwitz (2017). "16S rRNA gene sequencing on a benchtop sequencer: accuracy for identification of clinically important bacteria." J Appl Microbiol 123(6): 1584-1596.

105. White, J. R., N. Nagarajan and M. Pop (2009). "Statistical methods for detecting differentially abundant features in clinical metagenomic samples." PLoS Comput Biol 5(4): e1000352.

106. Whitmore, S. E. and R. J. Lamont (2014). "Oral bacteria and cancer." PLoS Pathog 10(3): e1003933.

107. Whitmore, S.E. and Lamont, R.J. (2014). "Oral Bacteria and Cancer." PLoS Pathog 10(3): e1003933

108. Wu, J., B. A. Peters, C. Dominianni, Y. Zhang, Z. Pei, L. Yang, Y. Ma, M. P. Purdue, E. J. Jacobs, S. M. Gapstur, H. Li, A. V. Alekseyenko, R. B. Hayes and J. Ahn (2016). "Cigarette smoking and the oral microbiome in a large study of American adults." ISME J 10(10): 2435-2446.

109. Yamamura, K., Y. Baba, S. Nakagawa, K. Mima, K. Miyake, K. Nakamura, H. Sawayama, K. Kinoshita, T. Ishimoto, M. Iwatsuki, Y. Sakamoto, Y. Yamashita, N. Yoshida, M. Watanabe and H. Baba (2016). "Human Microbiome *Fusobacterium nucleatum* in esophageal cancer tissue is associated with Prognosis." Clin Cancer Res 22(22): 5574-5581.

110. Yang, C. Y., Y. M. Yeh, H. Y. Yu, C. Y. Chin, C. W. Hsu, H. Liu, P. J. Huang, S. N. Hu, C. T. Liao, K. P. Chang and Y. L. Chang (2018). "Oral Microbiota community dynamics associated with oral squamous cell carcinoma staging." Front Microbiol 9: 862.

111. Yokoyama, S., K. Takeuchi, Y. Shibata, S. Kageyama, R. Matsumi, T. Takeshita and Y. Yamashita (2018). "Characterisation of oral microbiota and acetaldehyde production." J Oral Microbiol 10(1): 1492316.

112. Zaura, E., B. J. Keijser, S. M. Huse and W. Crielaard (2009). "Defining the healthy "core microbiome" of oral microbial communities." BMC Microbiol 9: 259.

113. Zini, A., R. Czerninski and H. D. Sgan-Cohen (2010). "Oral cancer over four decades: epidemiology, trends, histology, and survival by anatomical sites." J Oral Pathol Med 39(4): 299-305.

# Appendices

I. Characteristics of the patient population.

II. Poster presented at British Society for Oral Medicine Annual Scientific Meeting, Dublin, April 2017.

III. Published Abstract, European Oral Microbiology Workshop, Stockholm, Sweden, July 2017.

IV. Publication: **Amer Abdrazak**, Galvin Sheila, Healy Claire M., Moran Gary P., The Microbiome of Potentially Malignant Oral Leukoplakia Exhibits Enrichment for *Fusobacterium*, *Leptotrichia*, *Campylobacter*, and *Rothia* Species (2017). *Frontiers in Microbiology* 8, p2391

## **Appendix I.** Characteristics of the patient population

| Code | Age | OLK[a] | Normal[a] | Biopsy[b] | Smoker | Alcohol | Mouth Wash | Denture |
|---|---|---|---|---|---|---|---|---|
| **P1** | 65 | Rt LBT | Lt LBT | Mild dysplasia | NO | 1-2/wk | - | - |
| **P2** | 61 | Lt LBT | Rt LBT | Mild dysplasia | YES | 2-3u/mth | mouthwash 2x day | - |
| **P3** | 65 | Lt BM | Rt BM | Mod Dysplasia | YES | 1-2u/mth | - | - |
| **P4** | 85 | Lw LM | Up LM | Mod Dysplasia | NO | NO | - | Y |
| **P5** | 62 | Rt BM | Lt BM | Mild dysplasia | YES | NO | - | Y |
| **P6** | 58 | Rt FOM | Lt FOM | Sev dysplasia | YES | 10 u/wk | - | - |
| **P7** | 63 | Lt LBT | Rt BM | Mod dysplasia | YES | 8 u/wk | - | - |
| **P8(a)** **(b)** | 40 | Rt BM Up LM | NA | Mild dysplasia Mod dysplasia | NO | 1u/wk | - | - |
| **P9** | 55 | Lt BG | Rt BG | Mod dysplasia | YES | 1u/wk | mouthwash 1x day | Y |
| **P10** | 70 | Lt BM | Rt BM | Sev dysplasia | NO | 2-3u/mth | mouthwash 1 x day | Y |
| **P11** | 69 | D Ton | Rt BM | Mod dysplasia | NO | NO | | Y |
| **P12** | 58 | Lt LA | Rt LA | Mod dysplasia | NO | 4-5u/wk | - | - |
| **P13** | 56 | Rt BM | Lt BM | Mod dysplasia | YES | 10u/wk | - | - |
| **P14** | 58 | Lt BM | Rt BM | Mod dysplasia | YES | 32 u/wk | - | - |
| **P15** | 79 | Lt HP | Rt HP | Mild dysplasia | NO | 6u/wk | - | Y |
| **P16** | 70 | Lt LBT | Rt LBT | Sev dysplasia | NO | 24u/wk | - | Y |
| **P17** | 66 | Rt FOM | Lt FOM | Mod dysplasia | YES | 20u/wk | - | - |
| **P18** | 65 | V Ton | Rt BM | Mod dysplasia | NO | NO | - | - |
| **P19 (a)** **(b)** | 58 | SP V Ton | NA | Mod dysplasia Mod dysplasia | NO | 3-4 u/wk | mouthwash 1 x day | - |
| **P20** | 40 | Up Rt BG | Lw Rt BG | Mild dysplasia | YES | 8 u/wk | | - |
| **P21** | 67 | FOM | D Ton | Sev dysplasia | YES | 1u/wk | - | Y |
| **P22** | 57 | Rt SP | Lt SP | Mild dysplasia | YES | NO | - | - |
| **P23** | 66 | V Ton | SP | Mod Dysplasia | YES | YES | - | Y |
| **p24** | 37 | Rt LBT | Lt LBT | Sev dysplasia | NO | NO | - | - |
| **p25** | 83 | Lt HP | Rt HP | Mod dysplasia | NO | NO | - | - |
| **p26** | 49 | Rt BM | Lt BM | Mod dysplasia | YES | NO | - | - |
| **p27 (a)** **(b)** | 70 | V Ton BG | NA | Mild dysplasia Mild dysplasia | NO | 5 /wk | mouthwash 1 x day | Y |
| **p28** | 40 | Lt LA | Rt BM | Mod dysplasia | YES | NO | - | - |
| **p29** | 58 | Rt LBT | Lt LBT | Mod dysplasia | YES | 5u/wk | - | - |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **p30** | 57 | Lt SP | Lt BM | Mild Dysplasia | YES | 20u/wk | mouthwash 1 x day | - |
| **p31** | 62 | Lt BM | Rt BM | Sev dysplasia | NO | 6u/wk | mouthwash 1 x day | - |
| **p32 (a) (b)** | 46 | Lt BM LA | Rt BM | Mod dysplasia Mod dysplasia | YES | 20u/w | mouthwash 1 x day | - |
| **p33** | 76 | Lt LA | Rt LA | Mod dysplasia | NO | NO | - | Y |
| **p34** | 65 | Lt LBT | NA | Mod dysplasia | NO | 10u/wk | mouthwash 1 x day | - |
| **P35 (a) (b)** | 63 | Lt LBT Rt BM | Rt BM | Mild dysplasia Mild dysplasia | NO | 6-10u/wk | - | - |
| **P36** | 44 | Rt BM | Lt BM | Mod dysplasia | NO | 2-4 u/wk | - | - |
| **CM1** | 59 | Control | BM | Control | NO | 6-10 u/wk | - | - |
| **CM2** | 42 | Control | BM | Control | YES | 2-3 u/wk | - | - |
| **CM3** | 67 | Control | BM | Control | NO | 12-14 u/wk | mouthwash 1 x day | - |
| **CM5** | 55 | Control | BM | Control | NO | 5u/wk | - | - |
| **CM6** | 36 | Control | BM | Control | YES | 2-3/wk | - | - |
| **CM7** | 45 | Control | BM | Control | YES | 3-4 u/wk | - | - |
| **CM8** | 48 | Control | BM | Control | NO | 2u/wk | - | - |
| **CM9** | 42 | Control | BM | Control | NO | NO | - | - |
| **CM10** | 41 | Control | BM | Control | NO | 6u/wk | - | - |
| **CM11** | 66 | Control | BM | Control | NO | 7-9u/wk | - | - |
| **CM12** | 39 | Control | BM | Control | NO | NO | - | - |
| **CM13** | 41 | Control | BM | Control | NO | NO | - | - |
| **CM14** | 65 | Control | BM | Control | NO | 2 u/wk | - | - |
| **CM15** | 69 | Control | BM | Control | NO | 1 u/wk | mouthwash 1 x day | - |
| **CM16** | 66 | Control | BM | Control | NO | 10 u/wk | - | - |
| **CM17** | 46 | Control | BM | Control | NO | NO | - | - |
| **CM18** | 60 | Control | BM | Control | YES | 40 u/wk | mouthwash 1 x day | Y |
| **CM19** | 21 | Control | BM | Control | NO | NO | mouthwash 1x day | - |
| **CM20** | 65 | Control | BM | Control | NO | 2-3 u/wk | mouthwash 1x day | Y |
| **CM21** | 68 | Control | BM | Control | NO | 2-3 u/wk | - | |
| **CM22** | 53 | Control | BM | Control | NO | NO | - | - |
| **CM23** | 52 | Control | BM | Control | NO | 1 u/wk | - | - |
| **CM24** | 57 | Control | BM | Control | NO | 11 u/wk | mouthwash 1x day | - |
| **CT25** | 43 | Control | LBT | Control | NO | 2-3 u/wk | - | - |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **CT26** | 66 | Control | LBT | Control | NO | 5 u/wk | - | - |
| **CT28** | 51 | Control | LBT | Control | YES | 2-3 u/wk | - | - |
| **CT29** | 58 | Control | LBT | Control | YES | NO | - | - |
| **CT30** | 45 | Control | LBT | Control | NO | NO | - | - |
| **CT31** | 43 | Control | LBT | Control | NO | NO | - | - |
| **CT32** | 30 | Control | LBT | Control | NO | 10 u/wk | - | - |
| **CT33** | 35 | Control | LBT | Control | NO | 4 u/wk | - | - |
| **CT34** | 35 | Control | LBT | Control | NO | 2 u/wk | - | - |

