



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

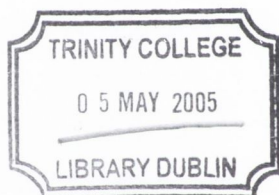
Y Chromosomes and Irish Phylogeography

Laoise Teresa Moore

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

Department of Genetics
University of Dublin
Trinity College

February 2005



THESIS
7624

Declaration

I declare that the work contained in this thesis has not previously been submitted for a degree at this or any other university. It is entirely my own work, except where otherwise acknowledged. I agree that Trinity College Library may lend a copy of this thesis upon request.

Signed

A handwritten signature in blue ink, appearing to read 'Laoise Moore', is written over a horizontal line.

Laoise Moore, January 2005

Acknowledgements

I would like to thank everyone who supported me during my Ph.D research.

In particular I would like to thank the following: Dr. Dan Bradley for his supervision, support, guidance, inspiration and patience and for giving me the opportunity to research such an interesting project; the Wellcome Trust for funding this research; everyone who donated DNA samples; the members and staff of the Genetics Department, particularly Sue, Brenda and Dave; also Brian, Emma, Caitríona, Ruth, Ceiridwen, Valeria and Dave, for all their help and for making the Bovine Lab such a pleasant place to work; my family Ciara, Fiona, Maeve, Simon, John C., Annika, Noah, Auntie Kate, Cathal, Tom, John M., Nana and Tigger for all their love and support, and for at least trying to understand what my Ph.D. is about; my friends, particularly Fiona D., Janna K. and Anne B.V. for keeping me sane (surely against the odds).

Finally and most importantly, I would like to thank my parents, Deirdre and Tony.

For everything.

Summary

A High-Resolution Y Chromosome Survey of Ireland

In order to investigate the origins of the Irish population and to describe the population genetic events that have shaped the extant Irish gene pool, Y chromosome genetic variation was analysed in 796 unrelated Irish males. High-resolution Y chromosome haplotypes were constructed by typing for 11 slowly-evolving SNP markers and 17 quickly-evolving microsatellite markers. Regional Irish genetic variation was examined and Irish Y chromosome genetic variation was placed in a European context by comparison with data from other European populations. The study revealed a high frequency of the putatively western European Palaeolithic Hg R1b3 in Ireland. When the Irish population is split into regional sub-populations a cline in Irish R1b3 frequency is revealed, which rises from the southeast to the northwest. This places northwestern Ireland on the extreme of a European-wide cline in the frequency of this Palaeolithic marker.

Investigating the Origins of the Irish Modal Haplotype

Within the Irish R1b3 population one 17-microsatellite haplotype has unusual features. It is the most common haplotype in the Irish sample, occurring 30 times. This Irish Modal Haplotype (IMH) with its one step neighbours accounts for ~8.2% of the Irish R1b3 sample and has a geographical distribution centred on the northwest, where it reaches frequencies of ~30%. The localisation of this haplotype in the northwest coincides with the territory of the most powerful and durable dynasty in Medieval Gaelic Ireland, the Uí Neill. Analysis of the Y chromosomes of 33 Irish males with surnames purported to derive from the Uí Neill lineage indicates the IMH is associated with the Uí Neill dynasty and suggests a rise in its frequency due to social selection associated with the hegemony of the Uí Neill and their descendant dynasties.

A Comparison of genetic diversity on the X and Y chromosomes in Ireland, Turkey and Togo.

Genetic diversity on the X and Y chromosome was examined in Ireland, Turkey and Togo. A comparison of the X haplotypes in the various populations suggests an Out of Africa bottleneck effect, whereby the Irish and Turkish X populations show markedly reduced diversity in comparison with the Togolese population. A comparison of the Y chromosome samples of each population suggests recent expansion in the Irish and Togolese populations. Estimates for the time frame of these expansions suggests they may have been due to the arrival of agriculture in these populations, which appears to have affected X and Y diversity to different degrees.

A Genetic Investigation of Newfoundland Origins

63 male samples from Newfoundland were genotyped for Y-linked loci in order to examine the genetic ancestry of this population. The Newfoundland haplotypes were compared with data from Irish and English populations in order to determine the relative contributions of these source populations to the Newfoundland Y chromosome population. The similarity of Y haplotypes among the three populations suggests that Y chromosome variation in Newfoundland is due to admixture of colonizing populations, which came mainly from southeast Ireland and England.

Autosomal microsatellite analysis of Irish, Danish and Basque samples.

78 Samples with ancestry from the east and the west of Ireland were analysed for 8 autosomal microsatellite loci in order to examine whether there is detectable regional genetic differences in the Irish population for these loci. Such underlying genetic structure may have epidemiological implications. Y chromosome analysis has suggested a possible Atlantic European versus Continental European gradient. The Irish data were compared with samples of English, Danish and Basque ancestry. Analysis using the Bayesian clustering algorithm implemented in Structure 2.1 (Pritchard *et al.* 2000), suggests no detectable genetic structure within these populations

Table of Contents

1	<u>General Introduction</u>	1
1.1	Approaches in the study of human population history	1
1.2	Modern Human Origins	1
1.3	Evolution of Modern Human Behaviour	2
1.4	Genetic evidence	3
1.5	Evidence from mtDNA and Y chromosomes	4
1.6	Expansions and dispersals of modern humans	4
1.7	Genetic variation in Eurasia	6
1.8	Anatomically Modern Humans in Europe	9
1.9	European Genetic Variation	9
1.10	The Last Glacial Maximum	10
1.11	Neolithic Y lineages?	13
1.12	Other European Population Movements	13
1.13	NR1 haplotypes and Irish male-specific population history	14
	<u>Chapter 2: A High-Resolution Y Chromosome Survey of Ireland</u>	16
2.1	Introduction	17
2.1.1	Western European Genetic Variation	17
2.1.2	Genetic Variation in the British Isles	18
2.1.3	The Common Genetic Heritage of the Atlantic Zone	18
2.1.4	Viking and Anglo-Saxon Genetic Influence in the British Isles	19
2.1.5	Irish Genetic History	22
2.1.6	Irish Prehistory	23
2.1.7	The Vikings in Ireland	24
2.1.8	The Norman Invasion	25
2.1.9	The Plantations	25
2.1.10	The Plantation of Ulster	26
2.2	Materials and Methods	27
2.2.1	Sample Collection	27
2.2.2	Single Nucleotide Polymorphism Analysis	28
2.2.3	SNP Genotyping Protocol	28

2.2.4	Microsatellite genotyping	31
2.2.5	Analysis	38
2.2.6	Fst pairwise population genetic distance analysis	39
2.2.7	Principal Components Analysis	40
2.2.8	Multidimensional Scaling Analysis	41
2.2.9	Median-joining Networks	41
2.2.10	Dating Estimates	41
2.3	Results	44
2.3.1	Haplogroup Frequencies	44
2.3.2	Microsatellite Haplotypes	45
2.3.3	Regional haplotype frequencies	46
2.3.4	Fst Pairwise Population Genetic Distances	47
2.3.5	Principal Components Analysis	49
2.3.6	Multidimensional Scaling Analysis	54
2.3.7	Network Analysis	55
2.3.8	BATWING posterior distributions	60
2.4	Discussion	63
2.4.1	Haplogroup frequency	63
2.4.2	Haplotype frequency	67
2.4.3	Genetic Distance Measures	69
2.4.4	Principal Components Analysis	69
2.4.5	Multidimensional Scaling Analysis	70
2.4.6	Median-joining networks	71
2.4.7	Dating Estimates	72
2.5	Conclusion	73
<u>Chapter 3: Investigating the Origins of the Irish Modal Haplotype</u>		75
3.1	Introduction	76
3.1.1	The regional frequency of IMH+1	76
3.1.2	Dating Expansion in the Northwest	76
3.1.3	The Genetic Legacy of the Mongols	77
3.1.4	Early Medieval Irish Society	80
3.1.5	The Uí Neill	81

3.2	Materials and Methods	83
3.2.1	Sampling Strategy	83
3.2.2	Genotyping and Analysis	83
3.3	Results	85
3.3.1	Batwing Posterior Distributions for Derry, Donegal, Tyrone R1b3 Samples	85
3.3.2	Network of Uí Neill R1b3 samples	88
3.3.3	Significant association of IMH+1 with the Uí Neill sample	89
3.4	Discussion and Conclusion	91

**Chapter 4: A Comparison of genetic diversity on the X and Y chromosomes
in Ireland, Turkey and Togo**

		92
4.1	Introduction	93
4.1.1	Ireland	93
4.1.2	Turkey	93
4.1.3	Togo	93
4.1.4	X and Y chromosome comparisons	94
4.1.5	The Irish Y chromosome sample	94
4.2	Materials and methods	95
4.2.1	X chromosome genotyping	95
4.2.2	Y Chromosome Genotyping	95
4.2.3	Analysis	96
4.2.4	Dating Estimates	97
4.3	Results	99
4.3.1	X Haplotype Analysis	99
4.3.2	Y Haplotype Analysis	99
4.3.3	Mismatch distributions	101
4.3.4	Median-joining networks	105
4.3.5	Irish X and Y Networks	108
4.3.6	Turkish X and Y median-joining networks	108
4.3.7	Togolese X and Y networks	109

4.3.8	Dating expansion in the Irish and Togolese populations	109
4.4	Discussion	111
4.4.1	Comparison of X haplotypes among the populations	111
4.4.2	Y haplotype comparisons among the populations	112
4.4.3	Bantu dispersals and the spread of agriculture	112
4.4.4	The Bantu Modal Haplotype	113
4.4.5	The onset of agriculture in Ireland	114
4.4.6	Discrepancy between the X and Y haplotypes	115

Chapter 5: A Genetic Investigation of Newfoundland Origins **118**

5.1	Introduction	119
5.1.1	The history of Newfoundland	119
5.1.2	The genetic history of Newfoundland	120
5.2.	Materials and Methods	121
5.2.1	Genotyping	121
5.2.2	Admixture Estimates	121
5.2.3	Irish and English source populations	122
5.3	Results	124
5.3.1	Haplotype frequency and sharing	124
5.3.2	Admixture Estimates	124
5.3.3	Principal Components Analysis	127
5.4	Discussion	128
5.5	Conclusion	131

Chapter 6 – Autosomal microsatellite analysis of Irish, Danish and Basque samples **132**

6.1	Introduction	133
6.1.1	Sex-specific loci and Irish population genetic history	133
6.1.2	Analysis of autosomal loci	133
6.1.3	Atlantic and Continental admixture?	133
6.1.4	Linkage disequilibrium, admixture and ethnic ancestry	134
6.1.5	Recent approaches to the analysis of autosomal loci	136
6.1.6	Detecting genetic structure within the Irish population	136

6.2	Materials and Methods	138
6.2.1	Selection of AIMs	138
6.2.2	Choosing samples for amplification	139
6.2.3	PCR amplification of loci	140
6.2.4	Analysis of genotype data	141
6.3	Results	143
6.3.1	Selection of loci with large Ds values	143
6.3.2	Genotype data	144
6.3.3	Structure Results	144
6.3.4	Genetic distance by population	148
6.3.4	MDS analysis	150
6.4	Discussion	152

Appendix

Section 1	Conversion from base pairs to repeat numbers
Section 2	Chapter 2 genotypes
Section 3	Chapter 3 genotypes
Section 4	Chapter 4 genotype
Section 5	Chapter 5 genotypes
Section 6	Chapter 6 genotypes
Section 7	Partial Y chromosome phylogeny

Literature Cited

Figures

Figure 1.7.1: The phylogeny and distribution of some Eurasian Y lineages	8
Figure 1.10.1: Geographic distribution of major Y lineages in Europe	12
Figure 2.2.2.1: Hierarchical SNP genotyping	28
Figure 2.2.5.1: Map of Ireland	39
Figure 2.3.3.1; Contour Map of IMH+1 frequency in Ireland	48
Figure 2.3.5.1: PC1 vs. PC2 of Irish and Capelli <i>et al.</i> (2003) data.	51
Figure 2.3.5.2: The impact of each Hg on the plot of PC1 vs. PC2	52
Figure 2.3.5.3: Plot of PC1 vs. PC3 of Irish and Capelli <i>et al.</i> (2003) data.	52
Figure 2.3.5.4: The impact of each Hg on the plot of PC1 vs. PC3	53
Figure 2.3.6.1 MDS plot of Irish and Capelli <i>et al.</i> 2003 populations	54
Figure 2.3.7.1: Median-Joining network of Northwest R1b3 samples. IMH	55
Figure 2.3.7.2: Median-Joining network of Northeast R1b3 samples	56
Figure 2.3.7.3: Median-Joining network of North Connaught R1b3 samples	56
Figure 2.3.7.4: Median-Joining network of Midlands North R1b3 samples	57
Figure 2.3.7.5: Median-Joining network of Dublin R1b3 samples	57
Figure 2.3.7.6: Median-Joining network of Galway, Clare, Tipperary R1b3 samples	58
Figure 2.3.7.7: Median-Joining network of Midlands South R1b3 samples	58
Figure 2.3.7.8: Median-Joining network of Southwest R1b3 samples	59
Figure 2.3.7.9: Median-Joining network of Southeast R1b3 samples	59
Figure 2.3.8.1: Posterior distribution for N_e , ancestral effective population size	60
Figure 2.3.8.2: Posterior distribution of alpha	61
Figure 2.3.8.3: Posterior distribution of TMRCA of Irish R1b3 samples	61
Figure 2.3.8.4: Posterior distribution of beta	62
Figure 3.1.1.1: Median-joining network of R1b3 samples from Donegal, Derry, Tyrone, Fermanagh, Leitrim, Mayo, Sligo, Roscommon.	79
Figure 3.1.1.2: Median-joining network of R1b3 samples from Cork, Kerry and Limerick	80

Figure 3.3.1.1: BATWING posterior distribution for estimate of N	85
Figure 3.3.1.2: BATWING posterior distribution for estimate of Alpha	85
Figure 3.3.1.3: BATWING posterior distribution for estimate of TMRCA	86
Figure 3.3.1.4: BATWING posterior distribution for Beta	86
Figure 3.3.2.1: Median-Joining Network of Uí Neill R1b3 samples	88
Figure 3.3.3.1: Number of differences from IMH of R1b3 chromosomes from the Uí Neill sample (blue) versus a general Ulster sample with no surname information attached (red)	89
Figure 4.3.3.1: X West Mismatch Distribution	101
Figure 4.3.3.2: X East Mismatch Distribution	101
Figure 4.3.3.3: X Turkey Mismatch Distribution	101
Figure 4.3.3.4: X Togolese Mismatch Distribution	101
Figure 4.3.3.4: Y West Mismatch Distribution	102
Figure 4.3.3.5: Y East Mismatch Distribution	102
Figure 4.3.3.6: Y Turkey Mismatch Distribution	102
Figure 4.3.3.7: Y Togolese Mismatch Distribution	102
Figure 4.3.3.8: Irish R1b3 Mismatch Distribution	103
Figure 4.3.3.9: Togolese E3a Mismatch Distribution	103
Figure 4.3.4.1: Median-joining network of Irish X chromosomes	105
Figure 4.3.4.2: Median-joining network of Irish R1b3 Y chromosomes	105
Figure 4.3.4.3: Median-joining network of Turkish X chromosomes	106
Figure 4.3.4.4: Median-joining network of Turkish Y chromosomes	106
Figure 4.3.4.5: Median-joining network of Togolese X chromosomes	107
Figure 4.3.4.6: Median-joining network of Togolese Y chromosomes	107
Figure 5.3.2.1: The relationship between pIre, pEng and pNf, the haplotype frequencies for Ireland, England and Newfoundland respectively	125
Figure 5.3.2.2: The slope of the least-squares regression line approximates Irish ancestry in the Newfoundland-Irish sample	125
Figure 5.3.2.3: The slope of the least squares regression line approximates the proportion of Irish ancestry in the Newfoundland-English sample	126

Figure 5.3.2.4. The slope of the least squares regression line approximates the proportion of Southeast Irish ancestry in the Newfoundland-Irish sample	126
Figure 5.3.2.5 The slope of the least squares regression line approximates the proportion of Southeast Irish ancestry in the Newfoundland-English samples	126
Figure 5.3.3.1: Principal Components plot of PC1 vs. PC2	127
Figure 6.3.1.1: The frequency distribution of Ds values obtained for the Basque versus the French samples of Rosenberg et al. (2002)	143
Figure 6.3.1.2: Scatter plot comparing Ds values obtained from Basque vs. French and Basque vs. Europeans	144
Figure 6.3.3.1: variation of estimated log probability varies for different values of K, the number of inferred clusters in the data	145
Figure 6.3.3.2: Structure outputs for K=1 to K=5	146-148
Figure 6.3.4.1: Ds distances for each locus calculated by population	149
Figure 6.3.4.2: Average Ds for Basques vs. other populations	150
Figure 6.3.5.1: MDS plot constructed from an allele sharing matrix of 8 unlinked microsatellite loci showing the relationship of individuals	150

Tables

Table 2.2.3.1 SNP Information	33
Table 2.2.3.2 Tetra-primer ARMS PCR Information	34
Table 2.2.4.1 Microsatellite Information	35-36
Table 2.3.1.1: Haplogroup frequencies in the Irish Sample	44
Table 2.3.1.2: Haplogroup frequencies in 9 geographic regions of Ireland	44
Table 2.3.4.1: Matrix of Fst Pairwise population genetic distances	47
Table 2.3.4.2: Matrix of significant Fst P-values	49
Table 2.3.5.1: PC1, PC2 and PC3	50
Table 2.3.5.2: The impact of each haplogroup on the PC plot	51
Table 2.3.8.5: Mode, Median and 95% Credible Intervals for Batwing outputs	62
Table 3.3.3.2: Asymptote 2-tailed significance results for Mann Whitney comparisons of Ui Neill R1b3 samples versus (i) an Ulster sample with no surname information, (ii) the Northwest sample of Chapter 2	90
Table 4.3.1.1: X chromosome pairwise population genetic distances	100
Table 4.3.1.2: Matrix of significant P-values for X Fst distances	100
Table 4.3.2.1: Y chromosome pairwise population genetic distances	100
Table 4.3.2.2: matrix of significant P-values for Y Fst distances	100
Table 6.2.3.1: Chromosomal location and primers for 10 loci selected for amplification	140

Abbreviations

AMH	Anatomically Modern Humans
bp	Base pairs
CI	Credible Interval
Hg	Haplogroup
Ht	Haplotype
MDS	Multidimensional Scaling
MtDNA	Mitochondrial DNA
NRY	Non-recombining portion of the Y chromosome
PC	Principal Component
TMRCA	Time to most recent common ancestor
YBP	Years Before Present

1 General Introduction

1.1 Approaches in the study of human population history

What is known of the evolution and prehistory of humans has come from several disciplines; including palaeontology, archaeology, climatology, linguistics and genetics. Since the late twentieth century the study of extant genetic variation in particular has enhanced our understanding of the origins and dispersals of modern human populations. This has been due to an explosion in available population genetic research tools, including the development of assays for many informative loci in the genome. The synthesis of genetic evidence with that from other disciplines has allowed a multidisciplinary approach to the study of human population history.

1.2 Modern Human Origins

Palaeontological, archaeological and genetic evidence support a sub-Saharan African origin for *Homo sapiens sapiens*, or anatomically modern humans (AMH) within the last 200,000 years (Foley 1998, Stringer 2000, Cann 2001, Cavalli-Sforza and Feldman 2003). Features of modern human anatomy are evident in the African fossil record by ~160,000 years before present (YBP) (White 2003, Stringer 2003). The African fossil record indicates skull evolution from *Homo erectus* through archaic *Homo sapiens* to modern *Homo sapiens*. The presence of the earlier types in Africa, together with the continuity from ancient to modern types, further support the hypothesis that the evolution of anatomically modern humans occurred in Africa.

1.3 Evolution of Modern Human Behaviour

The dispersal of modern humans from Africa to colonise the rest of the world has often been linked with the evolution of modern human behaviour. Klein (1999) has argued that the beginning of modern human behaviour coincided with and facilitated the main dispersal of modern humans from Africa after the Late Stone Age (equivalent to the Upper Palaeolithic in Eurasia ~40,000 – 50,000 YBP).

Besides large scale migrations, evidence for the evolution of modern human behaviour is usually considered to be manifested in the archaeological record as an increasingly varied Aurignacian-like stone toolkit, the diversification of materials, including bone, used for tools, trade, the exploitation of marine resources, long-distance transport, including by boat and the proliferation of art and decoration, among other things.

The concept that modern human behaviour arose suddenly throughout the Old World ~45,000 – 50,000 YBP through a cognitive advance that occurred long after the morphological evolution of the anatomically modern human form in Africa is referred to as the ‘human revolution’ model. Isaac (1976) suggested that these archaeological manifestations of modern behaviour were accompanied by the development of modern language.

However this view has become increasingly disputed. Archaeological evidence indicates that there was at least one dispersal of modern humans from Africa during

the Middle Palaeolithic and that both outside and within Africa aspects of modern human behaviour existed by then (Stringer 2000). The 'human revolution' model, it has been argued, is due to a Eurocentric bias and the failure to appreciate the depth and breadth of the African archaeological record, which indicates the presences of bone tools, demic expansion, use of aquatic resources, long distance trade, and uses of art and decoration in Africa during the Middle Stone Age (Middle Palaeolithic in Eurasia) (McBrearty and Brooks 2000).

1.4 Genetic evidence

There is strong genetic evidence based on multiple loci for a single, recent (<200,000 YBP) origin of modern humans in Africa. This data comes from Single Nucleotide Polymorphisms (SNPs), microsatellites and Y chromosome and mtDNA haplotypes. Compared with chimps and other apes, humans are relatively lacking in diversity. Variation levels in mtDNA, Y chromosomes, human lymphocyte antigen (HLA), beta globin, the zinc finger gene on the X chromosome (ZFX) intron and microsatellites all suggest an effective human population size of about 10,000 in Africa during the Upper Pleistocene, through a long population bottleneck (Hewitt 2000, Cann 2001). African populations are more diverse genetically than those found anywhere in the rest of the world and for the most part, non-African genetic variation can be treated as a subset of African variation (Foley 1998). Loci which do not conform to this system are usually targets of natural selection (Cann 2001).

1.5 Evidence from mtDNA and Y chromosomes

The reconstruction of gene genealogies of non-recombining regions of the genome, such as mtDNA and the non-recombining portion of the Y chromosome (NRY), has been useful in the study of human population history. The mtDNA and Y chromosome phylogenies are reasonably well resolved. The evolutionary relationship among lineages, together with their geographic distribution (the study of phylogeography), allows the inference of population history and past migrations. For example, the roots and deepest branches in the phylogenies of mtDNA and NRY both occur in sub-Saharan African populations, whereas all non-African lineages are located at later branches in the phylogeny, consistent with an African origin of the species, and subsequent dispersal out of Africa (Underhill *et al.* 1997, Forster *et al.* 2001, Maca-Meyer *et al.* 2001, Jobling and Tyler-Smith 2003). Dating estimates of the time to most recent common ancestor (TMRCA) of all extant mtDNA and NRY lineages fall consistently around ~100,000 to ~200,000 YBP, supporting a recent origin of modern humans.

1.6 Expansions and dispersals of modern humans

Archaeological and palaeontological evidence suggests that modern humans dispersed from Africa to the rest of the world on multiple occasions and by multiple routes (Foley 1998). The earliest evidence for AMH presence outside Africa comes from fossil and archaeological remains at Mount Carmel and Qafzeh in Israel, which have been dated to ~100,000 YBP (Cavalli-Sforza *et al.* 1994). This expansion out of Africa appears to have been temporary, however, as Neandertals reoccupied the

region from ~70,000 YBP. This early expansion of AMH to the Middle East is coincident with a period of climatic amelioration from 130,000 – 71,000 YBP. The return of Neandertals to the region is coincident with the onset of a cold spell initiated by the eruption of mount Toba in Sumatra (Hewitt 2000). Archaeological evidence indicates that anatomically modern populations using Middle Palaeolithic technology probably reached Australia by ~60,000 YBP (Thorne *et al.* 1999). Populations using modern technology are detectable in the Middle East from ~47,000 YBP and in Western Europe from ~43,000 YBP (Mellars 1992, Jobling and Tyler-Smith 2003).

Evidence from mtDNA and NRY also indicates early expansions of modern humans. The roots and early branches of both phylogenies occur in Africa, and appear to represent hunter-gatherer populations (Cavalli-Sforza and Feldman 2003). Early expansions across Africa appear to have left traces in the mtDNA of Bushmen and West African Pygmies (Watson *et al.* 1997). A subsequent expansion in East Africa between 80,000 – 60,000 YBP appears to have spread throughout Africa and to have ultimately led to the migration of one or two mtDNA types out of Africa and into Asia, represented by the M and N lineages (Forster *et al.* 2001, Maca-Meyer *et al.* 2001). Forster *et al.* (2001) calculated the age of the most recent common ancestor of all extant non-African mtDNA lineages to ~54,000 YBP. One calibration of NRY evidence indicates an early expansion within Africa ~100,000 YBP, followed by a later expansion ~60,000 – 40,000 YBP, which spread throughout much of Africa and also into Asia (Cavalli-Sforza and Feldman 2003).