[a]Rt=right, Lt=Left, LBT=lateral border tongue, V Ton=ventral tongue, D Ton=Dorsum Tongue, BM=Buccal mucosa, HP=Hard palate, SP=Soft palate, LA=Lingual Alveolus, FOM=Floor of Mouth, BG=Buccal gingiva.
[b]Mod=Moderate, Sev=Severe

# Appendix II. Poster presented at BSOM, April 2017

# The microbiome of oral leukoplakia shows enrichment in *Fusobacteria*

**Abdrazak Amer, Sheila Galvin, Claire Healy, Gary Moran**
Dublin Dental University Hospital and School of Dental Science, TCD

## *Introduction*

Bacteria have been associated with several human cancers, such as gastric cancer. World-wide, 350,000-400,000 new cases of Oral Squamous Cell Carcinoma (OSCC) are diagnosed every year. Despite the rich microbial diversity of the human oral cavity (> 600 different bacterial species) there have been few investigations of the possible role of these organisms in oral mucosal diseases

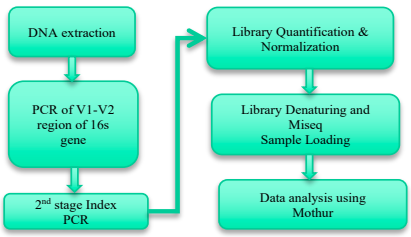
**Aims**

-To identify bacterial species associated with lesions of high malignant potential (leukoplakia)

## *Methods*

110 swabs were collected from patients attending the Oral Medicine clinic at DDUH. Ethical approval was granted by the Joint Hospitals Research Ethics Committee (JREC)

- Swabs were obtained from 47 leukoplakia sites (Diseased mucosa) 31 contralateral healthy sites (normal mucosa)
- Swabs were also recovered from 32 control patients




## *Results*

The pie chart in Figure 1 shows the distribution of sequences classified to the phyla level in all samples, buccal and lingual including controls.

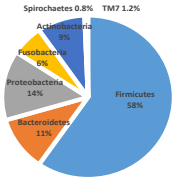



**Figure 1.** Pie chart showing the most abundant phyla (n=7) identified in all Samples (Buccal & Lingual). Six other Phyla were represented in the data but Made up less than 1% of sequences.

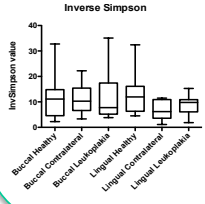



**Figure 2.** Biodiversity was assessed using the Inverse Simpson Index. Lingual samples from patients had Reduced biodiversity relative to healthy control patients

## *LEfSe*

Linear Discriminant Analysis was used to identify significant enrichments in species.

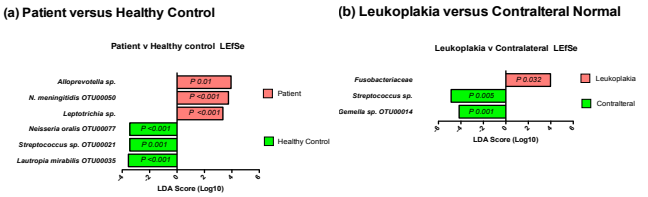



**Figure 3**. (a) Comparison of patients and healthy controls using LEfSe identified enrichments for *Alloprevotella*, *N. meningitidis* and *Leptotrichia sp*. in patients. (b) Comparison of leukoplakia samples and contralateral normal identified enrichments for *Fusobacteria*.

## Abundance of *Fusobacterium spp*. and *Rothia mucilaginosa*

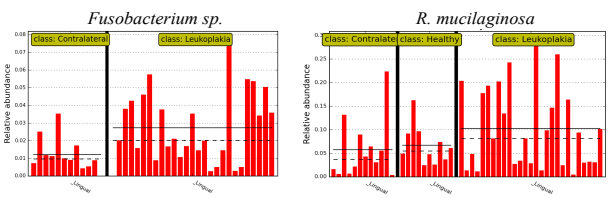



**Figure 4**. Plot showing the abundance *Fusobacterium sp*. and *R. mucilaginosa* in lingual Samples from leukoplakia and contralateral normal sites

## *Leukoplakias exhibit a variety of enrichment patterns*

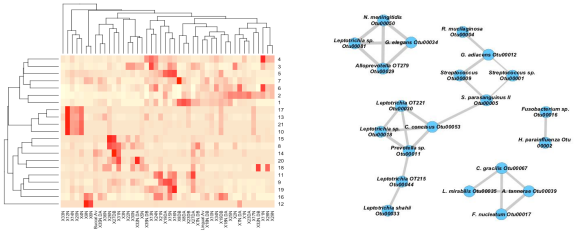


**Figure 5.** Enrichment levels of bacteria identified by LEfSe were compared uisng a Heatmap cluster. Several enrichment patterns could be seen. Analysis of Pearson correlation scores (P 0.01) identified several significant associations. These were combined in a network using Cytoscape

## *Future work*

- Longitudinal studies are required to determine the significance of the enrichment patterns identified in Figure 5 to determine if any of these communities are associated with progression to OSCC.
- The significance of Fusobacteria (both *Fusobacterium spp*. and *Leptotrichia spp.* required investigation to determine whether these organisms can drive malignant transformation of OLK.

## *Acknowledgement*

AA is supported by a Scholarship from the BTRC, Libya. Support was provided by the Dublin Dental University Hospital.

# Appendix III. Published Abstract, European Oral Microbiology Workshop, Stockholm, Sweden, July 2017.

JOURNAL OF ORAL MICROBIOLOGY, 2017
SUPPLEMENT, 1325253
https://doi.org/10.1080/20002297.2017.1325253

Taylor & Francis
Taylor & Francis Group

EOMW Stockholm 2017 – 12th European Oral Microbiology Workshop



## The microbiome of oral leukoplakia shows enrichment in *Fusobacteria* and *Rothia* species

Abdrazak Amer, Sheila Galvin, Claire Healy and Gary P. Moran

Trinity College, University of Dublin, Dublin, Ireland

**ABSTRACT**

The current study was carried out to determine if changes in the oral microbiome were associated with oral leukoplakia. Swabs of oral leukoplakias were taken from 36 patients. Contralateral normal tissue was also swabbed. Swabs from 35 control patients without symptoms of leukoplakia were also taken. DNA was extracted and the V1V2 region of the 16s rRNA gene was sequenced using the Illumina MiSeq and analysed using the Mothur software package.

The structure of oral mucosal communities was most affected by smoking and the location of the site (AMOVA $p < 0.01$). Analysis of the constituents of these communities using LEfSe showed that *Fusobacterium sp.* and *Leptotrichia sp.* were enriched on leukoplakia sites. Patients with leukoplakia also showed enrichment for *Rothia mucilaginosa* and *Campylobacter sp.* Quantitative RT-PCR also showed that leukoplakias from lingual sites were more likely to be colonised by *Candida sp.* Analysis of these enrichments identified specific co-localisation patterns (Pearson correlation $P < 0.01$) including *Leptotrichia sp., Prevotella sp.* and *Campylobacter concisus; F. nucleatum, Alloprevotella tannerae* and *C. gracilis*, amongst others.

*Fusobacteria* have been implicated in the progression of colorectal carcinoma and further studies are now required to determine if these microorganisms are linked to the development of OSCC.

**CONTACT** Abdrazak Amer gpmoran@dental.tcd.ie

# Appendix IV

frontiers in Microbiology

ORIGINAL RESEARCH
published: 01 December 2017
doi: 10.3389/fmicb.2017.02391

# The Microbiome of Potentially Malignant Oral Leukoplakia Exhibits Enrichment for *Fusobacterium, Leptotrichia, Campylobacter,* and *Rothia* Species

*Abdrazak Amer[1], Sheila Galvin[2], Claire M. Healy[2] and Gary P. Moran[1]**

*[1] Division of Oral Biosciences, School of Dental Science, Trinity College Dublin, Dublin Dental University Hospital, Dublin, Ireland, [2] Division of Oral and Maxillofacial Surgery, Oral Medicine and Oral Pathology, School of Dental Science, Trinity College Dublin, Dublin Dental University Hospital, Dublin, Ireland*

Oral leukoplakia presents as a white patch on the oral mucosa and is recognized as having significant malignant potential. Although colonization of these patches with *Candida albicans* is common, little is known about the bacterial microbiota of these patches. In the current study we analyzed the microbiome of oral leukoplakia in 36 patients compared to healthy mucosal tissue from the same patients and healthy control subjects to determine if specific microbial enrichments could be identified early in the malignant process that could play a role in the progression of the disease. This was carried out by sequence analysis of the V1–V2 region of the bacterial 16S rRNA gene using the Illumina MiSeq. Oral leukoplakia exhibited increased abundance of Fusobacteria and reduced levels of Firmicutes (Metastats $P < 0.01$). *Candida* colonization was also more prevalent in leukoplakia patients relative to healthy controls ($P = 0.025$). Bacterial colonization patterns on oral leukoplakia were highly variable and five distinct bacterial clusters were discerned. These clusters exhibited co-occurrence of *Fusobacterium*, *Leptotrichia,* and *Campylobacter* species (Pearson $P < 0.01$), which is strikingly similar to the microbial co-occurrence patterns observed on colorectal cancers (Warren et al., 2013). Increased abundance of the acetaldehydogenic microorganism *Rothia mucilaginosa* was also apparent on oral leukoplakias from lingual sites ($P$ 0.0012). Severe dysplasia was associated with elevated levels of *Leptotrichia* spp. and *Campylobacter concisus* ($P < 0.05$). Oral leukoplakia exhibits an altered microbiota that has similarities to the microbiome of colorectal cancer.

**Keywords: microbiome, oral leukoplakia, oral cancer, *Fusobacteria, Campylobacter, Rothia mucilaginosa***

OPEN ACCESS

***Edited by:***
*John R. Battista, Louisiana State University, United States*

***Reviewed by:***
*Nicola Segata, University of Trento, Italy*
*Donnabella Castillo Lacap-Bugler, Auckland University of Technology, New Zealand*

****Correspondence:***
*Gary P. Moran*
*gpmoran@dental.tcd.ie*

***Specialty section:***
*This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology*

***Received:*** *15 August 2017*
***Accepted:*** *20 November 2017*
***Published:*** *01 December 2017*

***Citation:***
*Amer A, Galvin S, Healy CM and Moran GP (2017) The Microbiome of Potentially Malignant Oral Leukoplakia Exhibits Enrichment for Fusobacterium, Leptotrichia, Campylobacter, and Rothia Species. Front. Microbiol. 8:2391. doi: 10.3389/fmicb.2017.02391*

## INTRODUCTION

Oral squamous cell carcinoma is the most common oral malignancy and is the eight most common cancer worldwide (Scully and Bagan, 2009). As with most cancers, failure to diagnose in the early stages of tumor development can have a dramatic impact upon long-term prognosis, with 5 year survival of late stage OSCC being less than 40% (Yanik et al., 2015). Early detection can be aided

**Abbreviations:** ACH, acetaldehyde; OLK, oral leukoplakia; OSCC, oral squamous cell carcinoma.

by the identification of potentially malignant precursors such as OLK, a condition that manifests as white, raised patches on the mucosal surface. The reported rate of transformation of OLK varies from 1 to >20% depending on the population and the length of follow up (Johnson et al., 2011; Yanik et al., 2015). Although smoking, alcohol and betel nut use are clearly associated with the development of OLK, factors that drive the malignant transformation of these lesions are poorly understood and it is difficult to accurately predict whether an OLK will resolve, persist or progress to OSCC (Liu et al., 2012). Malignant transformation of OLK is greater with increasing age, female gender, non-smoking, extent of OLK and may be affected by location, with some studies reporting that OLK on the lateral and ventral surfaces of the tongue and the floor of mouth having a higher rate of transformation (Silverman et al., 1984; Ho et al., 2012; Liu et al., 2012; Yanik et al., 2015). The degree of dysplasia is at present the most reliable indicator of the likelihood of transformation (Scully and Bagan, 2009; Liu et al., 2012).

A variety of hypotheses have been put forward to link microorganisms and their products with oral cancer (Perera et al., 2016). The production of known carcinogens such as *N*-nitroso compounds (Mills and Alexander, 1976) and acetaldehyde have been proposed as a way that microbes can induce OSCC (Marttila et al., 2013; Moritani et al., 2015). *Candida* colonization is often associated with OLK, which is referred to as "candidal leukoplakia" and is characterized by hyphal infiltration of the tissue (Abdulrahim et al., 2013; Alnuaimi et al., 2015). Some studies have associated *Candida* carriage with the degree of dysplasia (McCullough et al., 2002). Human papilloma virus is strongly associated with oro-pharyngeal cancer, but no clear role for it in the development of OSCC has been identified (Gillison et al., 2000).

More recently, microbiome studies have been carried out to identify changes in the bacterial microbiota in OSCC in the hope of identifying biomarkers of malignant transformation (Mager et al., 2005; Pushalkar et al., 2011; Hu et al., 2016). Nagy et al. (1998) carried out direct culture of OSCC lesions from 21 patients and identified high levels of *Porphyromonas, Prevotella,* and *Fusobacterium* species. Schmidt et al. (2014) analyzed the microbiome of OSCC in 27 patients by sequence analysis of the V4 region of bacterial DNA. OSCC patients exhibited reduced *Streptococcus* sp. and *Rothia* sp. but elevated levels of Bacteroidetes and Fusobacteria. Very recently, an enrichment for *F. nucleatum* and *P. aeruginosa* was identified in OSCC lesions from Yemeni patients (Al-hebshi et al., 2017). Although *F. nucleatum* is common in dental plaque and associated with oral infections, recent studies have also identified enrichment for *F. nucleatum* in colorectal cancer tissue (Castellarin et al., 2011; Kostic et al., 2011). Further studies have shown co-occurrence of *F. nucleatum*, *Leptotrichia* spp., and *Campylobacter* spp. on these tissues (Warren et al., 2013). *Fusobacterium* spp. have also been associated with cancers of the esophagus and were shown by PCR to be significantly enriched in esophageal cancer samples compared to normal esophageal mucosa (Yamamura et al., 2016). In the murine colon, it has been shown that *F. nucleatum* adheres to E-cadherin and Gal-GalNac expressed on tumors and this may also be the case in the oral cavity (Rubinstein et al., 2013; Abed et al., 2016). Molecular studies have shown that *F. nucleatum* can potentially promote tumor growth through activation of the Il-6-STAT3 axis and via activation of β-catenin signaling via the FadA adhesin (Rubinstein et al., 2013; Binder Gallimidi et al., 2015).

Although microbes such as *F. nucleatum* could potentially accelerate tumor development, most studies of the tumor microbiome have examined these lesions late in the malignant process. The current study was designed to examine the microbiome of potentially malignant OLK to determine if specific microbial enrichments could be identified early in the malignant process that could play a role in progression. Our study identifies a specific enrichment in Fusobacteria (both *Fusobacterium* spp. and *Leptotrichia* spp.) and *Campylobacter* spp. that bears similarity to recently identified enrichments on colorectal carcinoma.