The male and female genealogies suggest two possible routes taken by modern humans dispersing from Africa. The first is a southern route through the Horn of Africa, through India to South and Southeast Asia. These humans reached Oceania by ~40,000 – 60,000 YBP. The second is through the Middle East to Central Asia and from there to Europe, East and Northeast Asia ~40,000 YBP (Cavalli-Sforza and Feldman 2003). However it must be noted that the study of extant genetic variation can only illuminate the routes taken by those populations whose descendants have survived to the present day. Genetics studies (except possibly those of ancient DNA) cannot inform us of possible earlier dispersals of groups who did not leave descendants.

1.7 Genetic variation in Eurasia

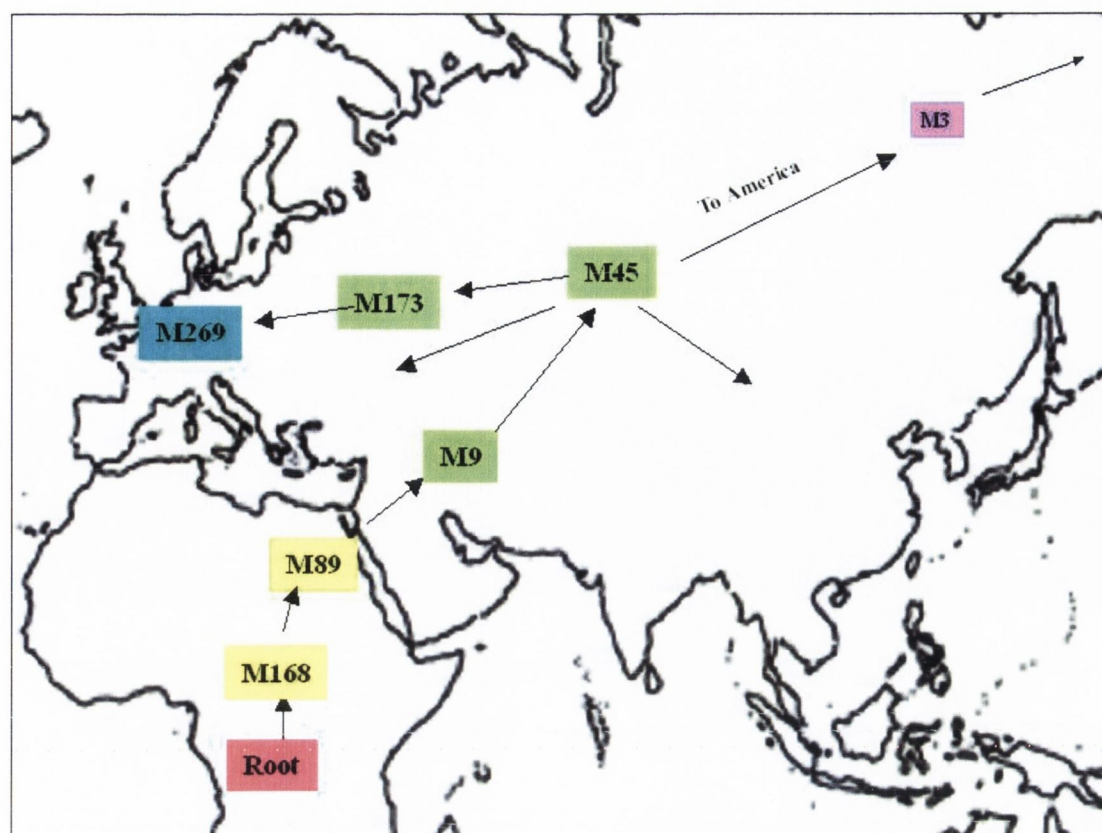
The high diversity levels of mtDNA and NRY lineages in central Asia support the theory of early dispersal of humans into this continent upon leaving Africa (Wells *et al.* 2001, Quintana-Murci *et al.* 2004). A recent study of Eurasian Y chromosome variation has shown that central Asia is a major reservoir of Y chromosome genetic diversity and that this region was the likely source of major migrations to Europe, India and the Americas (Wells *et al.* 2001).

Central Asian populations have the highest Y chromosome SNP and microsatellite diversity in Eurasia, suggesting these populations are among the oldest on the continent. The ages of Y lineages in Eurasia, estimated by assessing the amount of

microsatellite variation within lineages, suggest an early settlement of central Asia by modern humans possibly 40,000 – 50,000 YBP.

On the NRY phylogeny M45 is the ancestor of both the major European M173 lineage and the major Native American M3 lineage. M45 is descended from M9, the Y lineage thought to have been present in a small population emerging from Africa via the Middle East (Figure 1.7.1). M45 is found at polymorphic frequency mainly in central Asia. Its descendants include the major western European lineage M173 and the major Native American lineage M3. This suggests that a central Asian population was the source of major dispersals both to Europe and the Americas. The age of M45 is estimated at ~40,000 YBP (Wells *et al.* 2001).

The age estimate (~30,000 YBP), phylogeny and high frequency in Europe (~50%) of M173 are consistent with it being brought to Europe in one of the earliest modern human migrations to this continent from Asia, during the Upper Palaeolithic. Descendants of the M173 lineage include M269 and M17, both found in Europe.



Partial Y chromosome phylogeny

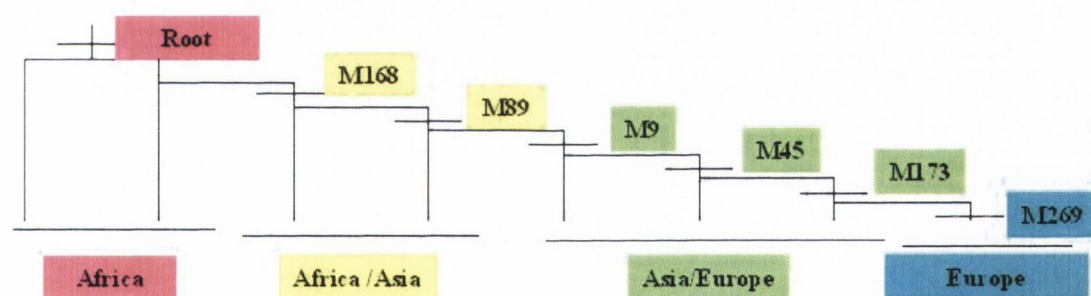


Figure 1.7.1: The phylogeny and distribution of some Eurasian Y lineages

Mitochondrial DNA diversity in Asia also provides evidence of dispersals of modern humans to populate the rest of the world. The expansion out of Africa of the L3 mtDNA lineage was followed by its split into the M and N lineages. Descendants of the M and N lineages expanded in Asia and went on to colonise East Asia, Oceania and America (represented by the mtDNA lineages A, B, C, D, E, F, G and M), and to colonise Europe, North Africa and Western Asia (represented by the mtDNA lineages H, I, J, K, T, U, V, W, X).

1.8 Anatomically Modern Humans in Europe

Archaeological and genetic evidence suggests that anatomically modern humans first reached the European continent ~40,000 YBP during the Upper Palaeolithic (Mellars 1992, Cavalli-Sforza *et al.* 1994, Hewitt 2000). The phylogeography of European Y and mtDNA lineages suggest that extant Europeans are largely the descendants of various Palaeolithic and Neolithic migrants (Semino *et al.* 2000, Rosser *et al.* 2000, Richards *et al.* 2000)

1.9 European Genetic Variation

Clines exist within Europe for many loci, which are likely the traces of past population movements. Notably, principal components analysis of 95 ‘classical’ (protein) polymorphisms by Cavalli-Sforza *et al.* (1994) revealed clines in gene frequency within and across Europe. More recently mtDNA sequence variation, Y chromosome and autosomal SNP and microsatellite variation have revealed non-random patterns within European genetic variation also (Richards *et al.* 2002,

McEvoy *et al.* 2004, Semino *et al.* 2000, Rosser *et al.* 2000). The first PC of Cavalli-Sforza *et al.*'s analysis shows a continent-wide southeast – northwest cline across Europe, which has been interpreted as reflecting the migration of farmers from the Near East into Europe after the advent of agriculture in the Fertile Crescent ~10,000 YBP. Similar southeast – northwest clines in gene frequency have been observed for other loci, most notably Y chromosome haplotypes (Semino *et al.* 1996, Semino *et al.* 2000, Rosser *et al.* 2000).

However the degree of demic diffusion of farmers from the Near East into Europe has been debated extensively (Cavalli-Sforza *et al.* 1994, Cavalli-Sforza and Minch 1997, Chikhi, *et al.* 1998, Richards *et al.* 1996, 1997, 2000). The degree of variation within mtDNA and Y chromosome lineages in extant European populations suggest that some lineages have been in Europe since the Palaeolithic, whereas others entered during the Neolithic, or with other migrational events (Richards *et al.* 1996, Richards *et al.* 2000, Semino *et al.* 2000, Rosser *et al.* 2000). European mtDNA and Y chromosome variation suggests that other population movements have also had major effects on the composition of the European gene pool, the most striking of these being the effects of the last ice age on the Palaeolithic European population (Semino *et al.* 2000, Torroni *et al.* 2001).

1.10 The Last Glacial Maximum

During the Last Glacial Maximum (LGM, ~20,000 – 13,000 YBP) human populations largely vacated the central European plain and retreated to a number of

refuge areas, possibly the Iberian Peninsula, Italy and the Balkans and present-day Ukraine. The archaeological record for this period indicates almost total vacation of central Europe by humans, in contrast with evidence for population persistence in the refuge areas (Otte 1990, Hewitt 2000). It is expected that this contraction would have produced severe population bottlenecks. Upon climatic amelioration from ~13,000 YBP, it is thought that these refuge populations expanded northwards and westwards to re-colonise the European plain. Patterns of expansion during this time are visible in the archaeological record. For example Otte (1990) has suggested an expansion from the Iberian refuge after the LGM, which is manifested in the archaeological record as the spread of the Magdalenian material culture from the Iberian Peninsula and southern France up through the Atlantic coast of Europe.

In terms of genetic variation, the 3rd PC of Cavalli Sforza *et al.*'s European data shows a cline with a similar pattern of expansion northward from the Iberian Peninsula, as does a cline in mtDNA haplogroup (Hg) V frequency. Torroni *et al.* (2001) have suggested that the age of mtDNA Hg V is consistent with this scenario. Semino *et al.* (2000) suggested the high frequency of M173 (Hg R1b) in Western Europe is due to a similar re-expansion from an Iberian refuge. These patterns suggest that climatic change had major effects on the genetic composition of Europeans. All Y chromosome Hg designations follow the nomenclature proposed by the Y Chromosome Consortium (2002).

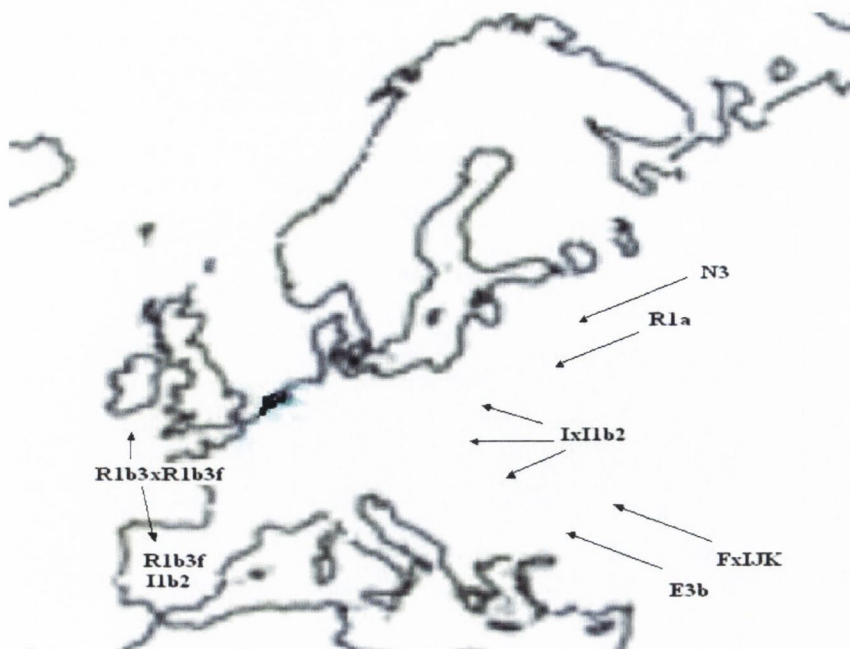


Figure 1.10.1: Geographic distribution of major Y lineages in Europe. The position of each Hg on the map approximates its region of highest frequency in Europe.

The geographic distribution of the major European Hg I, defined by the SNP M170, indicates that the carriers of this lineage were also affected by the LGM. This lineage is mainly confined to Europe, where it reaches a peak in frequency in central eastern Europe. Semino *et al.* (2000) have estimated the age of this lineage at ~22,000 YBP and have suggested that it originated in men whose descendants entered Europe from the Middle East 20,000 – 25,000 YBP. The geographic distribution of this lineage suggests that during the LGM some carriers of this lineage may have retreated to the Ukrainian/Balkan refuge. This lineage is also found in the Basque and Sardinian populations, which have accumulated a separate mutation, M26, defining Hg I1b2. The geographic distribution of I1b2 suggests that it formed part of the Iberian refuge population during the LGM (Semino *et al.* 2000, Rootsi *et al.* 2004).

1.11 Neolithic Y lineages?

There is a southeast – northwest decline in frequency of Y chromosomes carrying the SNPs M35 and M89, defining the Hgs E3b and FxIJK. The ages of these lineages in Europe are estimated at 15,000 – 20,000 YBP (Semino *et al.* 2000). Together with the southwest – northeast clinal pattern, this indicates that these lineages arrived in Europe from the Middle East after the LGM and may have been introduced to Europe in the Neolithic.

1.12 Other European Population Movements

Other more regional patterns of genetic variation in the European population suggest that more local population events have affected the composition of the European gene pool. For example the M17 mutation, found on some M173 Y chromosomes, defines Hg R1a. This lineage occurs at highest frequencies in central Europe (~50% in Russia, Poland and Slovakia), but is almost absent in the southeast and southwest (Rosser *et al.* 2000, Semino *et al.* 2000). It has also been observed at substantial frequency in northern India, Pakistan and Central Asia (Semino *et al.* 2000). Its distribution in Europe is similar to the 3rd PC of European classical gene frequencies which has been interpreted as marking the expansion of the Kurgan people from the north of the Caspian sea ~7,000 YBP, although other population movements may have been responsible (Cavalli-Sforza *et al.* 1994)

Rosser *et al.* (2000) observed a high frequency of Hg N3, defined by the tat mutation in the north, east of the Baltic Sea. They found that this marker was relatively frequent in all Uralic-speaking populations, except Hungarians who are proposed to

have acquired the language from the Magyars. This marker has been dated to ~4,000 YBP (Zerjal *et al.* 1997) and is proposed to mark the Uralic migration to northern Europe.

1.13 NRY haplotypes and Irish male-specific population history

It is clear that many genetic systems show regional patterns for European variation. While these patterns have been illustrated across multiple loci, including protein polymorphisms, disease gene frequency and autosomal DNA polymorphisms, the most detailed picture of European genetic history has come from the study of polymorphisms on mtDNA and NRY. These systems are particularly useful in deciphering human population history for numerous reasons. They do not recombine, enabling detailed haplotype reconstruction from the assaying of multiple markers. Their relatively small effective population size accentuates the effects of genetic drift, and their inheritance mechanism facilitates the reconstruction of sex-specific population history.

The study of NRY variation allows the reconstruction of aspects of male lineage population history. Y chromosomes show greater between-population differences and a stronger geographic signal than other genetic systems. This is possibly due to ethnic and social differences in male and female behaviour, such as greater male than female variability in reproductive success, higher rates of polygamy than polyandry and historic patterns of patrilocality (Seielstad *et al.* 1998) (however higher rates of female mobility on a global scale have recently been questioned (Wilder *et al.* 2004)).

Nevertheless, recent acceleration in Y chromosome polymorphism discovery means that the NRY is now the best-characterised haplotypic system in the human genome (Jobling and Tyler-Smith 2000).

Studies of European NRY variation have attested to the power of NRY haplotype reconstruction in the deciphering of male population history. Ireland occupies a peripheral position on the northwestern edge of Europe. Previous studies have indicated that its Y chromosome population is peripheral in a European sense also (Hill *et al.* 2000, Wilson *et al.* 2001b). A high-resolution reconstruction of Irish NRY haplotype variation using informative markers should illuminate aspects of Irish male-specific population history and place them in a European context.

Chapter 2: A High-Resolution Y Chromosome Survey of Ireland

In order to investigate the origins of the Irish population and to describe the population genetic events that have shaped the extant Irish gene pool, Y chromosome genetic variation was analysed in 796 unrelated Irish males. High-resolution Y chromosome haplotypes were constructed by typing for 11 slowly-evolving SNP markers and 17 quickly-evolving microsatellite markers. Regional Irish genetic variation was examined by subdividing samples into geographic regions, and Irish Y chromosome genetic variation was placed in a European context by comparison with data from other European populations.

2.1 Introduction

2.1.1 Western European Genetic Variation

It appears that the combination of migrations to Europe from the Near East and northeast Asia, together with the postglacial re-expansion of populations bearing Palaeolithic Y-lineages from an Iberian refuge into western Europe, have produced an Atlantic Y chromosome population that is remarkably homogeneous.

Genetic similarities among these populations, which include the Basque, British and Celtic populations, have been noted. For example these populations cluster together in the first Principal Component of European autosomal genetic variation (Cavalli-Sforza *et al.* 1994). The Basques, thought to be descended from the Palaeolithic inhabitants of Europe because they speak a non-Indo European language, are one of the consistent genetic outliers with respect to European genetic variation. Their mtDNA is thought to have the lowest Neolithic input for any European population (Wilson *et al.* 2001b). With respect to Y chromosome variation, the Irish, Basque and Welsh populations have a very high frequency of the putatively Palaeolithic Hg R1b. MtDNA lineages, when analysed in sufficient numbers, also show significantly similar European-wide east – west patterns to Y chromosome and autosomal variation. MtDNA variation indicates shared genetic ancestry throughout the Atlantic zone, with relatively low Neolithic input in maternal lineages in these populations

(Mc Evoy *et al.* 2004). This concordance in mtDNA, Y-chromosome and autosomal genetic variation suggests a common demographic history for these populations.

2.1.2 Genetic Variation in the British Isles

Wilson *et al.* (2001b) used the comparison of signature haplotypes to identify possible source populations for the extant Y chromosome gene pool in the British Isles. They compared Y samples from Orkney, Gaelic Ireland and Wales to Basque, Norwegian and Frisian samples, and argued that the striking similarity of the Welsh and Irish to the putatively Palaeolithic Basque sample indicated a substantial Palaeolithic component in these Atlantic Celtic Y chromosome populations.

2.1.3 The Common Genetic Heritage of the Atlantic Zone

The Basque, Gaelic Irish and Welsh samples all contained the putatively Palaeolithic Hg R1b at high frequency (89-90%), with most remaining samples belonging to Hg Ix11b2. When haplotypes were constructed from analysis of 6 Y-linked microsatellite markers, one haplotype occurred at high frequency on R1b chromosomes, the Atlantic Modal Haplotype. This haplotype, together with its one-step neighbours (AMH+1), accounted for 56% of the Basque, 70% of the Welsh and 44% of the Irish samples. It occurred at substantially higher frequency in these Atlantic populations than in mainland European and Near Eastern samples. The similarity of these Y chromosome populations suggests that a substantial proportion of the gene pool of the British Isles is western European Palaeolithic in origin, has been relatively undisturbed by population movements following initial colonization, and that the

indigenous Y chromosome population of the British Isles is likely to have carried haplogroup R1b at high frequency.

2.1.4 Viking and Anglo-Saxon Genetic Influence in the British Isles

The occurrence of signature haplotypes also suggested Norwegian and Anglo-Saxon genetic input in regions of the British Isles. A comparison of British and Norwegian haplotypes indicated that a haplotype on the R1a lineage, haplotype 3.65 and a haplotype on the Ix11b2 lineage, haplotype 2.47, are suggestive of Viking genetic introgression in Orkney, based on their high frequency in the Norwegian sample (2.47+1 38% and 3.65 22%) and in the Orcadian sample (2.47+1 7% and 3.65 20%). Hg 2.47+1 was also found at substantial frequency in the Frisian sample (16%), thought to be representative of the source population for supposed Anglo-Saxon migration to Britain, indicating that it may not be possible to distinguish whether the presence of this haplotype is due to Anglo-Saxon or Norwegian Viking genetic influence in regions of Britain where there is historical evidence for the presence of both groups. The low frequency of haplotypes 2.47 and 3.65 in the Gaelic Irish and Welsh samples suggested that Anglo-Saxon and Viking genetic input to these populations was modest.

Weale *et al.* (2002) argued for a substantial Anglo-Saxon contribution to the east and central mainland English gene pool. They sampled Y chromosomes from 7 towns on an east-west transect from East Anglia to Northern Wales and compared these samples with Frisian and Norwegian datasets which represented source populations

for continental immigrants. They found that while the central English samples were statistically indistinguishable from one another and from the Frisian sample, they were significantly different from the Welsh samples. They then used novel model-based population genetic methods to make inferences about both the possible timing and extent of male continental migration into Central England. They developed an inference method that explored a range of historical scenarios involving continuous gene flow and mass migration to evaluate whether or not a large Anglo-Saxon migration event is needed to explain the extremely high central English – Frisian affinity. They concluded that the observed patterns of genetic variation were best explained by substantial migration of Anglo-Saxon Y chromosomes, which replaced 50 – 100% of the indigenous Y chromosomes of Central England, but not of North Wales.

A more comprehensive survey of British Y chromosome variation was carried out by Capelli *et al.* (2003), who sampled 1772 Y chromosomes from small urban locations in the British Isles, Norway, Denmark and North Germany. The Norwegian sample represented the source population for Norwegian Vikings, the Danish sample represented the source population for Danish Vikings, and the North German sample represented the source population for the Anglo-Saxons. The Danish and North German samples were statistically indistinguishable, indicating the difficulty of distinguishing Danish Viking from Anglo-Saxon genetic influence in the British Isles. Two samples from Ireland (Rush on the east coast and Castlerea in the midlands) were used to represent the indigenous population of the British Isles.

By comparison of the signature haplotypes discussed above, they concluded that Norwegian Vikings had significant genetic influence on the northern areas of the British Isles, including Orkney, Shetland and the extreme north of Scotland, whereas evidence for Anglo-Saxon / Danish Viking influence was strongest in central England and in the east, in York, Norfolk and Southwell.

When they used Principal Components plots to summarise the variation they found that the position of a population on the first 2 axes of the plot provided an indicator of the degree of continental input into an indigenous background, with the Celtic samples skewed to the right of axis 1, Norwegian input moving populations left along axis 1 and German/Danish influence moving populations at an angle through both axes.

The Y chromosome gene pool of southern England appeared to be predominantly indigenous, judging by the high frequency of AMH+1 in these populations and their positions to the right of axis 1 on the PC plot (Capelli *et al.* 2003). Given that historically Anglo-Saxons are thought to have occupied southern England while the activities of the Danish Vikings were mainly in the east, Capelli *et al.* argued that in England the Danes had a greater demographic impact than the Anglo-Saxons, contrary to the finding of Weale *et al.*, whose analysis did not include a Danish sample. Capelli *et al.*'s survey showed that the frequency of AMH+1 was significantly higher in all British populations (always above 33%) than in mainland

Europe (always below 26%). If the presence of AMH+1 is accepted as diagnostic of the indigenous British population, this suggests that there is some indigenous component in all British regions and argues against population replacement events.

2.1.5 Irish Genetic History

Capelli *et al.*'s analysis of the central Irish Castlerea sample suggested a predominantly indigenous Y chromosome population, indicated by the high frequency of R1b and AMH+1. This is consistent with the lack of historical evidence for Anglo-Saxon or Viking contact with this region. They concluded that the eastern Irish sample from Rush, on the other hand, did show some evidence of Viking influence, indicated by its position on the PC plot further to the left of axis 1 than other Celtic populations. A comparison of the Rush and Castlerea samples indicates that this is probably due to the higher proportion of AMH+1 in the Castlerea sample and the presence of R1a samples in the Rush sub-population.