# MATERIALS AND METHODS

## Sample Collection

Ethical Approval for this study was granted by the Joint Hospitals' Research Ethics Committee (Tallaght Hospital, Dublin). Following written informed patient consent, mucosal swabs were collected at the Dublin Dental University Hospital (DDUH) using Catch-all collection swabs (Epicentre, Madison, WI, United States). Patients presenting with OLK ($n = 36$, average age: 60.6) were swabbed at the site of the OLK and a contralateral normal site, where present (Supplementary Table S1). No normal sites were present in four patients and five patients presented with more than one OLK and were swabbed at both sites (Supplementary Table S1). Data on the degree of dysplasia identified on biopsy, the presence of dentures, smoking, alcohol consumption and oral hygiene measures were recorded. Patients having taken antibiotics or used topical steroids intra-orally in the past 6 months, were excluded along with patients with diabetes mellitus, Crohn's disease, ulcerative colitis, current viral infection (cold/flu), or history of gastrointestinal malignancy. Healthy controls ($n = 32$, average age: 50.3) were subject to the same exclusion criteria and included 23 buccal swabs and 9 lingual (lateral border tongue) swabs.

## DNA Extraction

Swabs were resuspended in 300 μl of TSE buffer (10 mM Tris-HCl [pH 7.8], 1 mM EDTA, 100 mM NaCl) containing 500 U Ready-lyse lysozyme (Epicentre, Madison, WI, United States) and incubated at room temperature for 15 min. DNA extractions were carried out using the MasterPure DNA Purification Kit (Epicentre, Madison, WI, United States) using the manufacturer's protocol with the addition of a bead disruption step, as follows: after Proteinase K treatment and before the addition of RNase, 0.25 glass beads (100 μM diameter) were added to the tube and the sample was disrupted in a Minibead beater (Biospec Products, Bartlesville, OK, United States) for 30 s. DNA pellets were resuspended in 35 μl TE buffer.

## DNA Sequencing

Amplification of the V1–V2 region of the 16S rRNA gene was carried using with the KAPA HiFi Hot start system (Kapa Biosystems) with the primers 27F-YM and 338R-R (27F-YM: 5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTCAG TCTGTCAGAGTTTGATYMTGGCTCAG; 338R-R: 5′ GTC TCGTGGGCTCGGAGATGTGTATAAGAGACAGTATGGTAA TTCATGCTGCCTCCCGTAGRAGT) (Frank et al., 2008; Marttila et al., 2013). The V1–V2 region was used as it has been shown that ∼90% of species in the Human Oral Microbiome Database (HOMD) can be correctly identified using this region (Diaz et al., 2014). Sample indexing was carried out with the Nextera XT Index Kit (Illumina) and library quantification and purification was carried out according to the Illumina protocol "16S Metagenomic Sequence Library Preparation" (Storey and Tibshirani, 2003; Abdulrahim et al., 2013; Illumina, 2013; Alnuaimi et al., 2015; Yanik et al., 2015). Size and integrity of indexed amplimers was determined using a Bioanalyzer (Agilent Technologies) and samples were normalized to 4 nM. Samples were combined to generate a pooled library, denatured and combined with PhiX control DNA (5%) and loaded at a concentration of 6 pM. 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), BioProject accession PRJNA394711.

## Sequence Analysis

Bacterial 16S rRNA sequences were analyzed using the Mothur pipeline (Schloss et al., 2009). Forward and reverse reads were aligned and filtered for quality and length (300–400 bp). Chimeric sequences were identified using Uchime (Edgar et al., 2011) and removed along with contaminating eukaryotic sequences and rare sequences (1–2 copies) prior to taxonomic classification. Sequences were classified in Mothur using the HOMD reference 16S rRNA gene set (V14.51). Operating taxonomic units (OTUs) were defined at a cutoff of 2% (i.e., 98% sequence identity), which our empirical analysis found was suitable for discrimination of the major oral taxa. OTU classifications from Mothur were supplemented by BLAST searches using consensus sequences to identify to the species level, where necessary. Mothur was used to calculate the Inverse Simpson index for each sample and to generate intra-sample rarefaction curves. For beta diversity analysis in Mothur, data were subsampled (normalized) to the smallest data set (3,709 sequences). Differences in community structure were inferred using distance matrices generated using the Bray–Curtis dissimilarity index calculated in Mothur (Schloss, 2008). Distance matrices were visualized using non-metric multidimensional scaling (NMDS) using the rgl package in R Studio.

## Statistical Methods

Statistical differences between microbial communities was identified using analysis of molecular variance (AMOVA). Taxonomic classifications that showed statistically significant differences in abundance in the study groups were identified using Metastats and LEfSe (White et al., 2009; Segata et al., 2011). Data on OTU abundance was formatted and normalized for LEfSe in Mothur and analyzed using the Galaxy server at https://huttenhower.sph.harvard.edu/galaxy/. *P*-values were corrected for multiple hypothesis testing using Bioconductor's *q*-value package to estimate the false discovery rate (FDR) and associated *q*-values (Storey and Tibshirani, 2003). Taxa showing significant differences in abundance were reconfirmed using a Wilcoxon matched pairs test. Heatmaps were generated in R studio using Vegan to carry out hierarchical clustering on a matrix of Bray–Curtis dissimilarity values. Further statistical analysis was carried out using Prism (Graphpad Software, La Jolla, CA, United States).

## Quantitative PCR

To estimate carriage levels of *Candida* spp., we carried out quantitative Real-time PCR using primers targeting the *Candida* ITS2 rDNA region described by Kraneveld et al. (2012) (ITSF: 5′ CCTGTTTGAGCGTCRTTT; ITSR: 5′ TTCTCCGCTTATTGATAT). Amplification of the *Candida* ITS rDNA was performed with Fast Sybr Green master mix (Applied Biosystems) using the ABI 7500 Real-Time PCR System. *Candida* levels in each sample (CFU/ml) were extrapolated from a standard curve generated using DNA extracted from a serially diluted culture of *Candida albicans* SC5314 ($10^8$ to 10 CFU/ml). CT values generated from the diluted *C. albicans* DNA samples were used to generate a standard curve using the Prism Software package and the number of *Candida* CFU/ml in all patient and control samples were determined by extrapolation from this standard curve. Data from the samples that were deemed *Candida* spp. positive contained DNA equivalent to at least 300 CFU/ml.

# RESULTS

## Microbiome Composition

Following sequence assembly and processing in Mothur, over 14 million sequences were included for analysis. These sequences were classified to 13 bacterial phyla representing 215 genera and 2,030 OTUs (clustered at 2% identity, Supplementary Table S2). Number of sequences in each sample are listed in Supplementary Table S3. Approximately 99% of sequences could be classified to the six most abundant phyla, Firmicutes, Bacteroidetes, Proteobacteria, Fusobacteria, Actinobacteria and Spirochaetes (**Figure 1**), with the remaining ∼1% belonging to the phyla *TM7, SR1,* Tenericutes, Chloroflexi and Synergistetes. Analysis of the abundance of these phyla using Metastats (White et al., 2009) revealed that leukoplakia samples had significantly lower levels of Firmicutes (*P* 0.0009) and higher levels of Proteobacteria (*P* 0.046) and Fusobacteria (*P* 0.0029) relative to contralateral normal sites. Levels of Firmicutes were also low relative to healthy control patients (*P* 0.008). Leukoplakia samples also exhibited higher levels of Bacteroidetes relative to healthy controls (*P* 0.0049), although these levels were not significantly higher than contralateral normal sites.

Levels of *Candida* colonization were assessed by quantitative PCR (**Figure 1B**). In total, 35% of samples from OLK sites showed significant colonization with *Candida* spp. (equivalent to at least

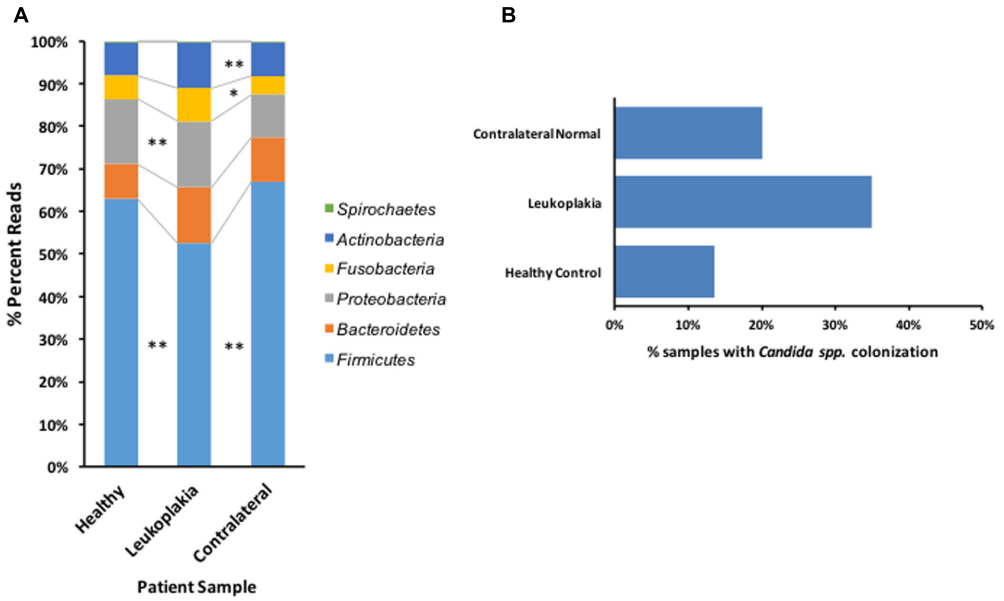



**FIGURE 1 |** Overview of bacterial phyla distribution and *Candida* carriage patterns in patients with OLK and healthy controls. **(A)** Graph showing the abundance of the six most important phyla in healthy control subjects and sites of OLK and contralateral normal sites in patients. * Indicates Metastats $P < 0.05$ and ** $P < 0.01$. **(B)** Percentage of *Candida* positive samples recovered from patients (OLK and contralateral sites) and healthy controls, determined by qPCR.

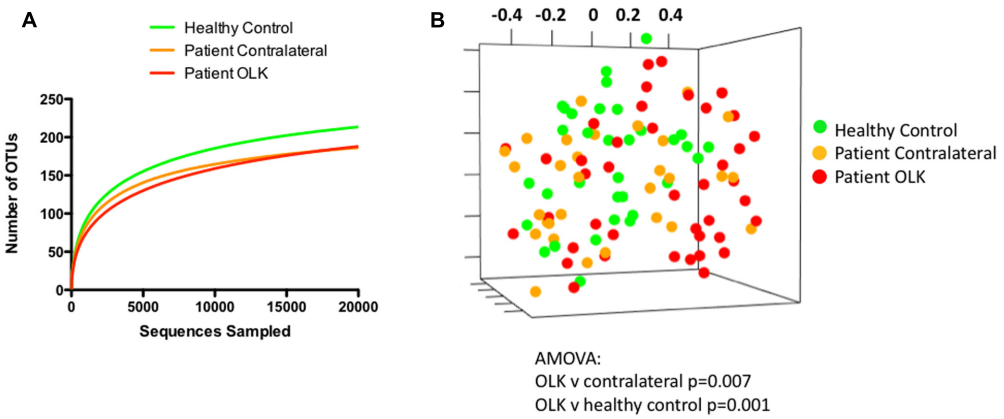



**FIGURE 2 |** Analysis community structure in samples from OLK and healthy mucosa. **(A)** Rarefaction curve showing analysis of OTUs identified in samples from healthy controls (green) and patients (red = OLK tissue, orange = contralateral normal tissue). **(B)** Non-metric multidimensional scaling (NMDS) plots generated from a matrix of Bray–Curtis dissimilarity values using Mothur (Schloss et al., 2009). Separation of OLK communities (red) from contralateral (orange; $P = 0.007$) and healthy control communities (green; $P = 0.001$) was determined using analysis of molecular variance (AMOVA).

300 CFUs/ml). Colonization levels were lower at contralateral healthy mucosa from the same group of patients (20%) and in healthy control patients (13.5%). The majority (84%) of *Candida* spp. positive OLK were located at lingual sites (tongue, palate).

The species richness of the mucosal communities from healthy and OLK tissue were compared using species-accumulation (rarefaction) curves (**Figure 2A**). Analysis of the curves show that the number of OTUs begins to plateau above 5,000 sequences, indicating that the sampling effort is sufficient (only one sample of 104 yielded < 5,000 sequence reads, Supplementary Table S3). The curves also indicated that healthy control subjects exhibited greater species richness compared to samples recovered from OLK patients. Biodiversity of the communities was estimated using the Inverse Simpson index (Supplementary Table S3). Mean inverse Simpson values from patients and healthy controls were not significantly different.

To compare the structure of the bacterial communities from each sample, we determined the Bray–Curtis dissimilarity values for each patient sample using normalized data and generated a distance matrix from these data. Visualization of these data was carried out using NMDS (**Figure 2B**). Statistical differences in community structure were assessed using AMOVA. Mucosal

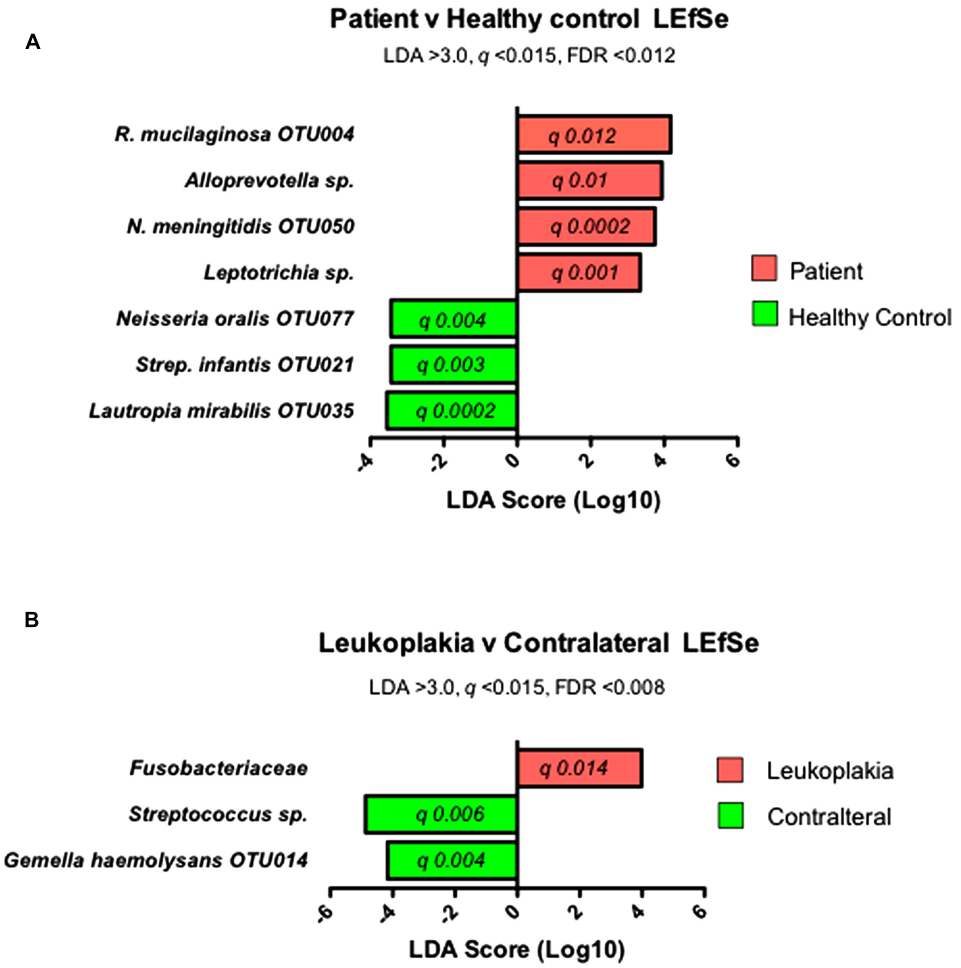



**FIGURE 3 |** Results of LEfSe analysis to identify significantly enriched bacterial taxa in **(A)** communities from patients and healthy controls and **(B)** in communities from sites of OLK and contralateral tissue from patients.

health status appeared to influence population structure as bacterial communities from OLK samples exhibited significant separation from contralateral and healthy control mucosa in AMOVA tests ($P < 0.007$). However, smoking and the site of sampling (i.e., whether a buccal or lingual site) were both shown to have a greater influence on community structure ($P < 0.001$, Supplementary Figure S1).

## Characterization of Specific OTU Enrichments

The data set of all OTU abundances (Supplementary Table S4) was analyzed using LEfSe to identify significant enrichments in specific species or OTUs. Confirmation of these results was carried out using paired analysis of OLK and contralateral tissue for specific OTUs or species.

Firstly, communities from OLK patients (combining both OLK and contralateral samples) were compared with communities from healthy controls. This analysis identified increased abundance of several taxa in OLK patients ($q < 0.015$, LDA > 3.0) including *Rothia mucilaginosa* (OTU004), *Alloprevotella* spp., *Neisseria meningitidis* (OTU050), and *Leptotrichia* spp. (**Figure 3A**). Next, we compared populations from OLK samples and contralateral normal sites from the same group of patients. This analysis identified increased abundance of Fusobacteriaceae on OLK (**Figure 3B**). Contralateral normal sites also exhibited increased abundance of *Streptococcus* spp. and *Gemella haemolysans* (OTU014).