A genetic survey of Irish Y chromosomes (Hill *et al.* 2000) which subdivided samples on the basis of ancient geographic information contained in surnames (largely surnames, like Y chromosomes, are paternally inherited) found that samples with Gaelic Irish surnames showed significant difference from samples with non-Gaelic surnames, with the Gaelic surname samples containing higher frequency of the putatively Palaeolithic haplogroup R1b (Hg1 in this reference). When the Gaelic surname samples were subdivided on the basis of geography into regions based on the four ancient provinces of Ireland, the frequency of Hg R1b followed an east – west

distribution across Ireland, reaching near fixation in Connaught, the westernmost province (98.3%). That R1b is the indigenous Irish variant was also suggested by the topology of a maximum parsimony network of Gaelic Irish Y chromosomes. Hg R1b samples occurred at high frequency and clustered tightly at the centre of the network, whereas non-R1b samples did not show this coherence, consistent with being immigrants. A higher frequency of non-R1b chromosomes in the eastern Gaelic subpopulations is consistent with a greater degree of admixture with non-indigenous groups in the east and south, or possibly with prehistoric genetic influx to the east of the island.

Hill *et al.* dated R1b coalescence in the Irish sample to ~4,200 YBP, roughly coincident with the spread of agriculture to the island, expected to cause population expansion. However confidence intervals associated with this estimate were large, encompassing a time range for numerous possible demographic events. The extremely high frequency of R1b in Ireland suggests that agriculture, when it came to Ireland, was not spread by demic diffusion, but rather may have been adopted by the indigenous Mesolithic population.

2.1.6 Irish Prehistory

Besides the introduction of agriculture, there are numerous candidate events for demographic change in Ireland. Archaeological evidence indicates that the earliest human presence in Ireland was postglacial. The last ice age ended in Ireland ~12,000 years BP and the earliest radiocarbon date for human occupation is ~9,000 years BP.

The earliest large-scale hunter-gatherer settlements date from ~7,500 – 5,000 years BP. The archaeological record does not provide evidence for a rapid transition to farming, but it does show evidence of agriculture from ~6,000 years BP, and for intensification of agriculture between 6,000 and 4,000 years BP, manifested in the archaeological record as domestic animal remains and signs of forest clearance. Megalithic tombs date from ~4,000 – 2,000 years BP. Other major prehistoric events indicated in the archaeological record include the Bronze Age, which dates from the end of the 3rd Century BC and the Iron Age, which dates from 600 BC to 400 AD and is associated with the La Tene cultural complex (Waddell 2000).

2.1.7 The Vikings in Ireland

In the historical period Ireland experienced numerous invasions. Unlike Britain, there is no evidence in Ireland for Roman or Anglo-Saxon settlement. The earliest historical evidence for foreign introgression is of Viking presence from the 8th century AD. The first recorded Viking attack on Ireland took place in 795 AD, with attacks becoming more intensive after 837 AD when much larger fleets sailed to the Liffey, Boyne and Shannon. However, by contrast with Britain, over half of which was under the control of the Vikings by the end of the 9th century, permanent Viking settlement in Ireland was confined to coastal areas and proved to be enduring only along the east and south coasts in Dublin, Waterford, Wexford and Limerick. By the mid 10th Century the Vikings had settled permanently and peacefully in Ireland and had converted to Christianity (Flanagan 1988, Davies 1999).

2.1.8 The Norman Invasion

The Normans reached Ireland *via* England and Wales, *circa* 1170 AD. The Norman invasion was a private enterprise, the settlements were established by freebooting adventurers whose activities were uncoordinated and uncontrolled. Their settlement pattern was fragmented, with dense colonization at the points of disembarkation in Dublin and the southeast, from where they radiated out along the river valleys and low ground to the south, west and northeast. The pattern that emerged was made up of a colonial heartland in Leinster, which became known as the Pale and maintained close ties with Britain, connected with numerous and extensive clusters of outlying settlements scattered amongst the native Irish (Warren 1988).

2.1.9 The Plantations

The political consequences of the Reformation in Europe were largely responsible for the next phase of foreign introgression in Ireland. Colonists who had been in Ireland since the Anglo-Norman invasion largely failed to adopt the Protestant religion. The Crown's policy in Ireland was ultimately governed by international pressure and in a Europe divided by religion the first line of English defence was control over Ireland; control preferably in Protestant hands. For this reason, from Tudor times, where and when the opportunity arose, migration to Ireland was encouraged and plantation (the systematized uprooting of the native population in favour of incoming colonists) of selected areas was undertaken. The first major plantation scheme occurred in the midland counties of Leix (Laois) and Offaly during the 1550s. A second scheme occurred in Munster in the 1580s. However these plantations lacked sufficient

numbers of English tenant farmers and artisans to ensure lasting stability (Clarke 1988, Davies 1999).

2.1.10 The Plantation of Ulster

This was followed by the plantation of Antrim and Down in the north by lowland Scots. After the 'Flight of the Earls' in 1607 the Crown embarked on the official plantation of Ulster. The O'Neill, Earl of Tyrone and the O'Donnell, Earl of Tyrconnell (Donegal) fled with about 100 lesser chieftains to Europe after their campaign to resist English authority in Ulster failed. Their lands were declared forfeit and a vast enterprise of colonisation and redistribution of the land in mid and west Ulster was initiated. The Ulster plantation attracted predominantly Anglican and Presbyterian English, Welsh and Lowland Scots, in sufficient numbers to ensure a lasting settlement, and whose descendants comprise a substantial portion of the Ulster population to this day (Davies 1999).

2.2 Materials and Methods

2.2.1 Sample Collection

846 DNA samples from males living in Ireland were genotyped. Samples were taken by buccal swab with appropriate informed consent. County of birth of individual, father and paternal grandfather were recorded, as well as surname in some cases. DNA was extracted by Dr. Emmiline Hill using a standard phenol-chloroform extraction protocol. Samples were collected at various locations in Ireland: Trinity College Dublin, University College Cork, University of Limerick, RTC Carlow, RTC Tallaght, RTC Waterford, RTC Dundalk, RTC Tralee, RTC Letterkenny, St. Patrick's College, Maynooth, Bonnicionollan, Co. Mayo, St. Muirideach's College, Co. Mayo, The King's Hospital School, Co. Dublin.

Samples were assigned geographic origin based on county of birth of paternal grandfather, or father or individual if this information was unknown. All 32 counties in the island of Ireland were represented in the sample but representation from the counties of east Ulster (Antrim, Down, Armagh) was poor. Therefore an additional 35 samples from males with surnames and geographic origin in these counties were provided by Brian McEvoy to supplement analysis. Samples with geographic origin outside the island of Ireland were not included in the analysis, leaving 796 Irish Y chromosomes for analysis (762 not including East Ulster additional samples).

2.2.2 Single Nucleotide Polymorphism Analysis

The DNA samples were assayed for 11 single nucleotide polymorphisms (SNPs) found commonly in European populations (Semino *et al.* 2000, Rosser *et al.* 2000, Y Chromosome Consortium 2002). SNPs were genotyped in a hierarchical order outlined in Figure 2.2.2.1 and were assigned to a haplogroup accordingly. The phylogenetic relationship of the SNPs typed is shown in Appendix Section 7.

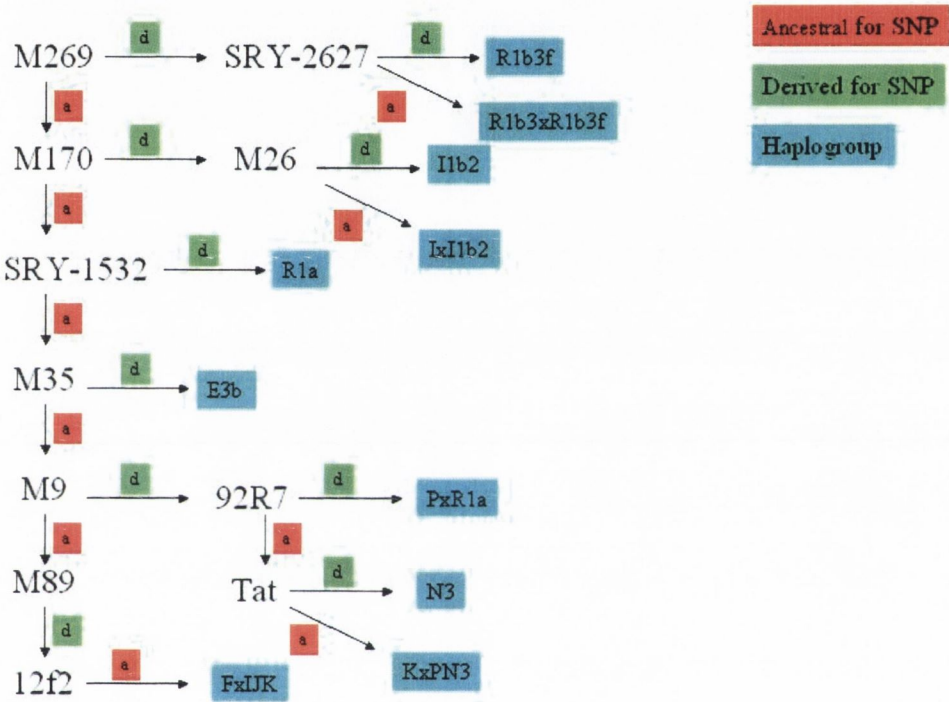


Figure 2.2.2.1: Hierarchical SNP genotyping. All samples were first assayed for the M269 SNP, which defines Hg R1b3. Samples that possessed the derived allele for M269 were then assayed for the SNP SRY-2627. If they possessed the derived allele for SRY-2627 they were assigned to Hg R1b3f, if they possessed the derived allele for M269 but the ancestral allele for SRY-2627 they were assigned to Hg R1b3xR1b3f. If samples possessed the ancestral allele for M269 they were assayed for the SNP M170, and so on, according to the scheme above, until all samples were defined by one of the Hgs indicated.

2.2.3 SNP Genotyping Protocol

SNPs were genotyped using two methods. M269, SRY-2627, SRY-1532, M35, M9 and tat were genotyped by RFLP analysis. M170, M26 and M89 were genotyped

using Tetra-primer ARMS-PCR (Ye *et al.* 2001). This method uses two primer pairs to amplify, respectively, the two different alleles of a SNP in a single PCR reaction. 12f2 is an insertion/deletion polymorphism and was analysed by sizing the amplicons on an agarose gel.

For those SNPs genotyped by RFLP analysis, regions around the SNP were amplified according to the following protocol:

RFLP PCR Mastermix Composition

A 10 μ l final volume PCR reaction mix contained:

1X PCR reaction buffer (500mM KCl, 100mM TrisCl (pH 9.0), 25 mM MgCL₂,

0.1% gelatin, 1.0% Triton X-100)

0.2 μ M forward primer

0.2 μ M reverse primer

200 μ M dNTPs

0.15 μ l enzyme

3.95 μ l dH₂O

2 μ l DNA (approx. concentration)

PCR cycling conditions

33 cycles of:

94°C 30sec

Annealing 30sec

72°C 30 sec

Primers and annealing temperatures are given in Table 2.2.3.1. Amplicons were digested with the appropriate restriction enzyme. The resulting pattern of restriction fragments was used to assign a genotype to each sample. Restriction enzymes, digestion temperatures and restriction fragment sizes are given in Table 2.2.3.1

The basic composition of a 10µl restriction digest reaction was as follows:

5 units restriction enzyme

1X appropriate restriction buffer (see Table 2.2.3.1)

0.1 µl BSA (100x), where appropriate

ddH₂O to bring final volume to 5µl

5µl PCR product

Samples were digested for 1 hour at the temperature given in Table 2.2.3.1.

For Tetra-primer ARMS-PCR a 10µl PCR reaction volume contained:

1X PCR buffer (as above)

10 pmol each inner primer (forward and reverse)

1 pmol of each outer primer (forward and reverse)

200µM dNTPs

0.15 µl enzyme

1.85µl dH₂O.

PCR cycling conditions:

95°C 2min

35 cycles of:

95°C 1min

Annealing 1min

72 °C 1min

72 °C 1min extension

Primers, annealing temperatures and amplicon lengths are given in Table 2.2.3.1.

All products were sized by electrophoresis on a 1.5% agarose gel alongside a 100bp ladder, visualised by Ethidium Bromide staining and photographed under UV light.

2.2.4 Microsatellite genotyping

All samples were typed for 16 Y-linked microsatellites: DYS19, DYS388, DYS390, DYS391, DYS392, DYS393, DYS434, DYS435, DYS436, DYS437, DYS438, DYS439, DYS389I, DYS389II, YA7.1 (DYS460), YA7.2 (DYS461), GO4911 (DYS462) and for amelogenin, a PCR-based sex-typing assay which differentiates a 6bp deletion in the X-linked copy with respect to the Y-linked copy. Microsatellites were amplified in three multiplex reactions, MS1, CTS and EBF, modified from the method of Bosch *et al.* (2002). The CTS multiplex (Ayub *et al.* 2000) amplifies DYS434, DYS435, DYS436, DYS437, DYS438, DYS439. The MS1 multiplex

(Thomas *et al.* 1999) amplifies DYS19, DYS388, DYS390 (Kayser *et al.* 1997), DYS391, DYS392 (Thomas *et al.* 1999) and DYS393 (Kayser *et al.* 1997). The EBF multiplex (Bosch *et al.* 2002) amplifies DYS389I/II (Kayser *et al.* 1997), YA7.1, YA7.2 (White *et al.* 1999), GO4911 (de Knijff, unpublished, see Bosch *et al.* 2002) and the Amelogenin sex test assay (Sullivan *et al.* 1993).

For each multiplex a 10 μ l final volume PCR reaction mix contained:

1X PCR buffer (as before)

2.5mM MgCl₂

1X primer mix

250 μ M dNTPs

0.15 μ l enzyme

3.85 μ l dH₂O

Table 2.2.3.1 SNP Information

SNP	Reference	Alleles	Primers, 5' – 3'	Annealing temp.	MgCl ₂ Mm	Enzyme	Digestion temp.	Product (BP)	Hg
M269	Cruciani <i>et al.</i> 2002	T>C	F:catgcctagcctcattcctc R:ggctgacgtagtcttgcctc	60°C	25	<i>BstNI</i>	60 °C	247, 169 T 94, 153, 169 C	R1b3
SRY-2627	Veitia <i>et al.</i> 1997	C>T	F:gaactcctttactggggtg R:gtgacaaccaagaagagga	52 °C	15	<i>BanI</i>	65 °C	111, 87, 115 C 111, 202 T	R1b3f
SRY-1532	Whitfield <i>et al.</i> 1995	A>G	F:ttcttagcaaccattaatctgg R:aaatagcaaaaaatgacacaaggc	59 °C	20	<i>DraIII</i>	37 °C	167 A 112, 55 G	R1a
M35	Underhill <i>et al.</i> 2001	G>C	F:taagcctaaagagcagtcagag R:agaggaggagcaatgaggaca	65 °C	25	<i>BsrI</i>	65 °C	345, 168 G 513 C	E3b
M9	Underhill <i>et al.</i> 1997	C>G	F:gcagcatataaaaactttcagg R:aaaacctaactttgctcaagc	58 °C	20	<i>HinfI</i>	37 °C	182, 93, 66 C 248, 93G	K
92R7	Mathias <i>et al.</i> 1994	C>T	F:gacccgctgtagacctgact R:gcctatctacttcagtgattct	62 °C	15	<i>HindIII</i>	37 °C	709 C 197, 512 T	P
Tat	Zerjal <i>et al.</i> 1997	T>C	F:gactctgagtgtagacttgta R:gaagggtgccgtaaaagtgtgaa	60 °C	20	<i>NlaIII</i>	37 °C	112 T 85, 27 C	N3
12f2	Rosser <i>et al.</i> 2000	present>absent	F:ctgactgatcaaaatgcttacagatc R:ggatcccttcttacacctatac	60 °C -	-	-	-	88bp present	J

Table 2.2.3.2 Tetra-primer ARMS PCR Information

SNP	Reference	Alleles	Primers, 5' – 3'	Annealing temp.	MgCl ₂ Mm	Product (BP)	Hg
M170	Underhill <i>et al.</i> 2001	A>C	Inner F:ctattttatttacttaaaaatcattgtccc Inner R:gagacacaacccacactgaaaaagat Outer F:ttattaactgtattacagggttgctagt Outer R:agtgatgttgactgagaaactcatagc	53 °C	25	446 large 290 A 212 C	I
M26	Underhill <i>et al.</i> 2001	G>A	Inner F:ttcaattttttctgaattagaattagcg Inner R:ttcataggccattcagtgttctcggt Outer F:atgctgtgtatgtttacaacgtgaactg Outer R:catacacaaggatgcagcactaagtttt	61 °C	25	318 large 170 G 203 A	I1b2
M89	Underhill <i>et al.</i> 2001	C>T	Inner F:tcttctcctaagggtatgtacaaaaatc Inner R:gctgcaactcaggcaaagtgcgagcta Outer F:ccttcccttttctcctaagaactctgc Outer R:agctgtgtgaagtcttggcagaatag	60 °C	25	305 large 152C 210 T	F

Table 2.2.4.1 Microsatellite Information

Locus	repeat	repeat no.	repeat size (BP)	final conc. uM	5' dye	Primer 5' – 3'
CTS multiplex:						
DYS434	ctat	8-11	110-122	0.2	tet	F:cac tcc ctg agt gct gga tt R:gga gat gaa tga atg gat gga
DYS435	tgga	11-13	220-228	0.5	tet	F:agc atc tcc aca cag cac ac R:ttc tct ctc ccc ctc ctc tc
DYS436	gtt	10-15	128-143	0.1	fam	F:cca gga gag cac aca caa aa R:gca atc caa ctt cag cca at
DYS437	tcata	8,2,4-12,2,4	186-202	0.1	hex	F:gac tat ggg cgt gag tgc at R:aga ccc tg cat tca cag atg a
DYS438	ttttc	6-12	203-233	0.25	hex	F:tgga gga ata gtt gaa cgg taa R:gtg gca gac gcc tat aat cc
DYS439	gata	9-14	238-258	0.2	tet	F:tcc tga atg gta ctt cct agg ttt R:gcc tgg ctt gga att ctt tt
MS1 multiplex:						
DYS19	ctat/c	10-19	173-209	0.35	tet	F:act gag ttt atg tta tag tgt ttt t R:atg gca tgt agt gag gac a
DYS388	ata	10-16	123-141	0.25	tet	F:gtg agt tag ccg ttt agc ga R:cag atc gca acc act gcg
DYS390	ctg/at	18-27	192-228	0.2	fam	F:tat att tta cac att ttt ggg cc R:tga cag taa aat gaa cac att gc
DYS391	ctat	8-13	156-176	0.23	fam	F:cta ttc att caa tca tac acc catat R:aca tag cca aat atc tcc agg g
DYS392	att	7-16	155-182	0.35	hex	F:cag tca aag tgg aaa gta gtctgg R:gtg gtc ttc tac ttg tgt caa tac
DYS393	gata	9-15	107-131	0.15	hex	F:gtg gtc ttc tac ttg tgt caa tac R:aac tca agt cca aaa aat gag g

Table 2.2.4.1 Microsatellite Information continued

Locus	repeat	repeat no.	repeat size (BP)	final conc. uM	5' dye	Primer 5' – 3'
EBF multiplex:						
DYS389I/II	ctg/at	I:10-16 II:26-34	142-166 258-300	0.35	fam	F:cca act ctc atc tgt att atc tat gt R:cct gag tag cag aag aat gtcata
YA7.1	gata	9-13	170-186	0.35	tet	F:gcc aaa ctc ttt cca aga ag R:tca tct atc ctctgc cta tca tt
YA7.2	gata	8-12	174-190	0.25	fam	F:agg cag agg ata gat gat atg gat R:ttc agg taa atc tgt cca gta gtg a
GO4911	tgta	10-14	179-199	0.3	hex	F:tgt gct gta cca gtt gcc ta R:cca gcc tga gca aga gag ta
AMEL	-	-	X:106 Y:112	0.15	hex	F:ccc tgg gcct ctg taa aga ata gtg R:act aga gct taa act ggg aa

Cycling conditions for each multiplex were as follows:

MS1

94 °C 2min
6 cycles of:
 94 °C 1min
 57.5 °C 1min (decreasing by 0.5 °C each cycle)
 72 °C 1min
33 cycles of
 94 °C 1min
 55 °C 1min
 60 °C 1min
final step 72 °C 10min

CTS

94 °C 2min
9 cycles of:
 94 °C 1min
 60 °C 1min (decreasing by 0.5 °C each cycle)
 72 °C 1min
31 cycles of:
 94 °C 1min
 56 °C 1min
 72 °C 1min
final step 72 °C 5min

EBF

95 °C 5min
6 cycles of:
 94 °C 1min
 60 °C 1min (decreasing by 0.5 °C each cycle)
 72 °C 1min
29 cycles of:
 94 °C 1min
 55 °C 1min
 72 °C 1min
final step 72 °C 10min

Details of primers, repeat types and sizes of microsatellite loci and alleles are given in Table 2.2.4.1.

Multiplex PCR products were electrophoresed on an 8% PAGE gel along with Rox-500 size standard (GeneScan-500 ROX size standard, Applied Biosystems) on an ABI377 DNA sequencer for ~3.5 hours. The microsatellite data was collected and analysed using GeneScan software (Perkin Elmer).

2.2.5 Analysis

Compound haplotypes were constructed for each sample using assigned haplogroup together with allele repeat number for each of the 17 microsatellite loci (presented in Appendix 2.1). Frequencies of each haplotype were calculated using ARLEQUIN 2.0 (Schneider *et al.* 2000). Overall frequencies of each haplotype in the Irish sample were calculated, as well as regional frequencies, obtained by splitting the Irish sample into 9 groups based on geographic origin as follows:

Region	Grandfather's County of Origin	N
Northwest	Derry, Donegal, Tyrone, Fermanagh	65
Northeast	Antrim, Down, Armagh	60
North Connaught	Sligo, Mayo, Leitrim, Roscommon	118
Midlands North	Cavan, Longford, Westmeath, Monaghan, Louth, Meath	103
Dublin	Dublin	108
Midlands South	Offaly, Laois, Kildare, Wicklow	74
CGT	Clare, Galway, Tipperary	75
Southeast	Wexford, Carlow, Kilkenny, Waterford	84
Southwest	Cork, Kerry, Limerick	109

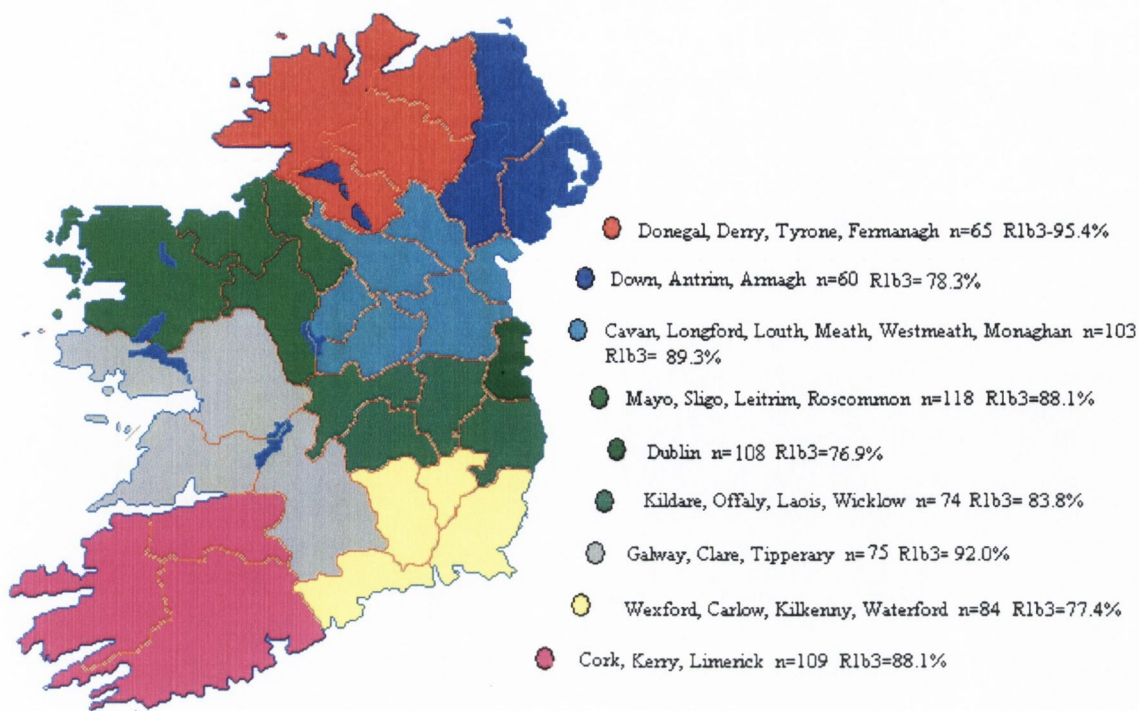


Figure 2.2.5.1: Map of Ireland
Irish Y samples were split into 9 regional sub-populations indicated above, based on grandparental geographic origin.