In order to confirm these enrichments in OLK communities, we carried out a paired analysis on matched leukoplakia and contralateral samples (**Figures 4A–F**). The Wilcoxon matched-pairs test confirmed that family Fusobacteriaceae (including *Fusobacterium* spp. and *Leptotrichia* spp.) were significantly greater in OLK communities relative to matched contralateral samples ($P$ 0.024). This was significant for the largest fusobacterial taxon identified, *F. nucleatum* OTU016 (**Figure 4A**; $P$ 0.039). Changes in abundance of *F. nucleatum* subsp. *vincentii* was site dependent with increased abundance in buccal OLKs, but conversely an increased abundance in lingual contralateral sites (**Figure 4B**). Increased abundance of *Leptotrichia* spp. ($P$ 0.039), *Campylobacter* spp. ($P$ 0.0069), and

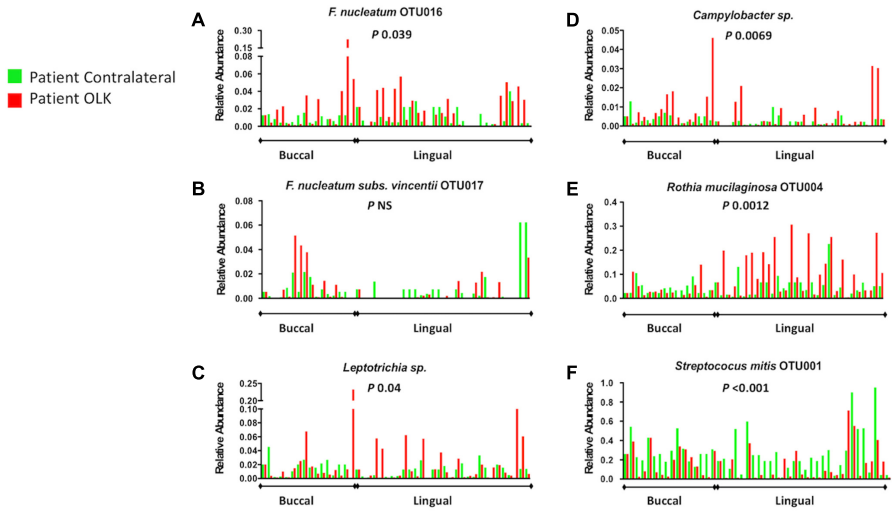



**FIGURE 4 |** Relative abundance of selected taxonomic groups identified by LEfSe analysis. The abundance of each organism **(A–F)** in OLK and contralateral healthy communities from each patient is plotted side by side. No healthy tissues were available from four patients (Supplementary Table S1) and these are plotted versus the average values for control tissue. *P*-values refer to Wilcoxon matched pairs test results.

*R. mucilaginosa* (*P* 0.0012) was also shown in OLK communities relative to contralateral samples (**Figures 4C–E**). Conversely, *Streptococcus mitis* was significantly enriched in contralateral healthy samples (**Figure 4F**, $P < 0.001$).

Patient meta-data was also used to interrogate the abundance data using LEfSe (Supplementary Figure S2). Levels *F. nucleatum* OTU0016 and *Leptotrichia* sp. OTU044 were reduced in smokers but elevated in patients that were *Candida* carriers ($P < 0.01$). Consumption of alcohol (>1 unit/week) was associated with increased abundance of *Campylobacter* spp. (*P* 0.018). Use of mouthwash, age and sex of the patients did not significantly affect the level of the taxa under investigation.

In order to determine whether any of these enriched bacterial taxa occurred together in specific communities on OLK, we generated a heatmap comparing the incidence of the 20 most abundant patient-enriched OTUs (Figure5A). Clustering of these profiles using Bray–Curtis dissimilarity values generated a dendogram with five major groups (labeled 1–5, **Figure 5A**) representing different enrichment patterns. Separation of these groups was statistically significant (UNIFRAC $P > 0.001$). Next, Pearson correlation coefficients were determined for each pair of OTUs and highly significant ($P < 0.01$) correlations were visualized using Cytoscape (**Figure 5B**). Five clusters representing the groups identified in **Figure 5A** were identified. Clusters (1) and (2) were found predominantly on buccal OLKs whereas Clusters (3), (4), and (5) were largely lingual. *S. mitis* OTU001 was negatively correlated with *S. parasanguinus* OTU005 (−0.444) and *G. adiacens* OTU012 (−0.375).

Patients in clusters (1), (3), and (5) were all *Candida* negative whereas 70% of patients in clusters 2 and 4 were *Candida* positive by qPCR. Severe dysplasia ($n = 6$) was recorded in biopsies from clusters (2), (3), and (4), with three cases (50%) of severe dysplasia in cluster (3). This cluster (3) exhibited enrichments for *Leptotrichia* spp. and *Campylobacter concisus* OTU0053. Comparison of the levels of these two organisms in all patients with mild, moderate, or severe dysplasia revealed that severe dysplasia was associated with elevated levels of *Leptotrichia* spp. and *C. concisus* OTU0053 (Supplementary Figure S3; Kruskal–Wallis $P < 0.044$).

## DISCUSSION

Cancer is generally a multifactorial disease involving the accumulation of multiple genetic lesions. The involvement of the oral microbiome in the etiology or progression of OLK to OSCC has received relatively little attention. Most microbiome studies of OSCC have investigated the microbiome late in the malignant process and it is unclear whether these organisms adhere to the malignant tissue post-transformation or are drivers of the transformation itself (Nagy et al., 1998; Mager et al., 2005; Pushalkar et al., 2011; Schmidt et al., 2014; Hu et al., 2016). While some OSCCs arise *de novo*, others arise out of pre-existing lesions with malignant potential, the most common of which are OLKs. This study was designed to investigate differences in the bacterial community structure of OLK by comparing the microbiome of OLK samples with healthy mucosa from the same patient and control samples from healthy individuals. This initial investigation was designed to determine whether the mucosal microbiome was altered on OLK and whether this could be a potential driver of malignant transformation. Phyla level analysis of our data showed that significant changes in the abundance of three of the six most abundant oral phyla occurred in OLK compared to healthy mucosa from the same patients. We observed a decrease in the abundance of Firmicutes and an increase in the abundance of Fusobacteria and Actinobacteria. A decrease in Firmicutes and increase in Fusobacteria have been observed previously in OSCC (Schmidt et al., 2014). Our data

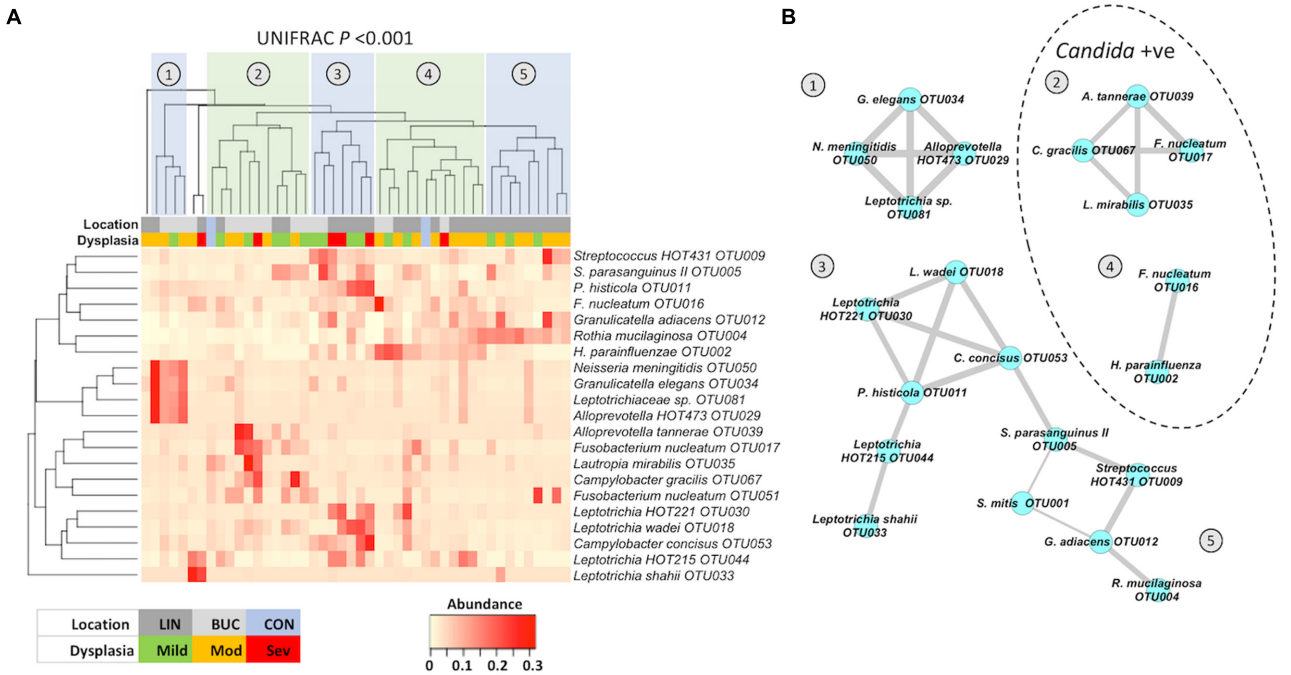



**FIGURE 5 | (A)** Heatmap showing the abundance of taxa identified by LEfSe analysis. Heatmap and dendogram were generated in Vegan using a matrix of Bray–Curtis dissimilarity values. Separation of patient clusters marked 1 to 5 in the dendogram was highly significant (Unweighted UNIFRAC $P < 0.001$). The color-coded legend indicates whether each sample is from a buccal or lingual site and the degree of dysplasia following biopsy (mild, moderate, or severe). CON (light blue) correspond to the average values in healthy buccal (left) and healthy lingual (right) samples. **(B)** Co-occurrence map generated from Pearson correlation coefficients ($r$) generated in Prism ($P < 0.01$) and graphically presented using Cytoscape 3.2.1. Thickness of edges (connecting lines) are proportional to $r$-values (0.4–0.99). Clusters within the dotted line were identified in patients who were also colonized with *Candida* spp.

indicate that these shifts occur early in the process of malignant transformation and may potentially play a role in its progression. As with previous studies, colonization with *Candida* spp. was also more prevalent on OLK compared to non-OLK sites.

Mucosal communities from OLK patients exhibited reduced species richness compared to healthy control subjects, but overall levels of biodiversity remained similar. Comparison of communities from healthy and diseased mucosa using the Bray–Curtis metric showed that community structure was significantly affected by the presence of leukoplakia (**Figure 2B**). However, our analysis indicated that the site (either buccal or lingual) and smoking had more significant impacts on community structure than whether the sample was recovered from OLK. In the current study smokers had significantly reduced levels of *Neisseria* sp. as recently reported (Wu et al., 2016). Smoking was also associated with reduced levels of *F. nucleatum* OTU016 and *Leptotrichia* sp. OTU044. The effects of smoking on the microbiome in OLK patients warrants further investigation as non-smokers are more likely to undergo malignant transformation than smokers (Silverman et al., 1984; Ho et al., 2012). Alcohol was also shown to affect the microbiome with alcohol consumption associated with enrichment for *Campylobacter* spp.

In general, OLK communities were found to be significantly enriched for Fusobacteriaceae, whereas contralateral sites exhibited higher levels of *Streptococcus* spp. and *Gemella* spp. Visualization of these data using a heatmap highlighted the heterogeneity in the patterns of enrichment. Five major clusters of OTUs could be identified (**Figure 5**), with clusters (1) and (2) generally associated with buccal OLK while clusters (3), (4), and (5) were largely lingual. Cluster (5) was exclusively lingual and exhibited enrichment for *R. mucilaginosa*, a normally abundant species in lingual communities. *R. mucilaginosa* typically accounted for ∼20% of sequences in enriched communities (compared to 6.7% of sequences on average in healthy patients). *Rothia* sp. are Gram positive species and are considered a part of the normal flora in the oropharynx and upper respiratory system (Kobayashi et al., 2012). One study examining the ability of oral microbes to produce acetaldehyde has shown that *R. mucilaginosa* is a potent producer of acetaldehyde (Moritani et al., 2015). ACH is a known carcinogen that can cause mutations in DNA and can cause sister chromatid exchanges and chromosomal defects in human cells. ACH has been found in the mouth after ethanol consumption (Homann et al., 1997). The high level of *R. mucilaginosa* identified in our study suggests that this organism could contribute to salivary ACH levels and studies are now ongoing to determine the ACH producing capacity of this species.

The remaining four clusters identified in **Figure 5B** all contained members of the Fusobacteriaceae, namely *F. nucleatum* and *Leptotrichia* spp. Cluster (2) included *F. nucleatum* subsp. *vincentii* (OUT017) and *Campylobacter gracilis*. Co-occurrence of Fusobacteria and *Campylobacter* sp. was also identified in

the predominantly lingual cluster (3), which was enriched for *C. concisus* and *Leptotrichia* spp. The co-occurrence of *Fusobacteriaceae* with *Campylobacter* spp. observed in clusters (2) and (3) are strikingly similar to the co-occurrence patterns observed by Warren et al. (2013) on colorectal carcinomas. They identified a similar polymicrobial signature in an analysis of 130 colorectal carcinoma samples. Co-occurrence of *Fusobacterium*, *Leptotrichia,* and *Campylobacter* species were overrepresented on colorectal tumors and this enrichment was associated with specific changes in host gene expression, including increased pro-inflammatory IL-8 expression. As the natural niche for many of these organisms is the oral cavity, it is perhaps not surprising that we should also identify similar co-occurrence patterns on OLK, which suggests a strong predilection for these organisms to adhere to dysplastic tissue throughout the GI tract. Although only six of the leukoplakias analyzed here exhibited severe dysplasia, 50% of these were found in cluster (3) and analysis of all samples showed that severe dysplasia was significantly associated with elevated *Leptotrichia* spp. and *C. concisus* OTU053. Further studies are required to determine if Fusobacteria and *Campylobacter* spp. can exert synergy in accelerating tumor development, both in the colon and the oral cavity.

Examination of these data show that the species most enriched in OLK include *Fusobacterium*, *Leptotrichia*, *Campylobacter,* and *Rothia* species. *Candida* carriage was also prevalent on lingual OLKs and may influence the microbiota. Fusobacteria were the most consistent enrichment in all OLKs (both buccal and lingual). The most significant question posed by these data is whether these organisms influence the malignant transformation of OLK. Co-occurrence of Fusobacteria and *Campylobacter* spp. on OLK suggest that similar processes in malignant transformation may be at work in the oral cavity and in the colon, namely Fusobacterial activation of the Il-6-STAT3 axis and activation of β-catenin signaling. Work is ongoing to determine the oncogenic effects of these communities on oral cells and larger studies are underway to determine whether the presence of these communities can influence malignant transformation. If any of these communities can be implicated in malignant transformation, it may in the future be possible to identify those patients most at risk of developing OSCC and to perhaps use topical antibiotic therapy to prevent malignant transformation.

## AVAILABILITY OF DATA AND MATERIAL

All sequence data has been submitted to the NCBI sequence read archive (SRA), BioProject accession PRJNA394711.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Trinity College Dublin "Good Research Practice" guidelines. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the 'Joint Hospitals' Research Ethics Committee, Dublin.'

## AUTHOR CONTRIBUTIONS

AA performed DNA extraction, PCR, DNA sequencing, data analysis, and manuscript preparation. SG was involved in data collection, study design, and manuscript preparation. CH was involved in study design, data collection, and manuscript preparation. GM was involved in study design, data analysis, and manuscript preparation.

## ACKNOWLEDGMENTS

The authors thank William Wade for helpful discussions and help with primer design. They wish to thank Elaine Kenny and Kathleen McGrath at TrinSeq, St. James's Hospital Dublin for their assistance with sample preparation for the Illumina MiSeq. They thank the Board of the Dublin Dental University Hospital for support. AA was supported by a scholarship from the Libyan Government.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02391/full#supplementary-material

## REFERENCES

Abdulrahim, M. H., McManus, B. A., Flint, S. R., and Coleman, D. C. (2013). 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. doi: 10.1371/journal.pone.0073738

Abed, J., Emgård, J. E. M., Zamir, G., Faroja, M., Almogy, G., Grenov, A., et al. (2016). Fap2 mediates *Fusobacterium nucleatum* colorectal adenocarcinoma enrichment by binding to tumor- expressed Gal-GalNAc. *Cell Host Microbe* 20, 215–225. doi: 10.1016/j.chom.2016.07.006

Al-hebshi, N. N., Nasher, A. T., Maryoud, M. Y., Homeida, H. E., Chen, T., Idris, A. M., et al. (2017). Inflammatory bacteriome featuring *Fusobacterium nucleatum* and *Pseudomonas aeruginosa* identified in association with oral squamous cell carcinoma. *Sci. Rep.* 7:1834. doi: 10.1038/s41598-017-02079-3