2.2.6 Fst pairwise population genetic distance analysis

The relationship of the 9 Irish subpopulations to one another were analysed using Arlequin 2.0 (Schneider *et al.* 2000). Haplotypes were converted to binary format as follows: Microsatellite data for each sample were converted to binary format using the program Factor implemented in the Phylip package (Felsenstein 1989). Then SNP data for each sample was included by assigning a binary code to each haplogroup. Fst pairwise population genetic distances were calculated between each pair of populations.

2.2.7 Principal Components Analysis

Principal Components Analysis of the data was done using the program POPSTR written by Henry Harpending. In order to examine the relationship of the Irish subpopulations to one another and to compare them with subpopulations from other regions of Europe, data from the 9 Irish subpopulations were added to the data of Capelli *et al.* (2003). The data were classified according to the method of Capelli *et al.* (2003), who analysed 17 populations from the British mainland (England, Scotland, Wales), samples from Orkney, Shetland, the Western Isles, the Isle of Man, the Channel Islands, two samples from Ireland (Rush on the east coast and Castlerea in the midlands), and samples from Norway, Germany and Denmark. These samples were typed for SNPs which define haplogroups largely analogous to those typed in this study.

Their samples were typed for 6 microsatellites, DYS19, 388, 390, 391, 392, 393, which were also typed in the Irish sample. When performing PC analysis, Capelli *et al.* defined their populations according to haplogroup, but samples belonging to haplogroups RxR1a1, IxI1b2 or RxR1a1 were split into 2 categories, based on whether they possessed the AMH+1, 2.47+1 or 3.65+1 haplotypes. The modal haplotype for IxI1b2 (2.47) was 14-14-22-10-11-13 and the modal haplotype for R1a1 was 16-12-25-11-11-13. POPSTR was used to calculate the first, second and third Principal Components of the dataset and these were plotted using Microsoft Excel. The impact of each haplogroup on the Principal Components were also plotted.

2.2.8 Multidimensional Scaling Analysis

The relationships of the Irish data and the data of Capelli *et al.* were also analysed using multidimensional scaling analysis. Data were converted to binary format as above. An Fst pairwise population distance matrix was created using Arlequin 2.0 (Schneider 2000). This matrix was summarised using the multidimensional scaling option in SPSS and plotted using Microsoft Excel.

2.2.9 Median-joining Networks

Median-joining networks of R1b3 samples in the 9 regional subpopulations were constructed using Network 4.0 (www.fluxus-engineering.com). Haplotypes were first analysed using the reduced median option and the *.rmf output file was used as an input file for the median-joining algorithm (Bandelt *et al.* 1999). Relative weights were applied to the loci on a 10-fold scale on the basis of the variance at each locus in order to minimize the number of reticulations in the network. Higher weights were given to the less variable loci (Hurles *et al.* 2002). The maximum parsimony post-processing option was used to display the network containing all the shortest trees

2.2.10 Dating Estimates

Hill *et al.* (2000) have argued that R1b3 is the earlier indigenous Y chromosome Hg in Ireland and that all non-R1b3 Irish Y lineages represent later genetic introgression. BATWING (Wilson and Balding 2002) was used to estimate TMRCA and date expansions in the Irish R1b3 population. BATWING reads multi-locus haplotype data along with model and prior distribution specifications and uses a

Markov Chain Monte Carlo method based on coalescent theory to generate approximate random samples from the posterior distributions of parameters.

10,000 runs of the program were performed and the first 1,000 were discarded as burn-in. The model used was exponential expansion from a single constant-sized ancestral population. Each locus was allowed to have a mutation rate independent of others. TMRCA and expansion time output were given in coalescent units and were converted to time in generations by multiplying by N , the effective population size. This was then converted to calendar years through multiplication by generation time, taken to be 25 years.

Prior distribution specifications were as follows:

Mutation rate: μ_{prior} : normal(0.00069,0.00057)

Expansion rate: α_{prior} : normal(0.01,0.01)

Ancestral effective population size: N_{prior} : gamma (1,0.001)

Expansion time: β_{prior} : uniform(0,10).

The prior distribution for the mutation rate was taken from Zhivotovsky *et al.* 2004, who estimated a Y microsatellite mutation rate of 6.9×10^{-4} per 25 years, with a standard deviation of 5.7×10^{-4} . The BATWING prior was based directly on this, being of the form normal($V1$, $V2$) where $V1$ = mean and $V2$ = standard deviation. The prior for ancestral effective population size (N_{prior}) was of the form gamma($V1$, $V2$) where mode = $(V1 - V2)/2$ and mean = $V1/V2$. The prior values were based on

example datasets for this program and were chosen to aid convergence of the posterior distribution.

Median, mode, 2.5% and 97.5% Credible Intervals of the posterior distribution for parameters were plotted and calculated using Microsoft Excel.

2.3 Results

2.3.1 Haplogroup Frequencies

All samples were categorised into haplogroups defined by the 11 SNPs

Haplogroup	No. Samples	Frequency
R1b3xR1b3f	680	0.854
IxI1b2	87	0.109
I1b2	8	0.010
E3b	7	0.009
KxPN3	3	0.004
Px1R1aR1b3	4	0.005
R1a	3	0.004
FxIJK	2	0.003
N3	1	0.001
R1b3f	1	0.001

Total 796

Table 2.3.1.1: Haplogroup frequencies in the Irish Sample

The most common haplogroup was R1b3xR1b3f, occurring at an overall frequency of 85.4%. The only other haplogroup occurring at substantial frequency was IxI1b2 at 10.9%. In order to examine regional haplogroup frequencies the Irish sample was split into 9 geographic regions as indicated previously.

Hg	W. Ulster	E. Ulster	North Conn	Midlands N.	Midlands S.	Dublin	CGT	S. West	S. East
R1b3xR1b3f	0.954	0.783	0.881	0.893	0.838	0.769	0.920	0.881	0.774
IxI1b2	0.03	0.20	0.085	0.087	0.135	0.176	0.053	0.083	0.143
I1b2					0.014	0.009	0.013	0.009	0.024
E3b						0.019		0.018	0.036
Kx PN3		0.015	0.008			0.009		0.009	
Px1R1aR1b3			0.008	0.010		0.009	0.013		
R1a					0.014	0.019			
FxIJK	0.017								0.012
N3									0.012
R1b3f				0.010					
Total	65	60	118	103	74	108	75	109	84

Table 2.3.1.2: Haplogroup frequencies in 9 geographic regions of Ireland

The frequency of R1b3xR1b3f varied from region to region. Broadly speaking its frequency rose from southeast to northwest. It peaked in the Northwest at ~95% and lowest frequencies occurred in Dublin and the Southeast (~76-77%). IxI1b2 was the only other haplogroup occurring at significant frequency. In a contrasting pattern to R1b3xR1b3f, it peaked in the east and southeast and occurred at lower frequencies in the Northwest. It reached frequencies of 17.4% in Dublin, 14.7% in the Southern Midlands, 14.3% in the Southeast and 20% in the Northeast. It accounted for ~3% of the Northwest samples. The other 8 haplogroups were found throughout the 9 regions of the country at frequencies too low to ascribe a geographic pattern. There is a regional contrast between the frequency of R1b3xR1b3f versus non- R1b3xR1b3f chromosomes in the Irish sample, with northwestern populations possessing a higher number of the former and the southern and eastern populations possessing a higher number of the latter.

2.3.2 Microsatellite Haplotypes

640 17-microsatellite haplotypes were observed overall in the Irish population. These are listed in Appendix Section 2. Of these 583 were singletons. Of the 57 haplotypes that occurred more than once, the most common occurred 30 times. This haplotype contained the alleles 14-12-25-11-14-13-9-11-12-15-12-12-13-16-11-10-11 at the loci DYS19-DYS388-DYS390-DYS391-DYS392-DYS393-DYS434-DYS435-DYS436-DYS437-DYS438-DYS439-DYS389I-DYS389b-YA7.1-YA7.2-GO4911. This accounts for 3.8% of the total Irish sample (8.2% together with its one-step neighbours) and is henceforth referred to as the Irish Modal Haplotype (IMH).

The second most common haplotype, 14-12-24-11-13-13-9-11-12-15-12-12-13-16-11-10-11, occurred 18 times in the Irish sample. This haplotype is equivalent to the Atlantic Modal Haplotype (AMH, Wilson *et al.* 2001b) at the loci DYS19-DYS388-DYS390-DYS391-DYS392-DYS393, as is the third most common haplotype, 14-12-24-11-13-13-9-11-12-15-12-11-13-16-11-10-11, which occurred 12 times. When these 6 loci only are considered, the Atlantic Modal Haplotype, 14-12-24-11-13-13, is the most common in the Irish sample. Together with its one-step neighbours it occurs 331 times and accounts for 41.6% of the Irish sample. Within a 17-locus haplotype system AMH broke down into 94 distinct haplotypes, 80 of which were singletons.

2.3.3 Regional haplotype frequencies

The regional frequencies of IMH and its one-step neighbours are given below.

Popn.	IMH+1 freq
Northwest (n=65)	0.246
North Connaught (n=118)	0.144
Midlands South (n=75)	0.120
Northeast (n=60)	0.083
Southeast (n=84)	0.060
Southwest (n=109)	0.046
Midlands North (n=103)	0.039
Dublin (n=109)	0.037
CGT (n=76)	0.013

IMH +1 frequency in the R1b3 population of Ireland was plotted as a contour map using ArcView. The Irish sample was split into populations as follows:

Population	Sample Size	IMH+1 freq	Degrees North	Degrees West
Antrim Down Armagh	47	0.106	54.55	-6.25
Derry Donegal Tyrone	59	0.237	54.8	-7.5
Cavan Monaghan	25	0.143	54.2	-7.25
Leitrim Sligo	25	0.160	54.3	-8.1
Louth	31		54	-6.5
Longford Roscommon Westmeath	24	0.083	53.6	-8
Mayo	70	0.186	54.1	-9
Galway	26	0.040	53.5	-8.75
Kildare Meath	36	0.083	53.3	-7
Dublin	83	0.060	53.4	-6.25
Laois Offaly	22	0.091	53.25	-7.5
Wicklow	22	0.182	53	-6.25
Wexford	27	0.000	52.25	-6.5
Carlow Kilkenny Waterford	38	0.079	52.5	-7.25
Tipperary	29		52.55	-7.8
Clare Limerick	34	0.118	52.55	-8.5
Kerry	30		52.1	-9.45
Cork	47	0.021	52	-8.5

The resulting contour map (Figure 2.3.3.1) shows a striking cline of IMH+1 frequency on the island, which peaks in the northwest, where it reaches frequency of 34.2% in Derry.

2.3.4 Fst Pairwise Population Genetic Distances

	nw	ne	nc	mn	cgt	ms	dub	se
nw	0							
ne	0.00616	0						
nc	0.00112	0.0013	0					
mn	0.01323	-0.00093	0.00457	0				
cgt	0.01443	0.00121	0.00647	0.00119	0			
ms	0.00914	-0.00081	0.00259	0.00078	0.00149	0		
dub	0.01153	-0.00033	0.00402	0.0001	0.0001	0.00069	0	
se	0.01428	0.00065	0.00615	0.00083	-0.00059	0.00181	-0.00003	0
sw	0.01136	0.00002	0.00431	0.00043	0.00148	-0.00003	0.00052	0.00115

Table 2.3.4.1: Matrix of Fst Pairwise population genetic distances

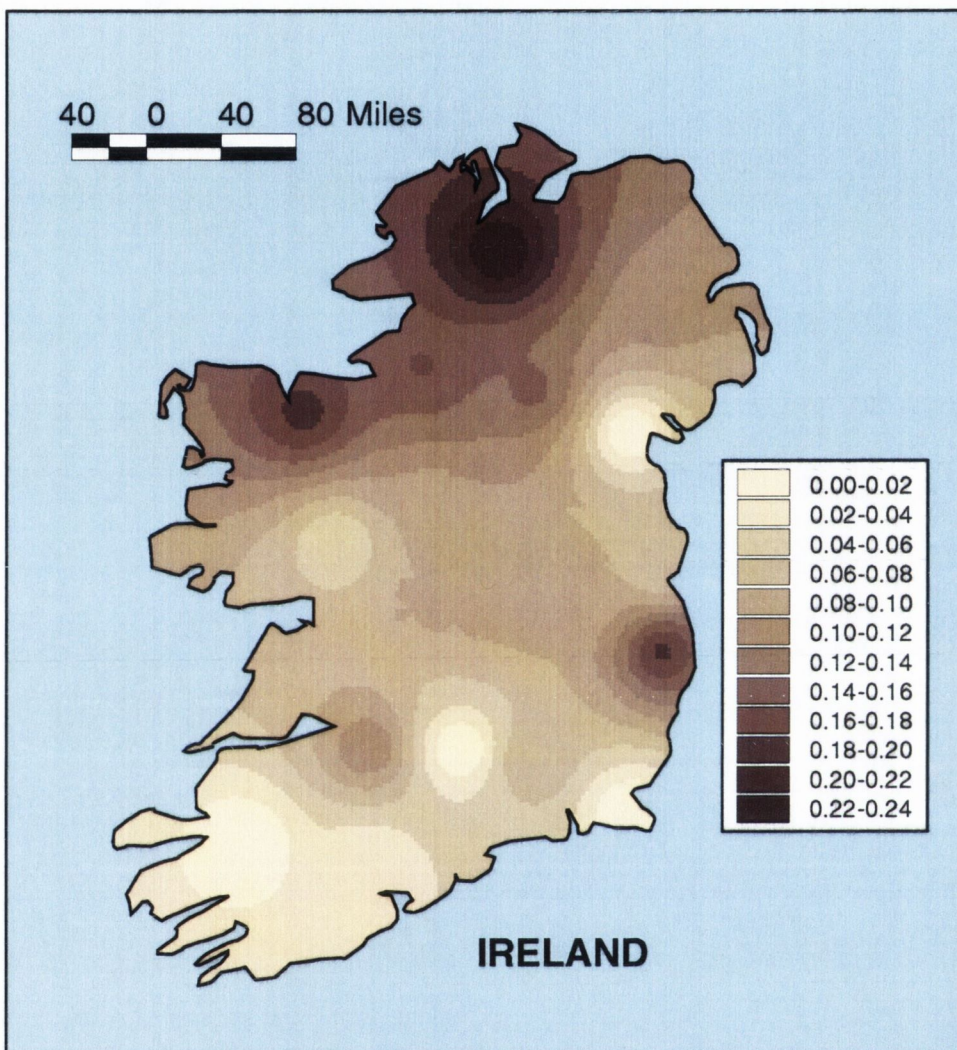


Figure 2.3.3.1; Contour Map of IMH+1 frequency in Ireland

	nw	ne	nc	mn	cgt	ms	dub	se
nw								
ne	+							
nc	-	-						
mn	+	-	+					
cgt	+	-	+	+				
ms	+	-	+	-	-			
dub	+	-	+	-	-	-		
se	+	-	+	-	-	+	-	
sw	+	-	+	-	+	-	-	-

Table 2.3.4.2: Matrix of significant Fst P-values

Table 2.3.4.1 shows Fst pairwise population genetic distances among the Irish regional sub-populations. The six largest genetic distances are between the Northwest population and others. The Northwest is most distant from CGT, then the Southeast, Midlands North, Dublin, Southwest, Midlands South and Northeast populations. It is significantly different from all populations except North Connaught. The North Connaught population is also significantly different from many of the populations, namely Midlands North, Dublin, CGT, Southwest and Southeast. It is not significantly different from Northeast or Northwest. Southwest and CGT are significantly different from one another, as are Midlands North and CGT and Midlands South and Southeast. The remaining populations do not show significant genetic distance from one another.

2.3.5 Principal Components Analysis

The first three principal components explain 82.6% of the variance, as indicated in Table 2.3.5.1. A plot of PC1 vs. PC2 for the Irish and Capelli data is shown in Figure 2.3.5.1. and a plot of PC1 vs. PC3 is shown in Figure 2.3.5.3. The Irish populations

are marked in blue. PC1 explains ~54% of the variance, PC2 ~19% and PC3 ~9%. Table 2.3.5.2 indicates the impact of each haplogroup on the principal components, and the impact of each hg on the PC plots are shown in Figures 2.3.5.2 and 2.3.5.4.

Population	PC1 (54%)	PC2 (19%)	PC3 (9%)
Norway	-0.11	-0.256	0.0527
Isle of Man	-0.101	-0.0121	0.0213
Shetland	-0.0998	-0.0834	0.099
Penrith	-0.0919	-0.00336	-0.000168
Chippenham	-0.0906	0.00241	-0.0248
Germany/Denmark	-0.0781	-0.102	-0.118
Western Isles	-0.0765	-0.0231	0.008
Llangefni	-0.0756	0.09	0.0333
Llanidloes	-0.0731	0.02	-0.0554
Midhurst	-0.0677	0.059	-0.0129
Morpeth	-0.0654	0.034	-0.017
Cornwall	-0.0639	0.019	0.019
Orkney	-0.0627	-0.09	0.0707
Southwell	-0.0621	0.0244	-0.0499
Pitlochry	-0.0589	0.0596	0.0116
Oban	-0.0563	0.0506	0.043
York	-0.0547	-0.0267	-0.0723
Haverforwest	-0.0536	0.0847	0.0488
Faversham	-0.0501	0.0354	-0.0119
Dorchester	-0.0479	0.0341	-0.0298
Uttoxeter	-0.0457	0.0229	-0.0354
Norfolk	-0.0439	-0.0204	-0.0628
Basques	-0.0374	0.0907	0.0481
Channel Islands	-0.0312	0.0129	-0.0432
Stonehaven	-0.0204	0.0147	-0.000655
Castlerea	-0.00239	0.0658	0.0421
Rush	0.0326	0.0263	0.0476
Durness	0.0328	-0.0208	0.0143
Midlands S.	0.123	0.00288	0.00282
Southeast	0.131	-0.00927	-0.0344
Northeast	0.152	-0.0228	-0.0272
Dublin	0.158	-0.0371	-0.0253
Southwest	0.164	-0.000266	-0.00361
CGT	0.187	-0.00597	0.0241
Midlands N.	0.202	-0.0213	0.0082
N. Connaught	0.206	-0.0175	0.011
Northwest	0.229	0.00233	0.0189

Table 2.3.5.1: PC1, PC2 and PC3

HG	PC1	PC2	PC3
AMH+1	-0.323	0.184	0.0831
3.65+1	-0.123	-0.23	0.125
2.47+1	-0.111	-0.114	-0.124
R1a	-0.102	-0.117	0.00841
J2	-0.0702	0.0427	-0.0702
E3b	-0.0399	0.0314	-0.0838
FxIJK	-0.0281	0.00435	-0.0446
JxJ2	-0.0168	0.0187	-0.0198
N3	-0.0107	-0.0627	-0.0416
IxIb2	-0.00954	-0.0709	-0.1
PxR1aR1b3	0.00956	-0.109	0.0449
I1b2	0.0168	0.0434	0.0264
KxPN3	0.0426	0.0032	-0.00236
R1b3xR1b3f	0.495	0.0206	0.035

Table 2.3.5.2: The impact of each haplogroup on the PC plot

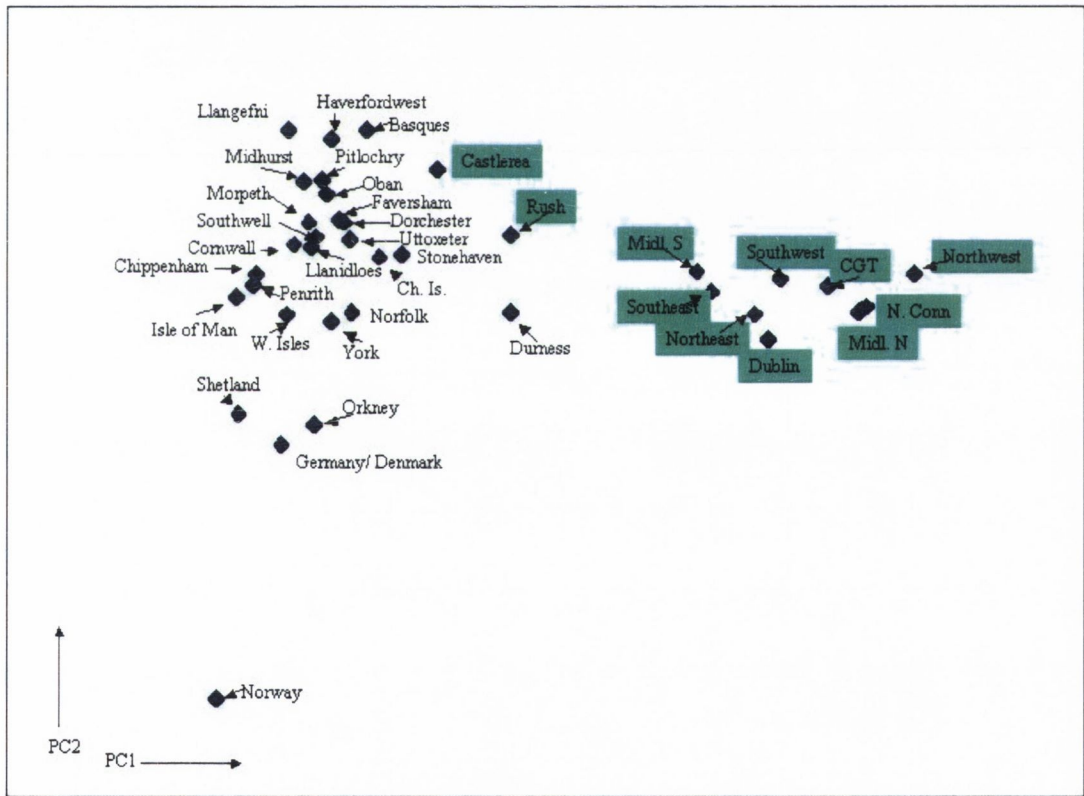


Figure 2.3.5.1: PC1 vs. PC2 of Irish and Capelli *et al.* (2003) data. Irish populations are marked in green

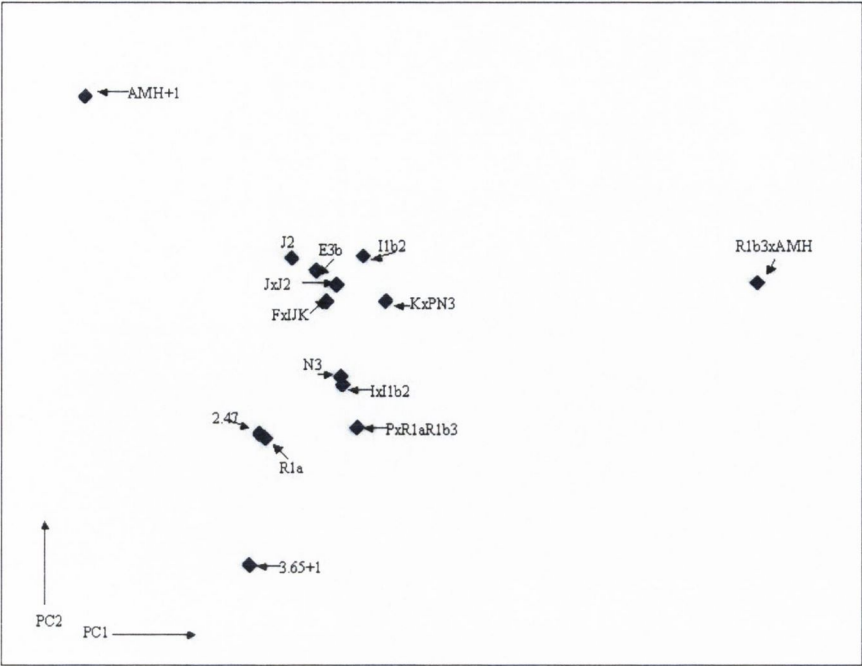


Figure 2.3.5.2: The impact of each Hg on the plot of PC1 vs. PC2.

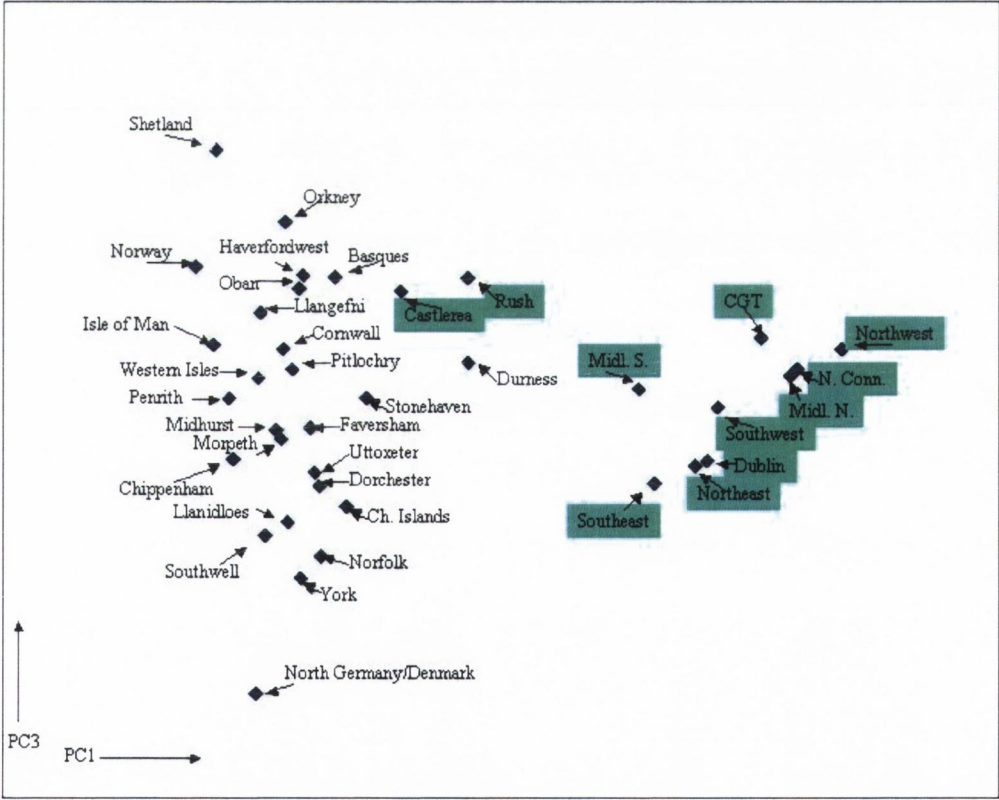


Figure 2.3.5.3: Plot of PC1 vs. PC3 of Irish and Capelli *et al.* (2003) data.

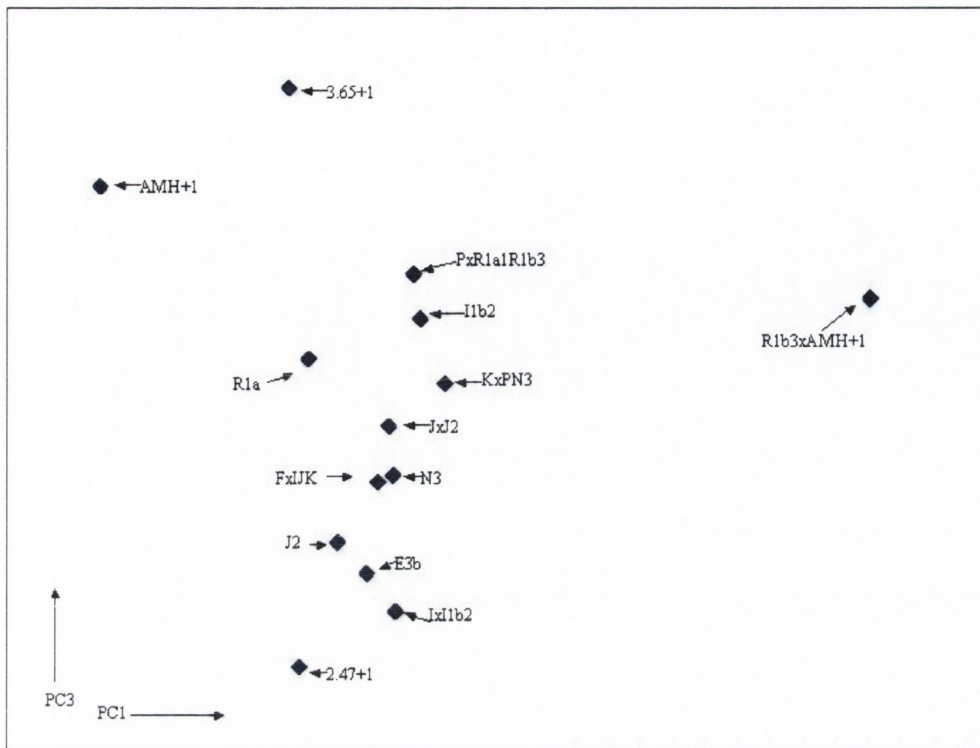


Figure 2.3.5.4: The impact of each Hg on the plot of PC1 vs.Pc3

In the plots of PC1 vs. PC2 and PC1 vs. PC3, the Irish populations fall to the right of the PC1 axis, relative to the other British and the continental populations. The plots of the impact of each Hg on the PC analysis indicate that this is due to their relatively high frequency of non-AMH R1b3 sequences. The populations from the north and west of Ireland occupy positions to the extreme right of the PC1 axis. The southern and eastern Irish populations fall closer to the British populations, consistent with the higher frequency of non-R1b3 chromosomes in these populations. This suggests that the southern and eastern Irish populations are closer genetically to the British populations than are the Irish populations of the north and west.

2.3.6 Multidimensional Scaling Analysis

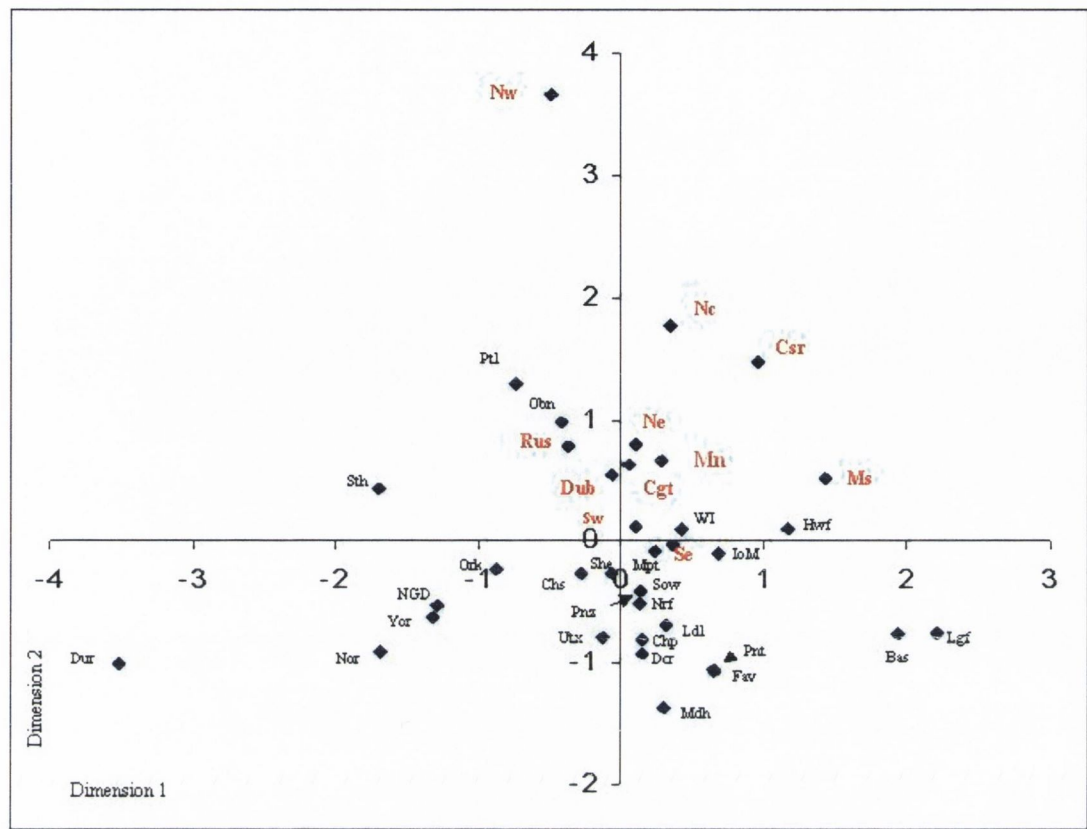


Figure 2.3.6.1 MDS plot of Irish and Capelli *et al.* 2003 populations. Nw = Northwest, Nc = North Connaught, Ms = Midlands South, Dub = Dublin, Sw = Southwest, Se = Southeast, Cgt = Clare, Galway, Tipperary, Ms = Midlands South, Ne = Northeast. For the populations of Capelli *et al.* (2003), Csr = Castlerea, Rus = Rush, Lgf = Llangefni, Bas = Basques, Ptl = Pitlochry, Obn = Oban, Hwf = Haverfordwest, WI = Western Isles, IoM = Isle of Man, Mpt = Morpeth, Pnt = Penrith, Fav = Faversham, Mdh = Midhurst, Pnz = Cornwall, Dcr = Dorchester, Chp = Chippenham, Utx = Uttoxeter, Ork = Orkney, Nrf = Norfolk, She = Shetland, Sow = Southwell, Ldl = Llanidloes, Chs = Channel Islands, Sth = Stonehaven, Dur = Durness, Yor = York, Nor = Norway, NGD = North Germany / Denmark. Irish populations are shown in red.

The MDS plot summarising the pairwise F_{st} genetic distances among the Irish and Capelli data (Figure 2.3.6.1) shows that the Irish populations cluster towards the top to the exclusion of other British Isles populations with the exception of Oban and Pitlochry. All Irish populations fall close together, apart from the Northwest population, which falls furthest to the top. Its closest neighbour is North Connaught. The other British Isles populations fall mostly below the Irish. Orkney and York fall

close to the continental European samples. Durness is an outlier to the bottom left. The Basque sample falls closer to the British than the Irish samples.

2.3.7 Network Analysis

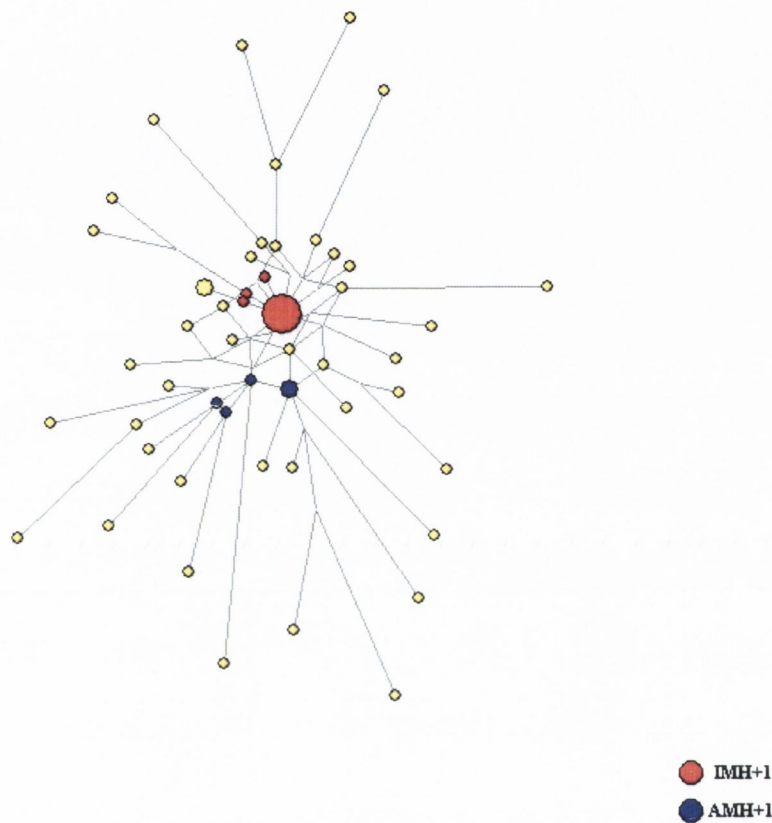


Figure 2.3.7.1: Median-Joining network of Northwest R1b3 samples. IMH+1 shown in red, AMH+1 shown in blue. Where a sample is one step from both IMH and AMH it is coloured in red.

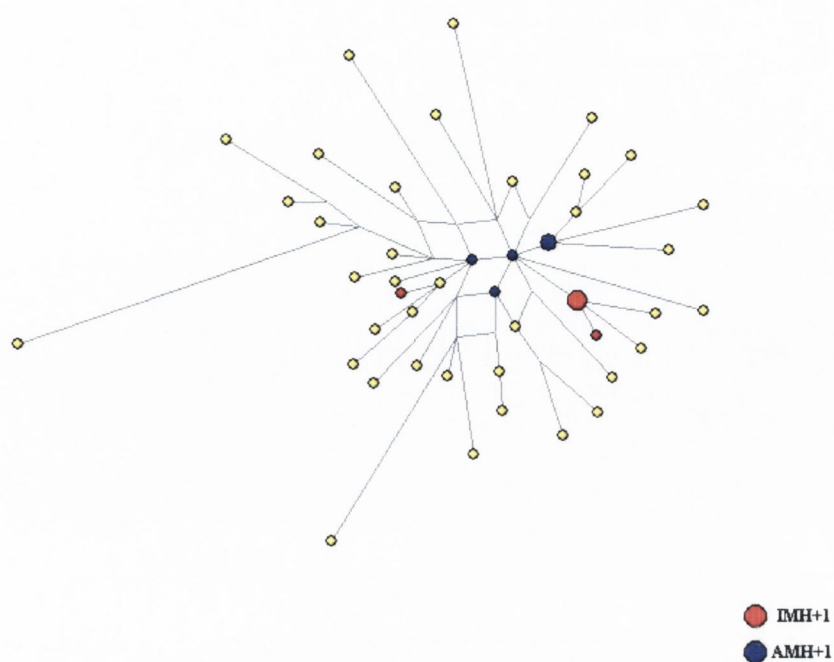


Figure 2.3.7.2: Median-Joining network of Northeast R1b3 samples. IMH+1 shown in red, AMH+1 shown in blue.

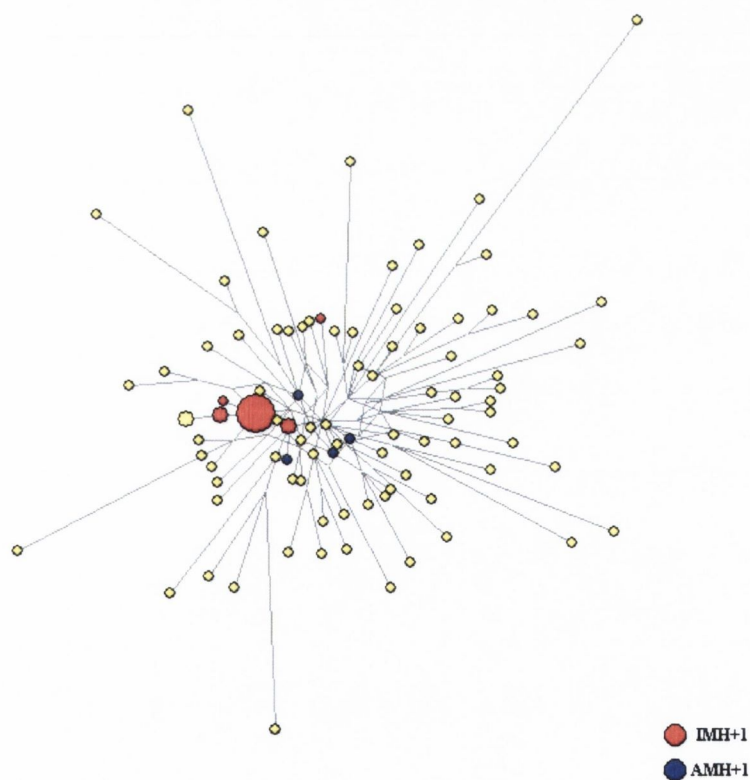


Figure 2.3.7.3: Median-Joining network of North Connaught R1b3 samples. IMH+1 shown in red, AMH+1 shown in blue.

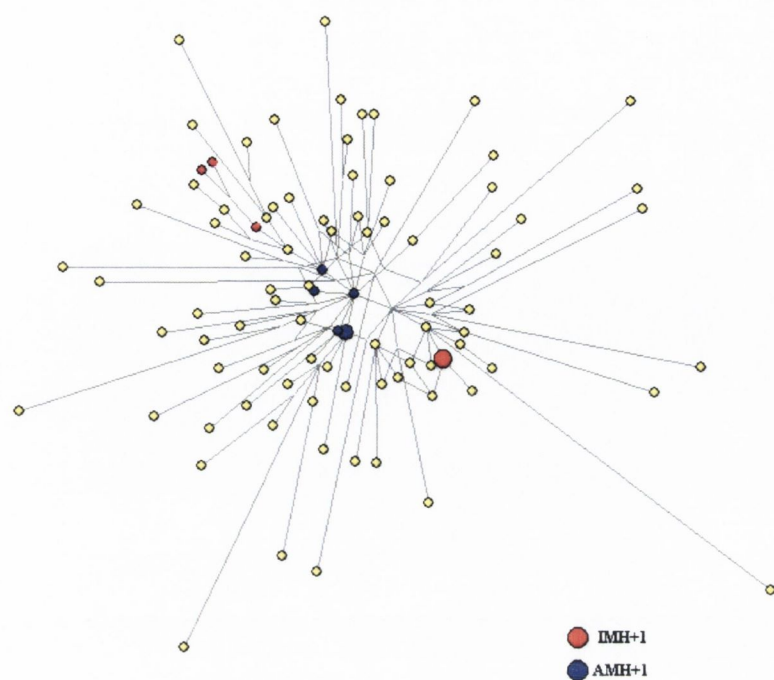


Figure 2.3.7.4: Median-Joining network of Midlands North R1b3 samples. IMH+1 shown in red, AMH+1 shown in blue.

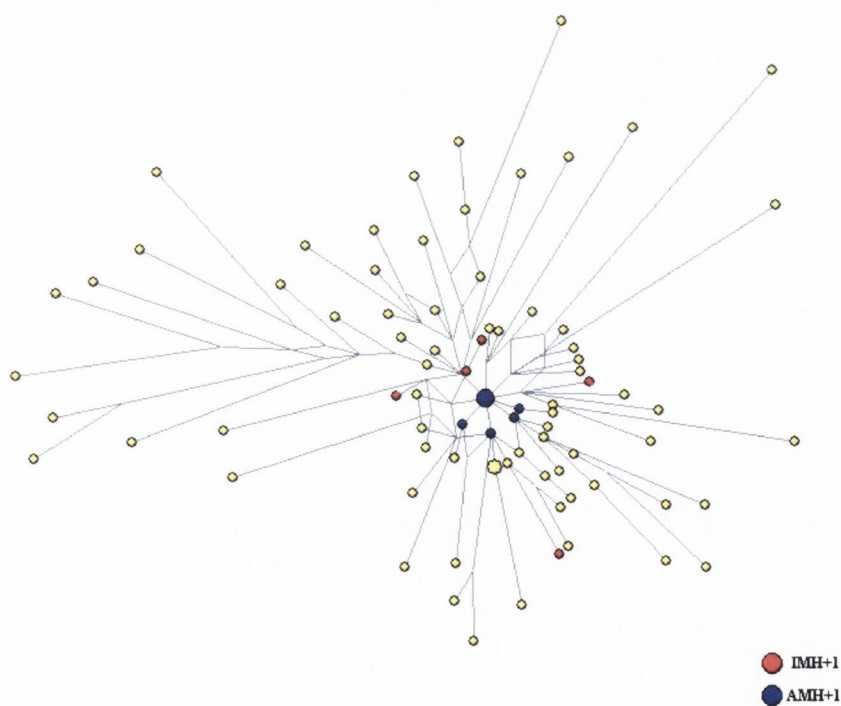


Figure 2.3.7.5: Median-Joining network of Dublin R1b3 samples. IMH+1 shown in red, AMH+1 shown in blue.

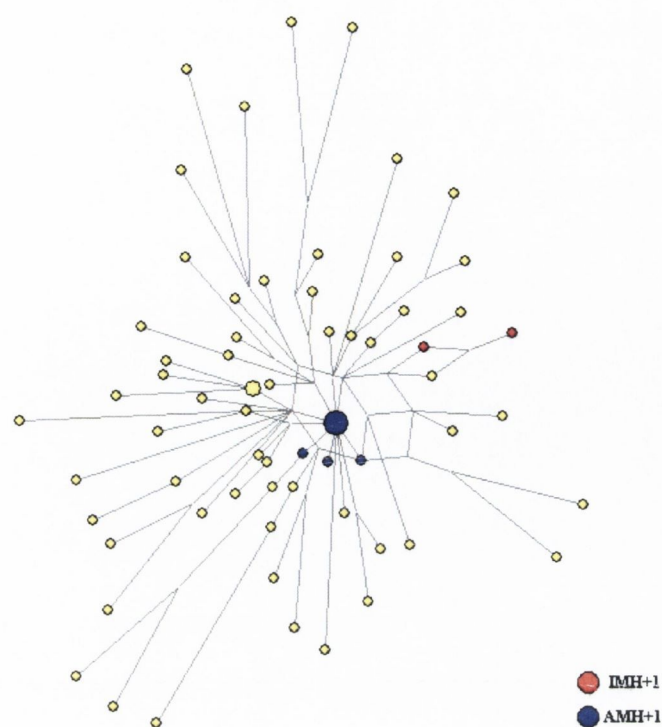


Figure 2.3.7.6: Median-Joining network of Galway, Clare, Tipperary R1b3 samples. IMH+1 shown in red, AMH+1 shown in blue.

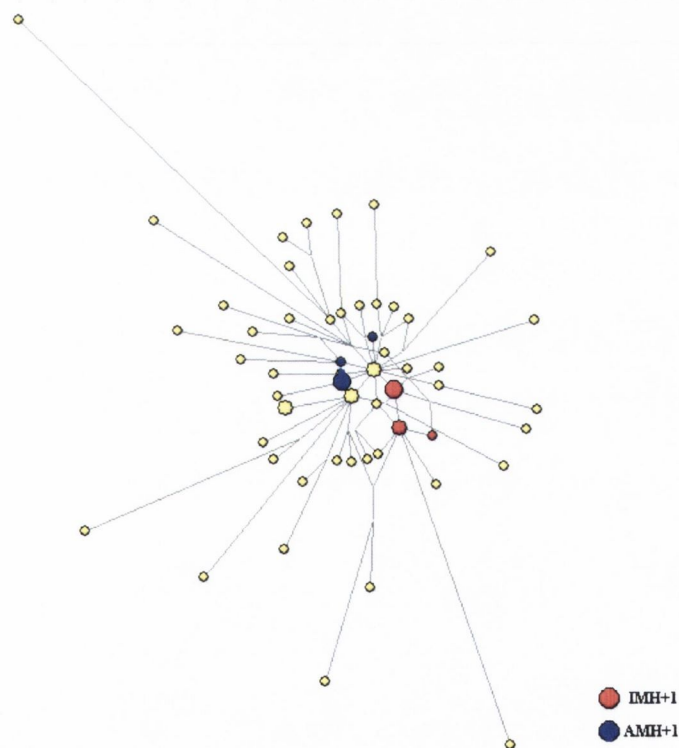


Figure 2.3.7.7: Median-Joining network of Midlands South R1b3 samples. IMH+1 shown in red, AMH+1 shown in blue.

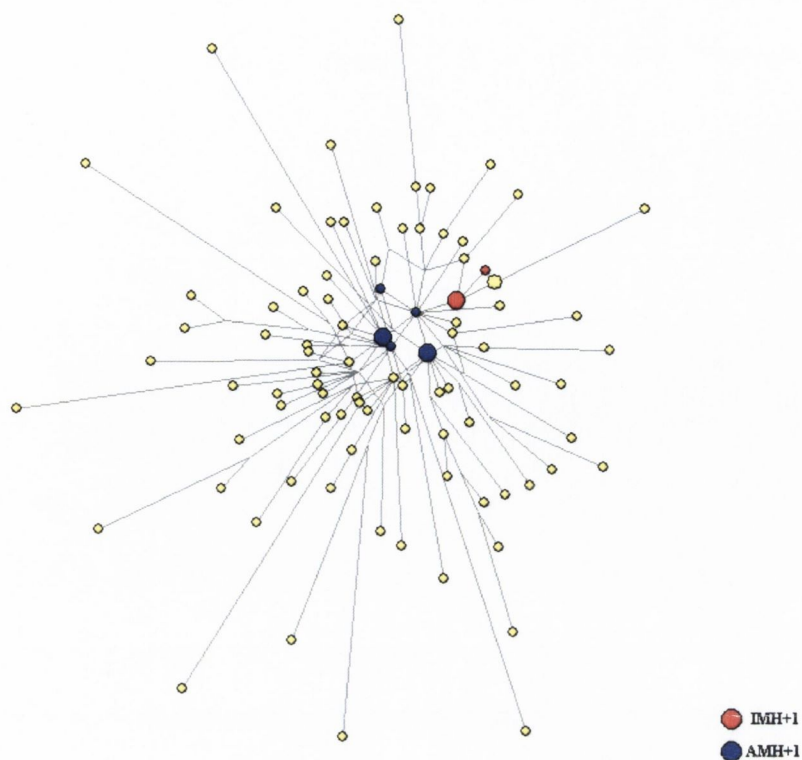


Figure 2.3.7.8: Median-Joining network of Southwest R1b3 samples. IMH+1 shown in red, AMH+1 shown in blue.