Alnuaimi, A. D., Wiesenfeld, D., O'Brien-Simpson, N. M., Reynolds, E. C., and McCullough, M. J. (2015). 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. doi: 10.1016/j.oraloncology.2014.11.008

Binder Gallimidi, A., Fischman, S., Revach, B., Bulvik, R., Maliutina, A., Rubinstein, A. M., et al. (2015). Periodontal pathogens *Porphyromonas gingivalis* and *Fusobacterium nucleatum* promote tumor progression in an oral-specific chemical carcinogenesis model. *Oncotarget* 6, 22613–22623. doi: 10.18632/oncotarget.4209

Castellarin, M., Warren, R. L., Freeman, J. D., Dreolini, L., Krzywinski, M., Strauss, J., et al. (2011). *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res.* 22, 299–306. doi: 10.1101/gr.126516.111

Diaz, P. I., Strausbaugh, L. D., and Dongari-Bagtzoglou, A. (2014). Fungal-bacterial interactions and their relevance to oral health: linking the clinic

and the bench. *Front. Cell Infect. Microbiol.* 4:101. doi: 10.3389/fcimb.2014.00101

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A., and Olsen, G. J. (2008). Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl. Environ. Microbiol.* 74, 2461–2470. doi: 10.1128/AEM.02272-07

Gillison, M. L., Koch, W. M., Capone, R. B., Spafford, M., Westra, W. H., Wu, L., et al. (2000). Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J. Natl. Cancer Inst.* 92, 709–720. doi: 10.1093/jnci/92.9.709

Ho, M. W., Risk, J. M., Woolgar, J. A., Field, E. A., Field, J. K., Steele, J. C., et al. (2012). The clinical determinants of malignant transformation in oral epithelial dysplasia. *Oral Oncol.* 48, 969–976. doi: 10.1016/j.oraloncology.2012.04.002

Homann, N., Jousimies-Somer, H., Jokelainen, K., Heine, R., and Salaspuro, M. (1997). High acetaldehyde levels in saliva after ethanol consumption: methodological aspects and pathogenetic implications. *Carcinogenesis* 18, 1739–1743. doi: 10.1093/carcin/18.9.1739

Hu, X., Zhang, Q., Hua, H., and Chen, F. (2016). Changes in the salivary microbiota of oral leukoplakia and oral cancer. *Oral Oncol.* 56, e6–e8. doi: 10.1016/j.oraloncology.2016.03.007

Illumina (2013). *16S Metagenomic Sequencing Library Preparation*, 1–28. Available at: http://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf [accessed January 8, 2017].

Johnson, N. W., Jayasekara, P., and Amarasinghe, A. A. H. K. (2011). Squamous cell carcinoma and precursor lesions of the oral cavity: epidemiology and aetiology. *Periodontol 2000* 57, 19–37. doi: 10.1111/j.1600-0757.2011.00401.x

Kobayashi, T., Uchibori, S., Tsuzukibashi, O., Goto, H., and Aida, M. (2012). A selective medium for *Rothia mucilaginosa* and its distribution in oral cavities. *J. Microbiol. Methods* 91, 364–365. doi: 10.1016/j.mimet.2012.09.011

Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., Earl, A. M., et al. (2011). Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res.* 22, 292–298. doi: 10.1101/gr.126573.111

Kraneveld, E. A., Buijs, M. J., Bonder, M. J., Visser, M., Keijser, B. J. F., Crielaard, W., et al. (2012). The relation between oral *Candida* load and bacterial microbiome profiles in Dutch older adults. *PLOS ONE* 7:e42770. doi: 10.1371/journal.pone.0042770.s008

Liu, W., Shi, L.-J., Wu, L., Feng, J.-Q., Yang, X., Li, J., et al. (2012). Oral cancer development in patients with leukoplakia–clinicopathological factors affecting outcome. *PLOS ONE* 7:e34773. doi: 10.1371/journal.pone.0034773

Mager, D. L., Haffajee, A. D., Devlin, P. M., Norris, C. M., Posner, M. R., and Goodson, J. M. (2005). 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. doi: 10.1186/1479-5876-3-27

Marttila, E., Uittamo, J., Rusanen, P., Lindqvist, C., Salaspuro, M., and Rautemaa, R. (2013). Acetaldehyde production and microbial colonization in oral squamous cell carcinoma and oral lichenoid disease. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* 116, 61–68. doi: 10.1016/j.oooo.2013.02.009

McCullough, M., Jaber, M., Barrett, A. W., Bain, L., Speight, P. M., and Porter, S. R. (2002). Oral yeast carriage correlates with presence of oral epithelial dysplasia. *Oral Oncol.* 38, 391–393. doi: 10.1016/S1368-8375(01)00079-3

Mills, A. L., and Alexander, M. (1976). N-Nitrosamine formation by cultures of several microorganisms. *Appl. Environ. Microbiol.* 31, 892–895.

Moritani, K., Takeshita, T., Shibata, Y., Ninomiya, T., Kiyohara, Y., and Yamashita, Y. (2015). Acetaldehyde production by major oral microbes. *Oral Dis.* 21, 748–754. doi: 10.1111/odi.12341

Nagy, K. N., Sonkodi, I., Szöke, I., Nagy, E., and Newman, H. N. (1998). The microflora associated with human oral carcinomas. *Oral Oncol.* 34, 304–308. doi: 10.1016/S1368-8375(98)80012-2

Perera, M., Al-hebshi, N. N., Speicher, D. J., Perera, I., and Johnson, N. W. (2016). Emerging role of bacteria in oral carcinogenesis: a review with special reference to perio-pathogenic bacteria. *J. Oral Microbiol.* 8:32762 . doi: 10.3402/jom.v8.32762

Pushalkar, S., Mane, S. P., Ji, X., Li, Y., Evans, C., Crasta, O. R., et al. (2011). Microbial diversity in saliva of oral squamous cell carcinoma. *FEMS Immunol. Med. Microbiol.* 61, 269–277. doi: 10.1111/j.1574-695X.2010.00773.x

Rubinstein, M. R., Wang, X., Liu, W., Hao, Y., Cai, G., and Han, Y. W. (2013). *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/b-catenin signaling via its FadA adhesin. *Cell Host Microbe* 14, 195–206. doi: 10.1016/j.chom.2013.07.012

Schloss, P. D. (2008). Evaluating different approaches that test whether microbial communities have the same structure. *ISME J.* 2, 265–275. doi: 10.1111/j.1365-294X.2007.03326.x

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Schmidt, B. L., Kuczynski, J., Bhattacharya, A., Huey, B., Corby, P. M., Queiroz, E. L. S., et al. (2014). Changes in abundance of oral microbiota associated with oral cancer. *PLOS ONE* 9:e98741. doi: 10.1371/journal.pone.0098741.s027

Scully, C., and Bagan, J. (2009). Oral squamous cell carcinoma: overview of current understanding of aetiopathogenesis and clinical implications. *Oral Dis.* 15, 388–399. doi: 10.1111/j.1601-0825.2009.01563.x

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60. doi: 10.1186/gb-2011-12-6-r60

Silverman, S., Gorsky, M., and Lozada, F. (1984). Oral leukoplakia and malignant transformation. A follow-up study of 257 patients. *Cancer* 53, 563–568. doi: 10.1002/1097-0142(19840201)53:3<563::AID-CNCR2820530332>3.0.CO;2-F

Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9440–9445. doi: 10.1073/pnas.1530509100

Warren, R. L., Freeman, D. J., Pleasance, S., Watson, P., Moore, R. A., Cochrane, K., et al. (2013). Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome* 1:16. doi: 10.1186/2049-2618-1-16

White, J. R., Nagarajan, N., and Pop, M. (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLOS Comput. Biol.* 5:e1000352. doi: 10.1371/journal.pcbi.1000352

Wu, J., Peters, B. A., Dominianni, C., Zhang, Y., Pei, Z., Yang, L., et al. (2016). Cigarette smoking and the oral microbiome in a large study of American adults. *ISME J.* 10, 2435–2446. doi: 10.1038/ismej.2016.37

Yamamura, K., Baba, Y., Nakagawa, S., Mima, K., Miyake, K., Nakamura, K., et al. (2016). Human microbiome *Fusobacterium nucleatum* in esophageal cancer tissue is associated with prognosis. *Clin. Cancer Res.* 22, 5574–5581. doi: 10.1158/1078-0432.CCR-16-1786

Yanik, E. L., Katki, H. A., Silverberg, M. J., Manos, M. M., Engels, E. A., and Chaturvedi, A. K. (2015). Leukoplakia, oral cavity cancer risk, and cancer survival in the U.S. elderly. *Cancer Prev. Res.* 8, 857–863. doi: 10.1158/1940-6207.CAPR-15-0091

**Conflict of Interest Statement:** 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.