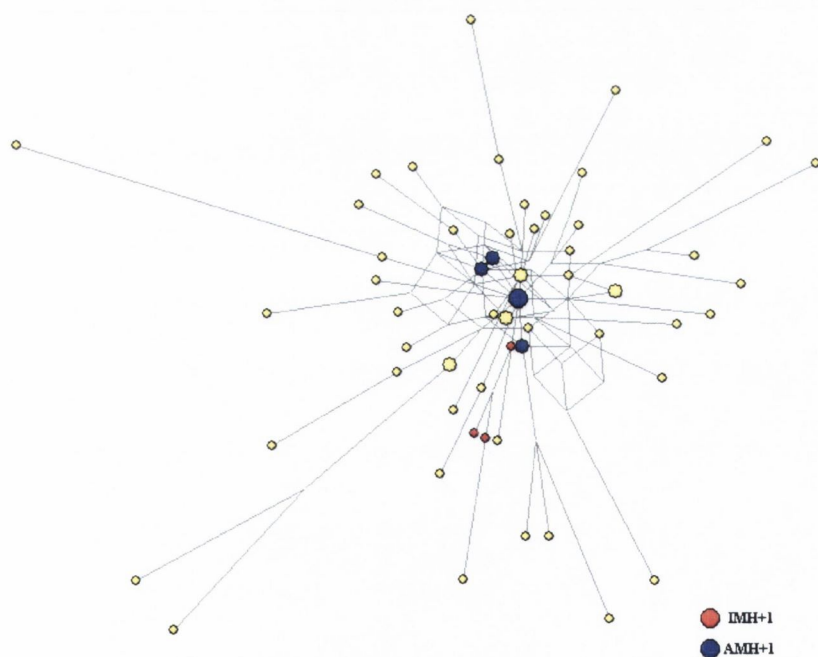


Figure 2.3.7.9: Median-Joining network of Southeast R1b3 samples. IMH+1 shown in red, AMH+1 shown in blue.

Median-joining networks of the R1b3 population of each of the 9 regions are shown in Figures 2.3.7.1 – 2.3.7.9. IMH+1 17-locus haplotypes and AMH+1 6-locus haplotypes in each population are indicated. The networks show that when 17-locus haplotypes are considered, most Irish R1b3 samples occur as singletons. They illustrate the breakdown of the 6-locus AMH haplotype. They also indicate the strikingly high frequency of the IMH haplotype in the Northwest and North Connaught populations.

2.3.8 BATWING posterior distributions

BATWING output parameter distributions for R1b3 samples in the Irish population are shown in Figures 2.3.8.1 – 4. Median, Mode and 95% Credible Intervals for N, ancestral effective population size, alpha, percentage population growth rate per generation, T, the TMRCA and Beta, the estimated time of expansion of Irish R1b3 samples are shown.

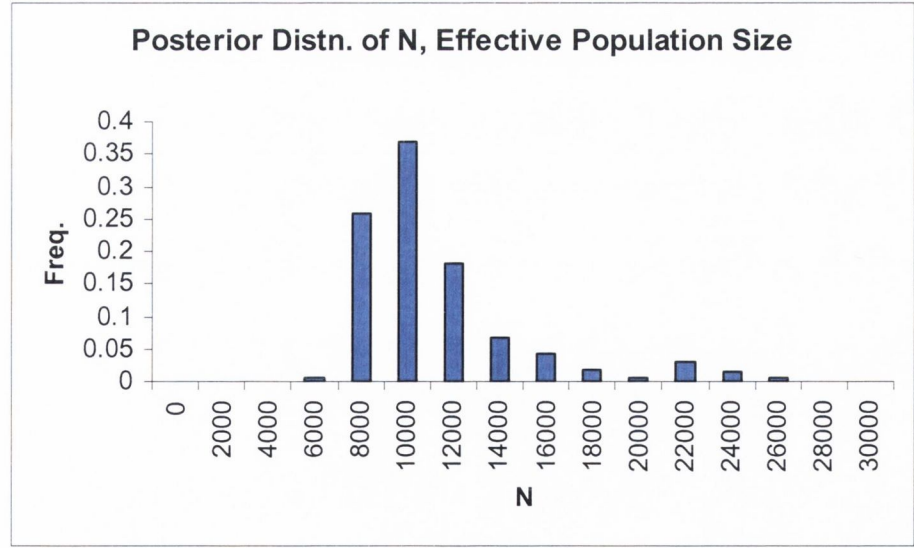


Figure 2.3.8.1: Posterior distribution for N, ancestral effective population size

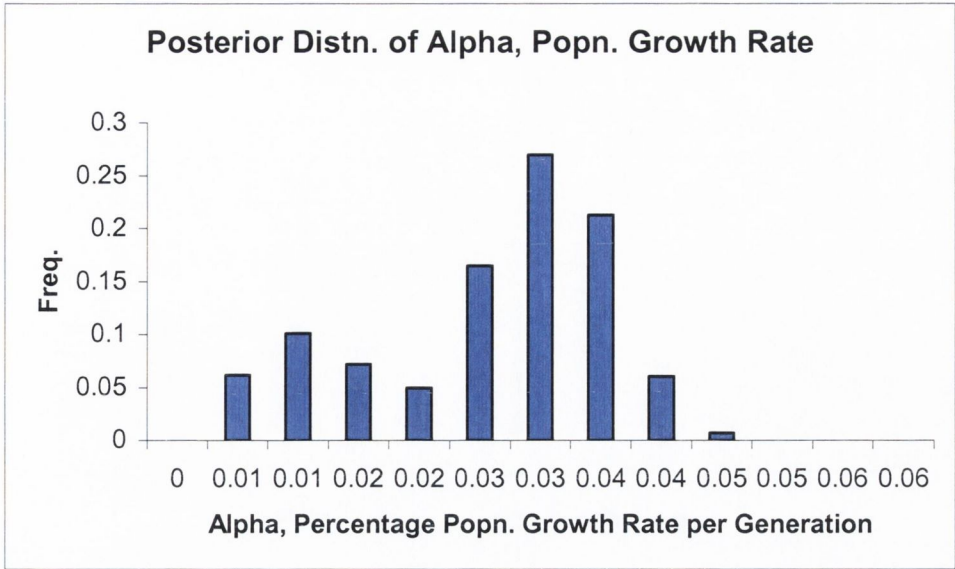


Figure 2.3.8.2: Posterior distribution of alpha, percentage population growth rate per generation.

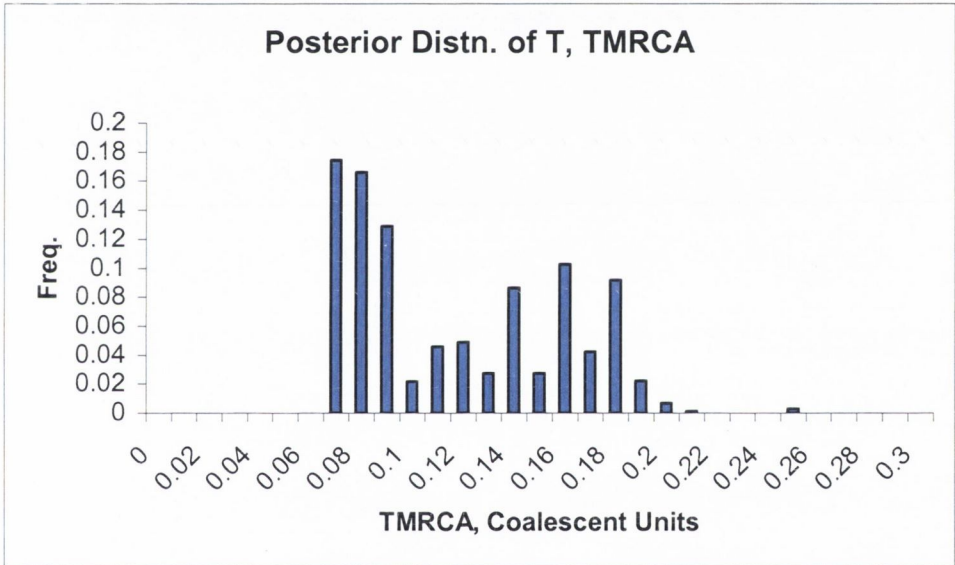


Figure 2.3.8.3: Posterior distribution of TMRCA of Irish R1b3 samples

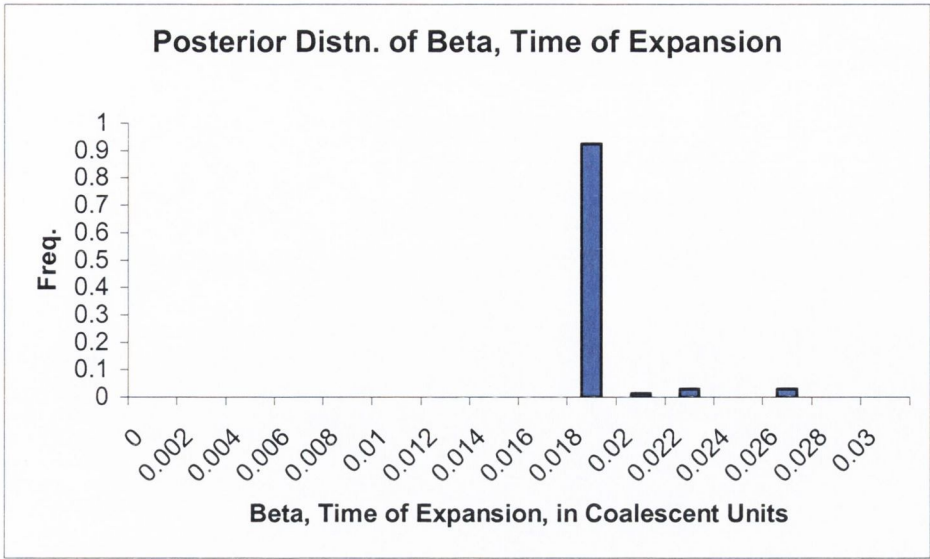


Figure 2.3.8.4: Posterior distribution of beta, estimated expansion time, of Irish R1b3 samples

	Mode	Median	2.50%	97.50%
N	11658	11295	8431	23976
Alpha	0.033	0.026	0.002	0.037
T (Yrs)	17610	28705	12735	108978
Beta (Yrs)	5226	5063	3747	15476

Table 2.3.8.5: Mode, Median and 95% Credible Intervals for Batwing outputs for Irish R1b3 data

2.4 Discussion

2.4.1 Haplogroup frequency

The high frequency of R1b3xR1b3f (henceforth referred to as R1b3, as there is only one R1b3f chromosome in the Irish sample) chromosomes in Ireland (85.4%) places the island near the extreme of the European northwest-southeast cline in R1b3 frequency, which peaks in the Spanish Basque country (88.9%, Semino *et al.* 2000). This haplogroup has been proposed as a Palaeolithic marker in Europe whose high frequency on the Atlantic Fringe of Europe has been attributed to the post-glacial re-expansion through the Atlantic coast of an R1b3 reservoir population from a Franco-Iberian refuge (Semino *et al.* 2000).

When the Irish population is split into regional subpopulations the northeast – southwest cline continues across the island, with R1b3 reaching frequencies of ~95% in the Northwest. R1b3 frequencies are considerably lower in the eastern and southern regions (76.1% in Dublin and 77.4% in the Southeast). It has been argued (Hill *et al.* 2000) that R1b3 (or Hg1) in Ireland is the older indigenous variant and that all non-R1b3 chromosomes are likely the result of later immigration following the initial settlement of the island. If this is the case, then the higher frequency of non-R1b3 chromosomes in the south and east would reflect a greater proportion of immigrant Y chromosomes to these regions than to the western and northern regions. Historical records of Viking and Norman settlements, as well as English and Scottish

Plantations, certainly suggest increased foreign introgression in the south and east. In light of this it is interesting that studies of Y haplogroup distribution in the British Isles (Wilson *et al.* 2000b, Capelli *et al.* 2003) have suggested that certain Y lineages defined by haplogroup IxI1b2 and R1a are the genetic legacy in Britain of Viking and or Anglo-Saxon incursion. Of the non-R1b3 Y chromosomes, only IxI1b2 occurs at substantial frequency in Ireland (10.9%). Splitting the Irish into regional subpopulations reveals a reciprocal west – east cline in frequency of this haplogroup, which peaks in the Northeast (20%), Dublin (17.4%) and the Southeast (14.3%) and with a low in Northwest Ulster 3.3%). The relatively high frequency of IxI1b2 in the Midlands South population (14.7%) may be the genetic legacy of the Leix-Offaly plantation in the 16th Century.

As other non-R1b3 chromosomes occurred at very low frequency it was not possible to detect geographic pattern of their incidence. R1a chromosomes, possibly diagnostic of Viking activity in the British Isles (Helgason *et al.* 2000, Wilson *et al.* 2001b, Capelli *et al.* 2003) are virtually absent (frequency 0.4%), perhaps suggesting that the Vikings did not leave an enduring Y chromosome genetic legacy in Ireland. However, Wilson *et al.* pointed out that it was not possible for them to distinguish whether IxI1b2 chromosomes in Britain were the result of Viking (Danish) or Anglo-Saxon introgression, so it is possible that at least some IxI1b2 Y chromosomes in Ireland are the result of Viking rather than British genetic input. This scenario is supported by the high incidence of IxI1b2 chromosomes in Dublin and in the Southeast where there is well-documented evidence of Viking settlement. However, historical records suggest

that while there was substantial Danish Viking activity in Britain, most Vikings in Ireland were of Norwegian descent. If the Ix11b2 chromosomes in the east and south of Ireland were of Norwegian Viking descent, it would be expected that they would be accompanied by R1a chromosomes at substantial frequency. That this is not the case suggests that non-R1b3 chromosomes in the south and east may rather be the genetic legacy of Anglo-Norman or later British colonial activity, or indeed prehistoric events.

Another subclade of haplogroup I, I1b2, is found at low frequency in the Irish sample (1%). It occurs at low frequency throughout the island. The geographic distribution and age estimates of this Hg in Europe suggest that it arose on a Hg I background in a Franco-Iberian glacial refuge population and from there spread through the Atlantic populations of Europe in an analogous fashion to R1b3 and mtDNA haplogroup V (Rootsi *et al* 2004).

The single N3 chromosome in the Irish population is possibly the result of Viking influence since this haplogroup is found at substantial frequency in Scandinavian populations (Rosser *et al.* 2000).

All R1b3 chromosomes were tested for the marker SRY-2627, defining the subclade R1b3f. This mutation has been shown to have recently arisen within the population of the Iberian Peninsula, where it occurs at frequencies of 11% (Hurles *et al.* 1999). It occurred only once in the Irish sample. This finding is significant because the genetic

similarity of Spanish (particularly Basque) and Irish Y chromosomes may be explained by either an ancient common heritage, whereby both populations possess a high frequency of Palaeolithic Y chromosomes due to postglacial expansion from the Iberian refuge and relative freedom from the population movements of mainland Europe. Alternatively it might be explained by more recent gene flow between the populations arising from navigation and trade networks among the populations of the Atlantic Fringe (Cunliffe 2001). If gene flow due to more recent contact were responsible for the genetic similarity, Hg R1b3f should occur at substantial frequency in the Irish population. Its virtual absence argues against substantial recent gene flow and instead suggests a more ancient Palaeolithic connection between the two populations.

Haplogroup E3b occurs at a frequency of 0.9% in the Irish population. E3b is a YAP+ lineage and its presence in Europe decreases from the southeast to the northwest. Haplogroup FxIJK is defined by the mutation M89 and is found at very low frequency (0.3%). Its frequency also decreases in a cline from southeast to northwest Europe. Semino *et al.* (2000) have argued that these haplogroups (along with J2 which is absent in the Irish sample) represent the male component of Neolithic demographic diffusion from the Fertile Crescent due to their clinal frequency pattern, age and increased diversity in Middle Eastern populations. If this is correct the male Neolithic component in the Irish population is very low, just ~1.2%, compared with ~22% in Europe as a whole, suggesting that agriculture, when it came to the Atlantic Fringe, may have been adopted by a substantially indigenous population.

The main Irish haplogroups, R1b3 and I, together accounting for ~97.3% of the Irish Y sample, are argued to be Palaeolithic in origin (Semino *et al.* 2000, Wells *et al.* 2001). The phylogeny, geographic distribution and age estimates of R1b3 suggest that this marker arose in a Central Asian population ~40,000 – 30,000 years ago, which subsequently migrated into Europe, characterising the earliest modern humans in Europe. The phylogeny, geographic distribution and age estimates of haplogroup I suggest that this marker arose in Europe ~20,000 – 25,000 years ago (Semino *et al.* 2000, Rootsi *et al.* 2004).

2.4.2 Haplotype frequency

When the Irish sample was analysed for 17 quickly-evolving microsatellite markers the 10 haplogroups represented in the Irish population were split into 640 high-resolution compound haplotypes, 583 of which occurred only once. Previous genetic analysis of the Irish NRY population revealed that the most common haplotype, defined by 6 microsatellite markers, was the Atlantic Modal Haplotype, which occurred at a frequency of 18%, or 44% if one-step neighbours are included (Wilson *et al.* 2001b). This haplotype, 14-12-24-11-13-13 at the microsatellite loci DYS 19- DYS 388- DYS 390- DYS 391- DYS 392- DYS 393, and its one-step neighbours, were found at high frequency in Atlantic populations (AMH+1 frequency 70% Wales, 56% Scots, 56% of Basques, compared with 15% in Anatolian Turks and 1% in Syrians) by Wilson *et al.*, who argued that the high frequency of this haplotype in these populations indicated that they formed a Y chromosome community with

members more closely related to one another than they are to the other European populations, and that Basque- and Celtic-speaking populations derive from common paternal ancestors. They suggest a northward expansion from a glacial refugium in Iberia may be responsible for this pattern, as has been suggested to account for the diffusion of Magdalenian industries, the geographic distribution of R1b3 and I1b2 chromosomes, and mtDNA lineage V (Otte 1991, Semino *et al.* 2000, Rootsi *et al.* 2003, Torroni *et al.* 2001).

The 6-locus AMH accounted for 18.3% of Irish chromosomes, or 41.6% together with its one-step neighbours. However when the 17-microsatellite haplotypes are considered, the coherence of AMH breaks down and in a 17-marker system AMH is no longer modal. The most common AMH haplotype occurs 18 times in the Irish sample, at a total frequency of 2.3%. The most common Irish haplotype (Irish Modal Haplotype or IMH) is 2 mutational steps removed from AMH in the 17 marker system and occurs 30 times in the Irish sample, at a total frequency of 3.8%, or 8.2% when one-step neighbours are considered. When the Irish sample is split into regional subpopulations, the frequency distribution of IMH and its one-step neighbours is highly non-random. The contour map of Irish IMH+1 frequency (Figure 2.3.3.1) shows that it peaks in the northwest, where it reaches a frequency of 21.5% in the Northwest population and 30.8% in the sample from Derry. It occurs at a frequency of 15.3% in the North Connaught sub-population. It occurs at lower frequency in populations outside the north and west of the country and in the populations of the British Isles.

2.4.3 Genetic Distance Measures

Fst pairwise population genetic distance analysis indicates that the greatest genetic distances among the Irish regional sub-populations are those between the Northwest sub-population and other populations. This may be due to the high frequency of IMH+1 in this sub-population, especially since the Northwest is not significantly distant from North Connaught, the other sub-population containing IMH+1 at high frequency. This suggests that populations in the North and West of Ireland may form a distinct coherent Y chromosome community. Another factor likely to contribute to the northwest / southeast divide in Ireland is likely to be the northwest – southeast decrease in R1b3 frequency.

2.4.4 Principal Components Analysis

PC analysis illustrates the genetic relationship of the regional Irish subpopulations with the British populations of Capelli *et al.* (2003), and samples from Norway, Germany and Denmark.

Capelli *et al.* found that the AMH+1 category had the largest impact on the PC plot. They found that those samples with the largest proportion of AMH+1 were those with higher ‘indigenous’ and lower continental input. These were geographically the more western populations, including the Irish, Welsh and Basque samples.

When the data from the 9 Irish subpopulations are added, the greatest impact on the PC plot no longer comes from AMH+1, but R1b3 (defined by all R1b3 data excluding AMH+1) as Figures 2.3.5.1 - 2.3.5.4 show. The northern and western Irish populations appear skewed to the right of the PC1 axis, with the eastern and southern Irish populations falling closer to the British and continental samples, reflecting their higher proportion of AMH+1 and non-R1b3 samples. This serves as further indication that NRY populations in the north and west of Ireland are genetically distinct from their Irish and British neighbours.

2.4.5 Multidimensional Scaling Analysis

The Hg classification method of Cappelli *et al.* (2003) introduced a bias to their analysis, because all R1b3 chromosomes were classified on the basis of whether or not they were AMH+1. The high frequency of IMH+1 in the north and west of Ireland suggests that this classification method may not be suitable for the analysis of Irish data. In order to eliminate this bias towards AMH, the data of Capelli *et al.* (2003) and the Irish data were analysed by multidimensional scaling of pairwise population genetic distances, as outlined. The MDS plot (Figure 2.3.6.1) shows that the Irish populations cluster together to the exclusion of other populations from Britain and Europe, with the exception of Oban which Possess the IMH at a frequency of 7.1% and Pitlochry which possesses IMH+1 at a frequency of 7.3%. These populations, Oban in particular, are located in Western Scotland. The close proximity of Western Scotland and Northern Ireland has facilitated migration between the two regions and this may explain the relatively high frequency IMH

chromosomes in these samples. For example, an Irish migration to Scotland in the 6th Century A.D. to found the kingdom of *Dal Riata* is hypothesised. Migration from northern Ireland to Scotland for economic reasons was common in the 19th and 20th Centuries also.

MDS analysis also indicates that the Northwest population is significantly distant from all other populations, including the Irish, possibly due to its high frequency of IMH+1. Its nearest neighbour is North Connaught, also containing a high frequency of IMH+1. York, Stonehaven and Orkney fall close to the Norwegian, North German and Danish samples, suggesting significant genetic input from Anglo-Saxons and Vikings in these regions.

However when the data are analysed by MDS, Capelli *et al.*'s tri-polar pattern of Basque, German and Danish influence breaks down, with the Basque and Celtic regional samples no longer clustering, suggesting that such a pattern may be of limited usefulness for assessing continental admixture in British Isles populations.

2.4.6 Median-joining networks

Median-joining networks of R1b3 samples in the Irish regional subpopulations illustrate the uneven nature of haplotype distribution. Most regional networks are comprised of singletons, with no haplotypes having numerical or topological predominance. In contrast, the networks of the Northwest, Northeast and North Connaught populations have the IMH as a major feature. All regional networks also

indicate the breakdown of the 6-locus AMH haplotypes when 17 markers are considered.

2.4.7 Dating Estimates

Batwing estimated the date of R1b3 coalescence in Ireland to ~17,600 years (modal value). The 95% credible intervals associated with this are very large (12,700 – 109,000 years) and indicate the large degree of uncertainty associated with this estimate, which can be visualised in the ragged distribution of the TMRCA output distribution in figure 2.3.8.3. It may be concluded from this that the TMRCA date contains little information.

BATWING estimates of expansion time, on the other hand, are more promising. In contrast with the TMRCA distribution, the expansion (beta) distribution is much more condensed, with a modal value and median value both of ~5,200 years (Figure 2.3.8.4). It is tempting to associate this expansion event with the introduction of agriculture to this region of the western Atlantic, which occurred around this period. This would further suggest that agriculture in Ireland was adopted by a substantially indigenous population. However the uncertainty associated with this date, indicated by the 95% Credible Intervals (lower 3,700 years, upper 15,400 years), means that a number of scenarios for population expansion are encompassed in this time range, the upper limit of which will accommodate the first settlement of the island or indeed the postglacial re-expansion of an R1b3-carrying population from the Franco-Iberian refuge.

2.5 Conclusion

This detailed, high-resolution investigation of Irish Y chromosome phylogeography has shown a high frequency of the putatively western European Palaeolithic Hg R1b3 in Ireland. When the Irish population is split into regional sub-populations a cline in Irish R1b3 frequency is revealed, which spreads from the east to the west. The frequency of R1b3 in the Dublin and Southeast sub-populations is comparable with that found in many British populations (Capelli *et al.* 2003), but the frequency of R1b3 in the northwestern populations is higher than that found in the Basque country (Semino *et al.* 2000). This places northwest Ireland on the extreme of a European-wide cline in the frequency of this Palaeolithic marker.

The higher frequency of non-R1b3 chromosomes to the east and southeast suggests a higher proportion of genetic introgression to these areas, consistent with historical evidence of Viking, Norman and British influence in these regions, although prehistoric genetic introgression to these areas may also explain this pattern.

The near absence of R1a chromosomes in the Irish sample suggests that the Vikings did not leave an enduring genetic legacy on this island. Viking activity was greatest in the east and southeast. Historical records suggest that the Vikings who came to Ireland were predominantly Norwegian in origin. Helgason *et al* (2000) and Wilson *et al.* (2001b) have shown that genetic records of Norwegian Viking incursion should indicate a higher frequency of both R1a and IxI1b2 Y chromosomes in the

population. The east and southeast do have a relatively high frequency of Ix11b2, but not of R1a, suggesting that non-R1b3 chromosomes in this region did not come from a Norwegian Viking source.

High-resolution microsatellite haplotype analysis has shown that within Ireland, R1b3 haplotypes have a non-random distribution. In a 17-marker system the Atlantic Modal Haplotype is no longer modal. However regional analysis of the Irish R1b3 sample shows that the predominance of the Irish Modal Haplotype is not a feature of all Irish populations, but is largely confined to the north and northwest. In a 17-marker system most haplotypes in the Irish sample occur as singletons. The high frequency of IMH in the northwest, its coherence relative to other haplotypes and its low frequency outside Ireland suggest that relatively recent population events specific to the northwest of the island have been responsible for its prevalence.

Chapter 3: Investigating the Origins of the Irish Modal Haplotype.

Within the Irish R1b3 population one 17-microsatellite haplotype has unusual features. It is the most common haplotype in the Irish sample, occurring 30 times. Together with its one step neighbours IMH accounts for ~8.2% of the Irish R1b3 sample. It has a geographical distribution centred on the northwest, where it reaches frequencies of ~31% in Derry.

Zerjal et al. (2003) found a lineage occurring at high frequency (8%) in Central Asian populations. They proposed that this lineage rose to high frequency in a short timeframe due to a novel form of selection based on social prestige.

This pattern of a single lineage occurring at strikingly high frequency is mirrored in the IMH and its localisation in the northwest coincides with the territory of the most powerful and durable dynasty in Medieval Gaelic Ireland, the Uí Neill.

Y chromosomes from 33 Irish males with surnames supposedly derived from the Uí Neill lineage were genotyped in order to investigate whether the IMH is associated with the Uí Neill dynasty

3.1 Introduction

3.1.1 The regional frequency of IMH+1

The high frequency the 17-microsatellite IMH haplotype (~8.2% with one step neighbours) in the Irish sample is striking. Median-joining networks of R1b3 samples have illustrated the dominance of the IMH in the samples from the northwest and North Connaught populations (Figures 2.3.7.1 and 2.3.7.3). Figure 3.1.1.1 below shows a median-joining network of R1b3 samples from the Northwest and North Connaught populations combined. IMH is a major feature of this network. In comparison, a median-joining network of R1b3 samples from a region in southwest Ireland (Cork, Kerry and Limerick, Figure 3.1.1.2), indicates that in this region no one haplotype has numerical or topological predominance. This, together with other regional networks of R1b3 samples shown in Chapter 2, indicates that IMH+1 is a major feature of the north and west only. The relative coherence of this haplotype and its confined geographic distribution suggest a recent origin for this haplotype.

3.1.2 Dating Expansion in the Northwest

In order to further investigate the origins of the IMH a BATWING expansion date estimate was performed on R1b3 samples from the counties where IMH is most prevalent, namely Donegal, Derry and Tyrone. Using prior and model specifications indicated in Chapter 2, population expansion in this region was dated to ~4,800 YBP (outputs for TMRCA, beta, alpha and N are given in Section 3.3.1). The wide Credible Intervals associated with this date indicate its uncertainty (95% CI ~2,400 to

19,000 YBP), and therefore another dating method, the Rho statistic (Morral *et al.* 1994), was applied. To calculate this, a median-joining network of R1b3 samples from Derry, Donegal and Tyrone was constructed using Network 4.0 as before. The dating option in Network 4.0 was used and descendant nodes were chosen as those linked to IMH without gaps. A rate of 1 mutation every 2,131 years was used, based on the rate of Zhivotovsky *et al.* 2004, which led to a TMRCA estimate for this lineage of 932 YBP (Std. Dev. 416 – 1,448 YBP (Saillard *et al.* 2000)). This date estimate together with the geographic distribution of IMH+1 suggest an early Medieval origin for this lineage in northwest Ireland.

3.1.3 The Genetic Legacy of the Mongols

A survey of Central Asian Y chromosomes (Zerjal *et al.* 2003) has revealed a haplotype with certain features similar to the IMH. Central Asian populations possess a haplotype at high frequency (8%) and with widespread geographic distribution from the Pacific to the Caspian Sea. Diversity levels suggest a Mongolian origin for this lineage, whose origins were dated using BATWING to ~1,000 YBP (95% CI 700 – 1,300 YBP) and using the Rho statistic (Morral *et al.* 1994) to ~860 YBP (590 – 1300 YBP).

The pattern of variation of this lineage does not fit with a general population expansion from Mongolia, since this would be expected to result in several Mongolian lineages being present in other Asian populations. Zerjal *et al.* therefore argued that its high frequency and geographic distribution were the result of social

selection acting on a particular Y chromosome lineage in such a way that this lineage, arising in Mongolia, has spread within the last 1000 years to represent ~8% of the males in a region stretching from northeast China to Uzbekistan.

Zerjal *et al.* argued that the most likely explanation for such a pattern was that the Y lineage in question was carried by male line descendants of Genghis Khan (1162 – 1277), who established the largest land empire in history. He and his close male relatives had many children and their male line descendants ruled large areas of Asia for many generations. In this scenario this Y lineage rose to prominence in Asia due to selection, which was not biological, but social, arising from human activities recorded in history.

There are similarities between the IMH and the Genghis Khan haplotype, which may suggest that the IMH arose to prominence in the Irish Y chromosome sample by a similar form of social selection due to linkage with an ancient and enduring dynastic hegemony.

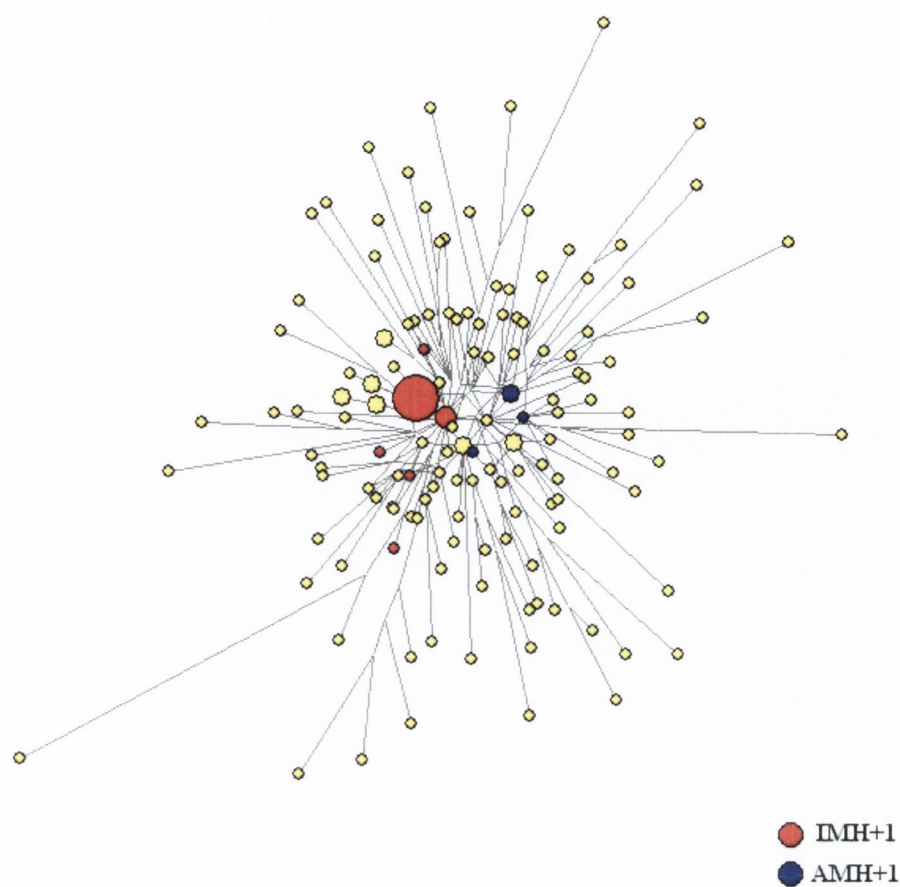


Figure 3.1.1.1: Median-joining network of R1b3 samples from Donegal, Derry, Tyrone, Fermanagh, Leitrim, Mayo, Sligo, Roscommon. IMH+1 samples are shown in red and AMH+1 samples are shown in blue.

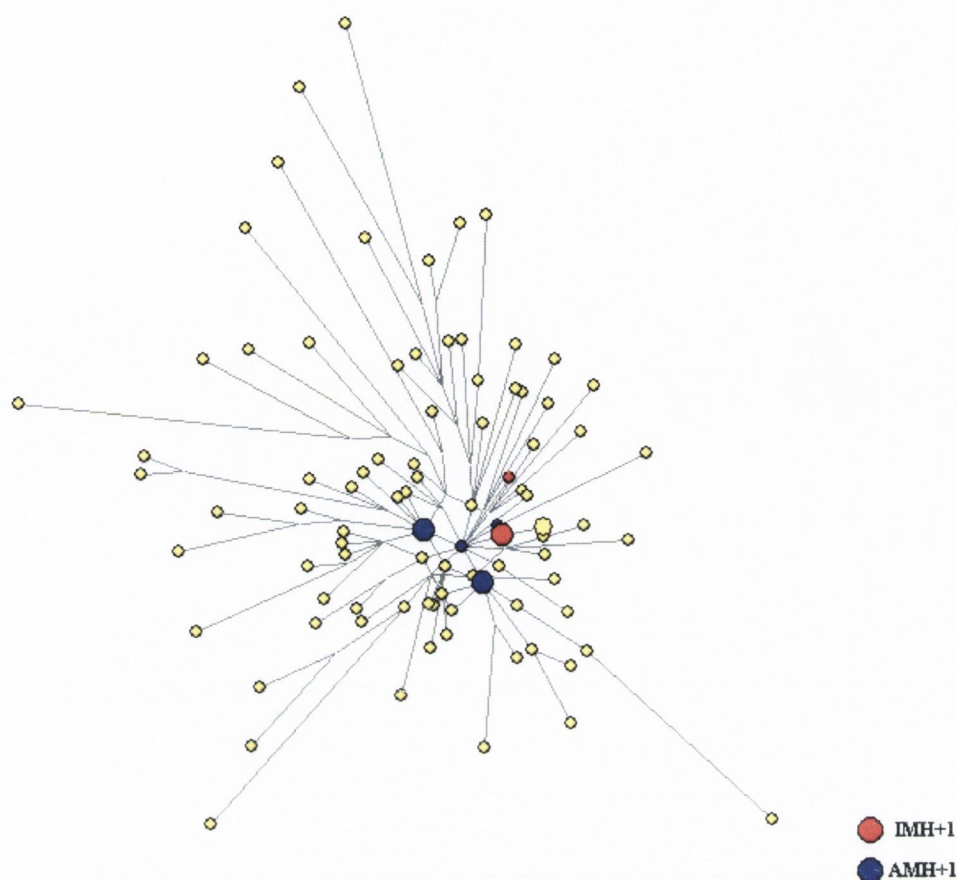


Figure 3.1.1.2: Median-joining network of R1b3 samples from Cork, Kerry and Limerick. IMH+1 samples are shown in red and AMH+1 samples are shown in blue.

3.1.4 Early Medieval Irish Society

Early Medieval Irish society was tribal and placed great emphasis on family relationships. These were organised around a strongly patrilineal system whose features acted to propagate a particular Y chromosome lineage: Land and title could be handed down to successors chosen from within a kin-group of male lineage relatives, the *derbhfine*, defined as the group of all males descended from the same paternal great-grandfather. This was a strong defence against the dissociation of lineage and power, and would have acted to preserve the social status of a particular Y chromosome lineage. Whereas medieval Ireland was Christian, earlier marriage

customs persisted and allowed divorce and concubinage. One of the features of this was that illegitimate sons were claimed and had rights protected in law. As in other polygamous societies, the siring of offspring was related to power and prestige (Mac Nicoll, 1972).

Early Medieval Irish history is semi-mythological. What is known of it comes from oral tradition, epic tales such as the *Táin*, the study of Early and Medieval Irish texts and genealogies such as the Annals of the Four Masters, The Book of Ballymote and the Book of Lecan. The oldest portion of the genealogies purport to deal with the 5th and 6th centuries. They were important records used to validate claims to prestige and power. They are not historical in the literal sense, but rather represent how subsequent Irish families saw their interrelationships and it is difficult to ascertain when these genealogies become an actual record of descent.

3.1.5 The Uí Neill

The localization of the IMH in the northwest coincides with the territory of the most powerful and durable dynasty in Medieval Gaelic Ireland, the Uí Neill. The ultimate origin of the Uí Neill is attributed to the conquering sons of the eponymous and possibly mythological 5th century Niall of the Nine Hostages. His descendants established a provincial kingdom in northwest Ulster into which they later incorporated the central part of the province. From this time until the Elizabethan Plantation in the 17th Century, Ulster was the leading Gaelic province of Ireland. The Uí Neill figured prominently in Irish history and were almost without interruption the

High Kings of Ireland from the 7th to the 11th Century (Mc Lysaght, 1960). The epicentre of the enduring Uí Neill power coincides with the northwest peak in the frequency of IMH+1.

The Uí Neill were a dynastic lineage and must be distinguished from the bearers of the later surname O'Neill. The Uí Neill lineages were much wider and included ruling families who took other surnames when patrilineal surnames were adopted in Ireland circa 950 AD. The prefix Uí (or O) denotes descent and Uí Neill as used in the genealogical tracts refers to those claiming descent from Niall of the Nine Hostages. It therefore embraced, as well as the O'Neills, families bearing surnames such as Quin, O'Doherty, O'Donnell and MacLochlainn.

3.2 Materials and Methods

3.2.1 Sampling Strategy

In order to investigate whether IMH is associated with the Uí Neill dynasty, a further 33 samples with surnames which, according to the genealogies, indicate descent from the Uí Neill lineage were analysed. There are ~30 such names, which can be linked with differing degrees of certainty to this patrilineage, and prior work has indicated that Irish surnames carry significant genetic associations (Hill *et al.* 2000).

Samples were collected for the following surnames: O'Boyle / Boyle (4), Bradley (2), Cannon (1), Devlin (1), Doherty (1), Donnelly (2), Egan (1), Flynn (2), Gormley (1), Hynes (2), McKee / Kee (2), McCaul (1), McGovern (1), McLoughlin (1), McManus (1), McMenamin (1), Molloy (1), O'Connor (3), O'Kane (1), O'Reilly (2), O'Rourke (1), Quinn (1).

3.2.2 Genotyping and Analysis

These 33 samples were analysed for Y-linked SNP and microsatellite markers as indicated in the materials and methods section of Chapter 2. The surname information, genotype and number of mutational steps from the IMH for each sample are indicated in Appendix Section 3. A median-joining network of Uí Neill R1b3 chromosomes was constructed using Network 4.0 as before (Figure 3.3.2.1).

The resulting Uí Neill R1b3 samples were then compared with 85 R1b3 samples with geographic origin in Ulster (from the Northwest and Northeast populations), but of which there is no surname data available. Using the number of mutational steps from the IMH as a metric, Mann Whitney tests implemented in the SPSS statistical package were used to test for significant association between the Uí Neill surname sample and the IMH. The Uí Neill R1b3 samples were also compared with the general R1b3 Northwest sample of Chapter 2.

3.3 Results

3.3.1 Batwing Posterior Distributions for Derry, Donegal, Tyrone R1b3 Samples

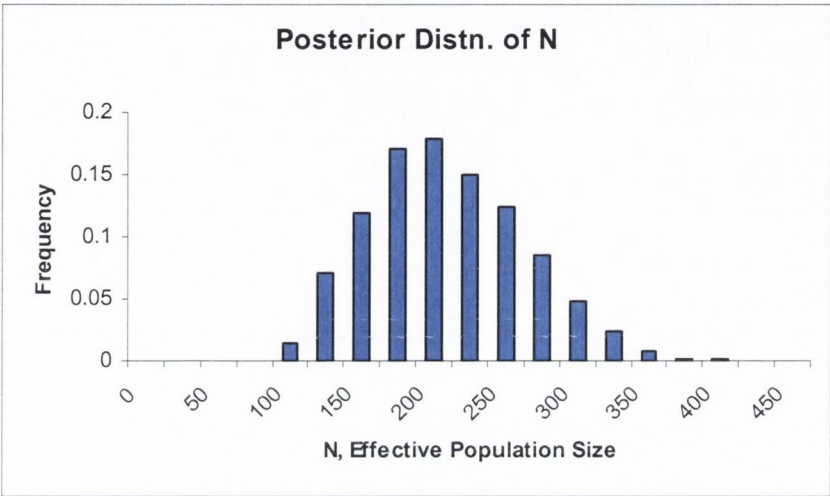


Figure 3.3.1.1: BATWING posterior distribution for estimate of N, ancestral effective population size

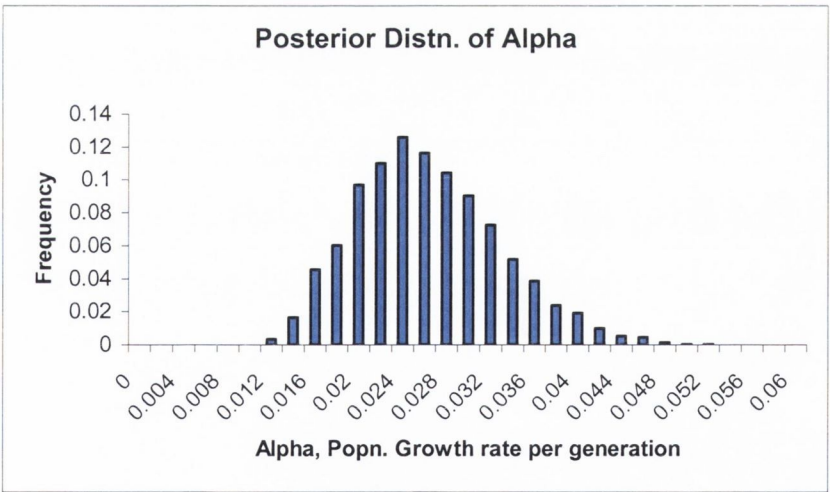


Figure 3.3.1.2: BATWING posterior distribution for estimate of alpha, percentage population growth rate

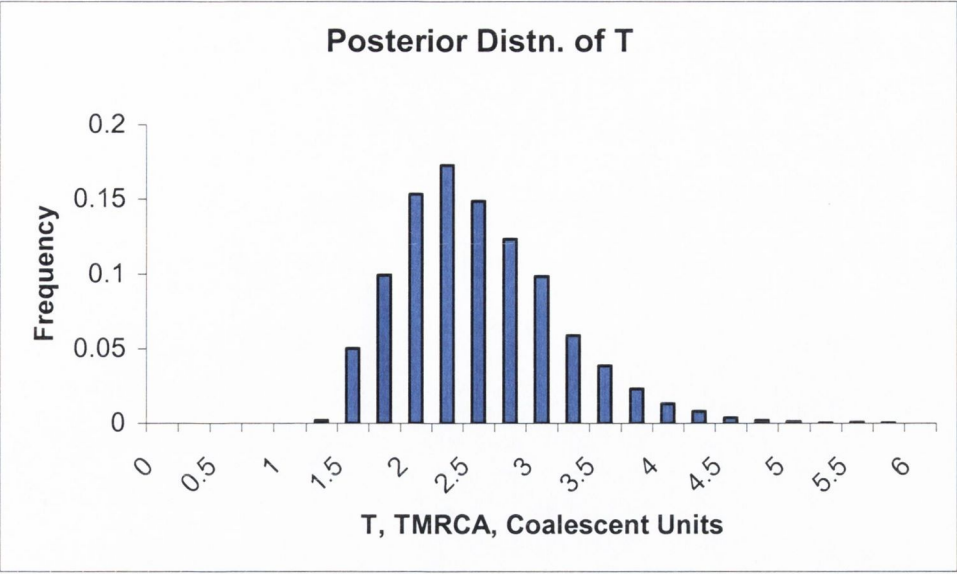


Figure 3.3.1.3: BATWING posterior distribution for estimate of TMRCA of Derry, Donegal, Tyrone R1b3 samples

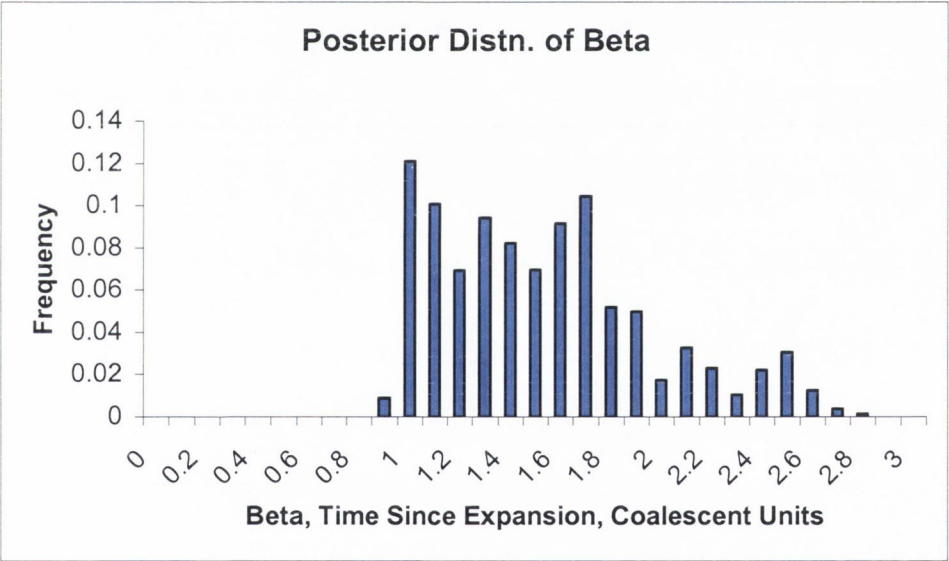


Figure 3.3.1.4: BATWING posterior distribution for Beta, estimate for time of onset of expansion in Derry, Donegal, Tyrone R1b3 samples

Mode, Median and 95% Credible Intervals for Batwing Outputs

	Mode	Median	2.5%	97.5%
N	192	126	104	307
Alpha	0.025	0.023	0.014	0.040
T (yrs)	10958	6787	3666	29728
Beta (yrs)	6817	4783	2435	19029

BATWING posterior distributions for N, effective population size, alpha, population growth rate, TMRCA and beta, time of population expansion for R1b3 samples from Derry, Donegal and Tyrone are shown above. The posterior distribution for beta, time of expansion, is ragged, indicating that estimates for expansion dates taken from his distribution should be interpreted with caution. The considerable difference between the median and modal estimates for expansion (~6,800 and ~4,700 YBP respectively) also highlights this. The median expansion date estimate for this regional sample is older than that for the entire Irish R1b3 sample (~5,100 YBP). Given these discrepancies an alternate dating method was applied, using the dating option in Network 4.0 (as described in section 3.1.2) and gave an earlier estimate of 932 YBP (std. dev. ~416 – ~1,448 YBP).

3.3.2 Network of Uí Neill R1b3 samples

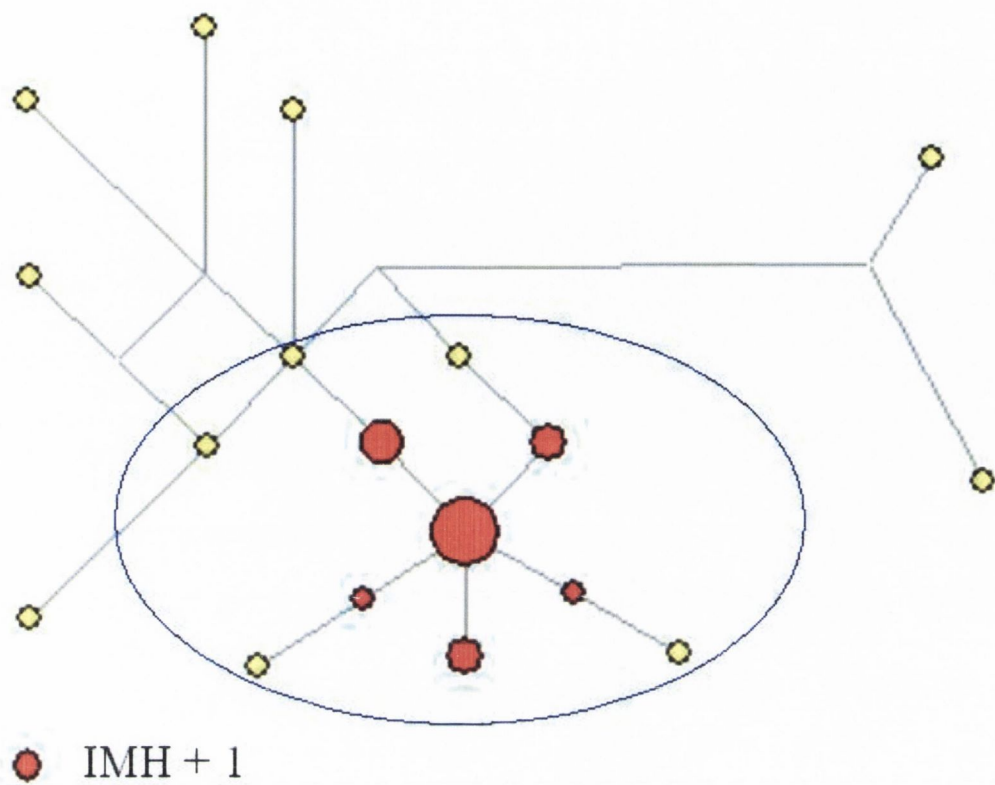


Figure 3.3.2.1: Median-Joining Network of Uí Neill R1b3 samples. Samples enclosed in the ellipse are those used in the Rho dating estimate.

A median-joining network of Uí Neill R1b3 samples is shown in Figure 3.3.2.1. It highlights a coherent cluster around IMH for this sample. When the diversity around this is dated (using the Rho statistic (Morral *et al.* 1994, Forster *et al.* (1996), a mutation rate of 1 every 2,131 years and specifying descendant nodes as those which are connected to the IMH without gaps, as before) it yields an estimate of common ancestry consistent with an early medieval time depth of ~1,900 YBP (Std. Dev. ~1,200 – 2,600 YBP (Saillard *et al.* 2000)). The distribution of this haplotype among the surnames also provides an independent date estimate for the origin of this

haplotype >1,000 years ago, since historical evidence suggests that surnames were adopted in Ireland ~1,000 YBP.

3.3.3 Significant association of IMH+1 with the Uí Neill sample

The frequency of IMH+1 in the northern sample was ~20%, or ~21% if only R1b3 chromosomes are considered; compared with a frequency in the Uí Neill sample of ~52%, or ~59% if only R1b3 chromosomes are considered. Mann-Whitney tests showed a significantly higher affinity of IMH with the Uí Neill R1b3 surname sample than to the Ulster R1b3 sample or to the Northwest R1b3 sample (Table 3.3.3.2).

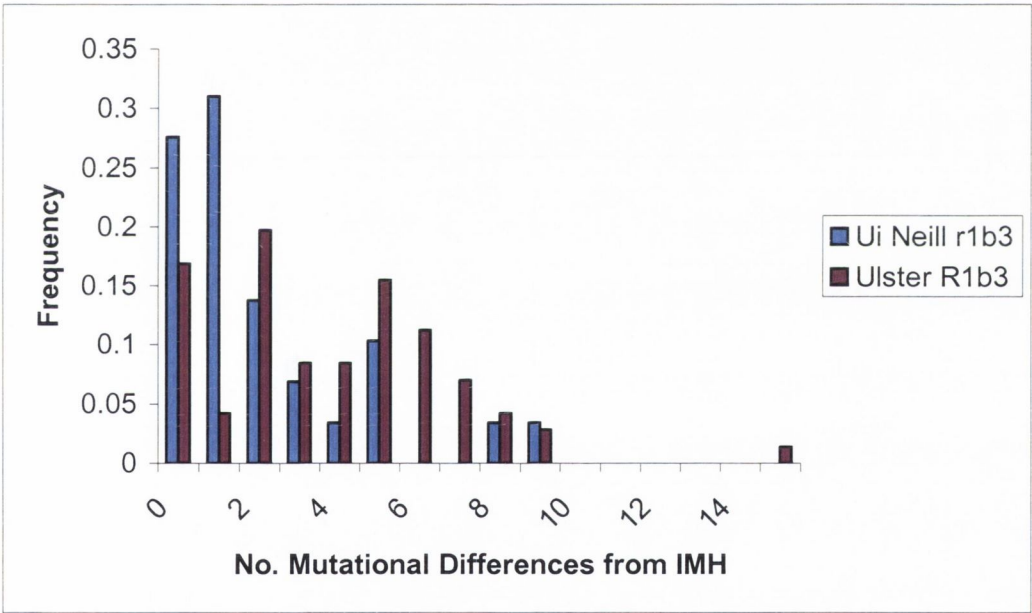


Figure 3.3.3.1: Number of differences from IMH of R1b3 chromosomes from the Uí Neill sample (blue) versus a general Ulster sample with no surname information attached (red).

Figure 3.3.2.1 shows a frequency distribution of the number of differences from the IMH haplotype of R1b3 samples in the Uí Neill sample versus the general northern R1b3 sample with no surname information which was used as the comparison dataset

for the Mann Whitney test. It shows that the IMH and its one-step neighbours occur at much higher frequency in the Uí Neill sample than in a general surname-blind northern sample.

Popn. 1	N	IMH+1	freq	Popn. 2	N	IMH+1	freq	Assymp. 2 tailed sig.
Ulster (no surname)	71	15	0.211	Ui Neill	29	17	0.586	0.002
Northwest	62	14	0.226	Ui Neill	29	17	0.586	0.007

Table 3.3.3.2: Assymptote 2-tailed significance results for Mann Whitney comparisons of Ui Neill R1b3 samples versus (i) an Ulster sample with no surname information, (ii) the Northwest sample of Chapter 2. In both cases there is highly significant association between IMH and the Ui Neill sample.

3.4 Discussion and Conclusion

The predominance and pattern of variation of this haplotype suggest a rise in its frequency due to social selection associated with the hegemony of the Uí Neill and their descendant dynasties. The extent to which figures of early Irish history are real or mythological is unknown, but this is some confirmation of the existence of a single early Medieval progenitor of the most powerful and enduring dynasty in Irish history. These findings suggest that the phylogenetic predominance of the IMH lineage in Ireland is a biological record of past hegemony and give some confirmation of the semi-mythological ties indicated in the Genealogies and in oral Gaelic tradition.

Chapter 4: A Comparison of genetic diversity on the X and Y chromosomes in Ireland, Turkey and Togo.

Genetic diversity on the X and Y chromosome was examined in three populations of different genetic backgrounds: Ireland, Turkey and Togo. The Irish and Turkish populations exist at opposite poles of a genetic gradient that runs from the northwest to the southeast of Europe. To provide a context for the Irish and Turkish populations, samples were compared with X and Y chromosomes from Togo in West Africa. 340 X and 211 Y microsatellite haplotypes were analysed from male DNA samples in these populations, comprised of 5 markers on the X and 16 markers on the Y chromosome, in order to assess the genetic diversity and relationships within and between these populations.

4.1 Introduction

4.1.1 Ireland

The most obvious genetic gradient in Europe runs from the southeast to the northwest and Ireland and Turkey are geographically situated at opposite poles of this gradient. Human history in Ireland is relatively recent, likely post-dating the Last Glacial Maximum, and genetic studies have indicated a relatively homogeneous population, with high frequencies of putatively Palaeolithic markers, such as the R1b3 Y Hg and the O blood group (Hill *et al.* 2000, Hackett and Dawson 1958).

4.1.2 Turkey

In contrast with the northwest of Europe, human presence in the Near East is ancient. The Turkish population has a complex and heterogeneous history. This region was one of the first settled by modern humans dispersing from Africa (Mellars 1992) and has been in the path of many subsequent migrations and dispersals. Reflecting this, the Turkish population is comprised of a number of ethnic groups of diverse origin. Turkey is also located in the region which has been proposed as a location for the invention of agriculture, an event that is likely to have had a profound population genetic impact.

4.1.3 Togo

To provide a context for the Irish and Turkish populations, samples were compared with X and Y chromosomes from Togo in west Africa. Anatomically modern humans have an ancient history in sub-Saharan Africa. There is reasonable consensus that modern humans originated there and genetic distance measures between Africans and

non-Africans are the largest in the species. Initially genetic studies on Africa focused on the question of human origins, but more recently the genetic effects of later demographic events have been considered (Salas *et al.* 2002, Pereira *et al.* 2003).

4.1.4 X and Y chromosome comparisons

This study compares diversity among populations with different genetic histories. Different loci in the human genome have different genealogies and record demographic history in different ways. The Y chromosome is used frequently in population genetic studies to trace the paternal genealogy of the species but it is only one locus in the genome. Analysis of the X chromosome should give an insight into how demographic events shape diversity in the autosomal part of the genome. X spends two thirds of its time in the female line and has a history that is influenced by both sexes, but more so by females. A comparison of X and Y diversity in the same populations allows the investigation of how demographic events in a single population are recorded by different loci.

4.1.5 The Irish Y chromosome sample

The Irish Y chromosome sample used for this study came from that used by Hill *et al.* (2000) and presented in Helgason *et al.* (2000). As Chapter 2 indicates, there is a large amount of Y chromosome data available on the Irish population. However since the Irish X chromosome sample and the X and Y samples of the other populations are relatively small, use of the smaller Irish Y sample facilitated direct comparison. The 135 Irish Y chromosome used in this study are a subset of the larger sample presented in Chapter 2. The smaller Irish Y sample has similar haplogroup and haplotype frequencies to the larger Y sample presented in Chapter 2.

4.2 Materials and methods

4.2.1 X chromosome genotyping

X chromosomes were genotyped for 5 microsatellite markers by Dr. Emmiline Hill (Hill 2000) and are subjected to novel statistical analysis here. X chromosomes were analysed in 340 males. Of these 262 were Irish and this sample was further subdivided into east and west sub-populations by birthplace of maternal grandmother, or if not available on the basis of maternal birthplace. The east sub-population was defined as those samples originating in the Irish provinces of Leinster and Ulster, the west sub-population was defined as those samples originating in the provinces of Connaught and Munster. The east sub-population contained 137 samples and the west 125. 42 samples were Turkish and 36 were Togolese.

Haplotypes from microsatellite loci on the X chromosome were constructed from 4 dinucleotide (F3, F7, F13 and F29) and 1 trinucleotide (F14) repeat loci located on cosmid hsu177e8 of the X chromosome (GenBank accession no. Z68694). Details of amplification and genotypes can be found in Hill (2000).

4.2.2 Y Chromosome Genotyping

16 Y microsatellites were assayed in 211 samples. 135 samples were Irish and were subdivided into east and west sub-populations based on birthplace of paternal grandfather or father as above. The east sub-population contained 54 samples and the west subpopulation contained 81. 40 samples were Turkish and 36 were Togolese.

Samples were typed for SNP markers defining major Hgs according to Rosser *et al.* (2000) and for 5 microsatellite loci DYS19, DYS390, DYS391, DYS392 and DYS393 by Dr. Emmiline Hill (Hill 2000).

For the present study samples were typed for an additional 9 Y-linked microsatellite markers, DYS434, DYS435, DYS436, DYS437, DYS438, DYS439, YA7.1, YA7.2 and GO4911 according to the methods in Chapter 2. Allele lengths for the locus DYS389b were obtained by subtracting the DYS389I allele length from the DYS389II allele length (Helgason *et al.* 2000).

4.2.3 Analysis

Fst pairwise population genetic distances and AMOVA for each X and Y population were calculated using ARLEQUIN 2.0 (Schneider *et al.* 2000). Mismatch distributions for the X and Y haplotypes in each population were created by converting allelic repeat numbers to binary characters, as before, which were then entered into ARLEQUIN 2.0.

Median Joining networks of the Irish, Turkish and Togolese X and Y samples were constructed using Network 3.111 (www.fluxus-engineering.com). Haplotypes were first analysed using the reduced median option and the *.rmf output file was used as an input file for the median-joining algorithm (Bandelt *et al.* 1999). Relative weights were applied to the loci on a 10-fold scale on the basis of the numbers of alleles at each locus. Higher weights were given to the less variable and thus slower-mutating loci (Hurles *et al.* 2002). For creation of the Y networks it was taken into account that these samples had been typed for SNPs by Hill (2000), which allowed the subdivision

of samples into haplogroups. Networks and mismatch distributions were created for the predominant haplogroup R1b3 in the Irish sample and the predominant haplogroup E3a in the Togolese sample.

4.2.4 Dating Estimates

Several methods were used to estimate expansion dates for the Irish and Togolese X and Y samples. BATWING (Wilson and Balding 2002) was used to create output distributions for various parameters including TMRCA and expansion time estimates, as described in Chapter 2. Irish and Togolese X data of Hill (2000) and the 16-locus Y haplotypes of the R1b3 Irish samples and E3a Togolese samples were analysed using BATWING. 10,000 iterations of the program were done and the first 1,000 were discarded as burn-in. The model used was exponential expansion from a constant-sized ancestral population. A generation time of 25 years was assumed. Prior distribution specifications for all populations were as follows: μ prior: $\text{gamma}(4,1482)$, α prior: $\text{normal}(0.1,0.1)$, N prior: $\text{gamma}(1,0.001)$ for Y and $\text{gamma}(3,0.001)$ for X, β prior: $\text{normal}(0,10)$. The prior values used were based on those used in example sessions of the programme and were chosen to aid convergence, as in Chapter 2.

In addition to this the coalescence times to the most recent common ancestor for R1b3 in Ireland and E3a in Togo were estimated using three different genetic distance measures. These were the average squared distance (ASD) measure (Goldstein 1995, Slatkin 1995), Goldstein's Variance Method (Goldstein *et al.* 1996) and the Bertranpetit and Calafell method (Bertranpetit and Calafell 1996). All of these methods rely on the diversity of microsatellite haplotypes that have arisen from one

ancestral chromosome and calculate the time taken to reach that diversity. In all cases a mutation rate of 0.21% per generation (Heyer *et al.* 1997) and a generation time of 25 years was assumed. The ancestral haplotype was determined by taking the modal haplotype at each locus for each lineage.

4.3 Results

4.3.1 X Haplotype Analysis

99 X haplotypes were found in total by Hill (2000). Of these ~60% were unique. The most frequently occurring haplotype occurred 48 times and was found in all populations. There is a greater proportion of shared haplotypes between Ireland and Turkey, suggesting a closer genetic affinity between these populations than between either of these and the Togolese.

Fst pairwise genetic distance measures (Table 4.3.1.1) indicate significant distance between both these populations and the Togolese but not between Ireland and Turkey. AMOVA analysis indicates that 29.5% of the variance is accounted for by variation between the Togolese and the other populations. 0% of the variance was accounted for by difference between the Irish and Turkish populations suggesting that for X haplotypes these populations are not structured.

4.3.2 Y Haplotype Analysis

147 Y haplotypes were found. These are shown in Appendix 4.1. Of these 90% are singletons. There was no haplotype sharing among populations, except the west and east of Ireland sub-populations.

Pairwise Fst genetic distance measures indicate that all populations are significantly different from one another (Table 4.3.2.1), including the west and east sub-populations of Ireland. The distances were correlated with geography, the greatest distance being found between the Irish and Togolese populations, followed by the

Turkish and Togolese populations. AMOVA analysis indicated a larger proportion of the genetic variance of the Y haplotypes (37.5%) was due to differences among populations than was the case for the X haplotypes, indicating that a stronger geographic signal is associated with the Y chromosome. AMOVA analysis indicated that 1.6% of the variance was accounted for differences among the east and west Irish Y sub-populations.

	West	East	Turkey	Togo
West	0			
East	0	0		
Turkey	0	0	0	
Togo	0.25	0.31	0.19	0

Table 4.3.1.1: X chromosome pairwise population genetic distances among the populations

	West	East	Turkey	Togo
West	-			
East	-	-		
Turkey	-	-	-	
Togo	+	+	+	-

Table 4.3.1.2: Matrix of significant P-values for X Fst distances among the populations

	West	East	Turkey	Togo
West	0			
East	0.0341	0		
Turkey	0.37425	0.25547	0	
Togo	0.58351	0.47442	0.31931	0

Table 4.3.2.1: Y chromosome pairwise population genetic distances among the populations

	West	East	Turkey	Togo
West	-			
East	+	-		
Turkey	+	+	-	
Togo	+	+	+	-

Table 4.3.2.2: matrix of significant P-values for Y Fst distances among the populations

4.3.3 Mismatch distributions

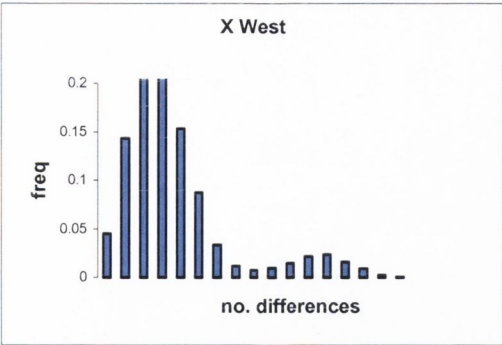


Figure 4.3.3.1: X West Mismatch Distribution. Mean = 3.7, Variance = 9.5, Mode = 2

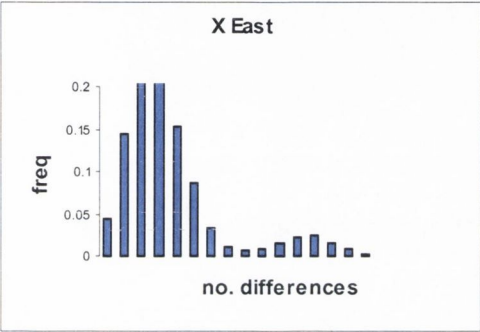


Figure 4.3.3.2: X East Mismatch Distribution. Mean = 3.2, Variance = 9.2, Mode = 3

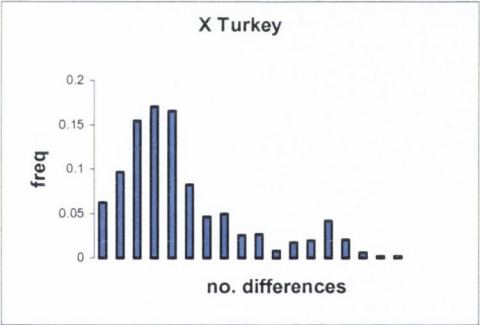


Figure 4.3.3.3: X Turkey Mismatch Distribution. Mean = 4.6, Variance = 13.5, Mode = 2



Figure 4.3.3.4: X Togolese Mismatch Distributio. Mean = 7.2, Variance = 12.9, Mode = 9

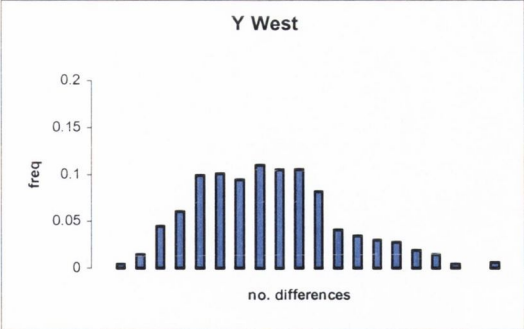


Figure 4.3.3.4: Y West Mismatch Distribution. Mean = 10.7, Variance = 17.7, Mode = 8

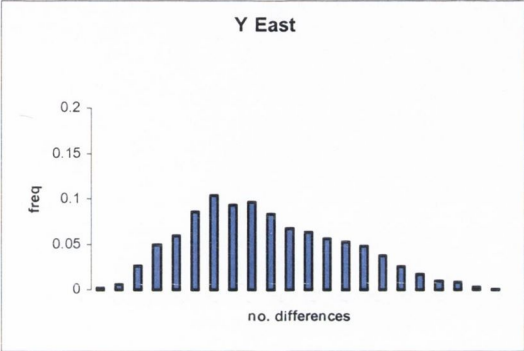


Figure 4.3.3.5: Y East Mismatch Distribution. Mean = 8.8, Variance = 16.7, Mode = 6

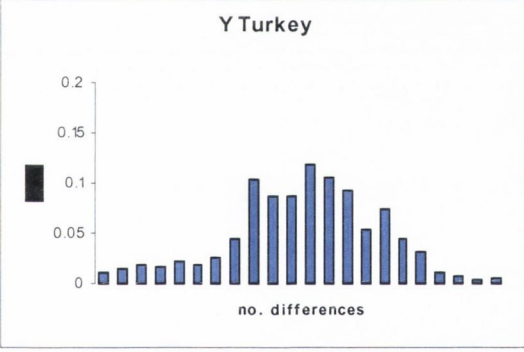


Figure 4.3.3.6: Y Turkey Mismatch Distribution. Mean = 10.7, Variance = 16.0, Mode = 12

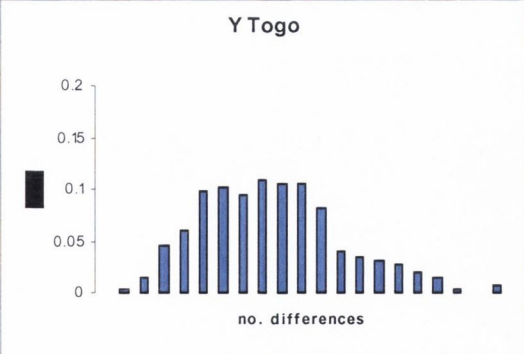


Figure 4.3.3.7: Y Togolese Mismatch Distribution. Mean = 8.5, Variance = 16.0. Mode = 12

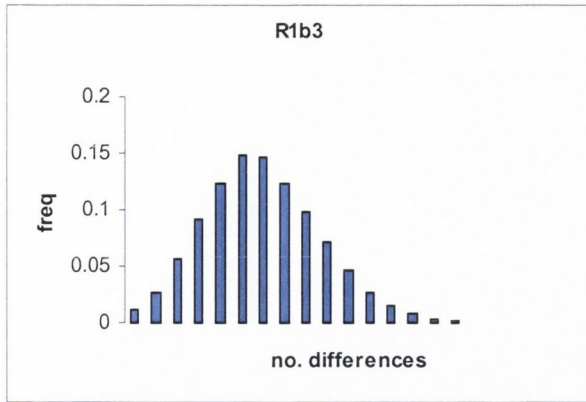


Figure 4.3.3.8: Irish R1b3 Mismatch Distribution. Mean = 6.0, Variance = 7.4, Mode = 5

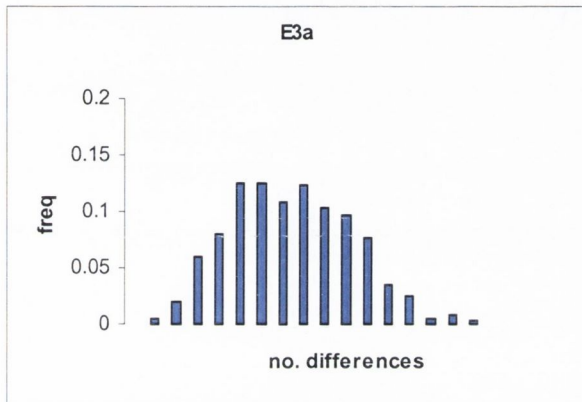


Figure 4.3.3.9: Togolese E3a Mismatch Distribution. Mean = 7.4, Variance = 8.3, Mode = 5

Mismatch distributions for X and Y haplotypes in each population were plotted to compare phylogenetic diversity in the samples. The shape and properties of such distributions contain a record of the ancient demographic events of a population. In particular, they have been used to infer population expansions (Sherry *et al.* 1994, Harpending 1994). Mismatch distributions of the X and Y haplotypes for each population are shown in Figures 4.3.3.1 – 4.3.3.9. In the X distributions the Togolese population shows the greatest diversity and highest modal value, suggesting that it is the oldest X population. The Turkish X distribution is intermediate between the Irish and Togolese in terms of diversity and modal value, but closer to the values of the Irish X distribution. The Irish X population is the least diverse and has the lowest modal value, indicating that it is the youngest population. All X populations have

bimodal tails. These are the result of the 152 bp allele at F7 and the 105 bp allele at F13, which represent a jump of several mutational steps from the other alleles in the populations and may indicate the presence of more than one lineage. Alternatively they could be due to a departure from the stepwise mutation model. When haplotypes containing these alleles are removed resulting distributions are unimodal, consistent with population expansion (Hill 2000).

Mismatch distributions for the Y samples show smooth unimodal curves for each population, consistent with expansion. In contrast with its X distribution, the Togolese Y distribution does not show the greatest diversity and modal value. These are found in the Turkish sample, indicating that this Y population is the oldest and most diverse. The Togolese sample has a similar diversity and modal value to the Irish sample, suggesting expansion at a similar time in these populations.

4.3.4 Median-joining networks

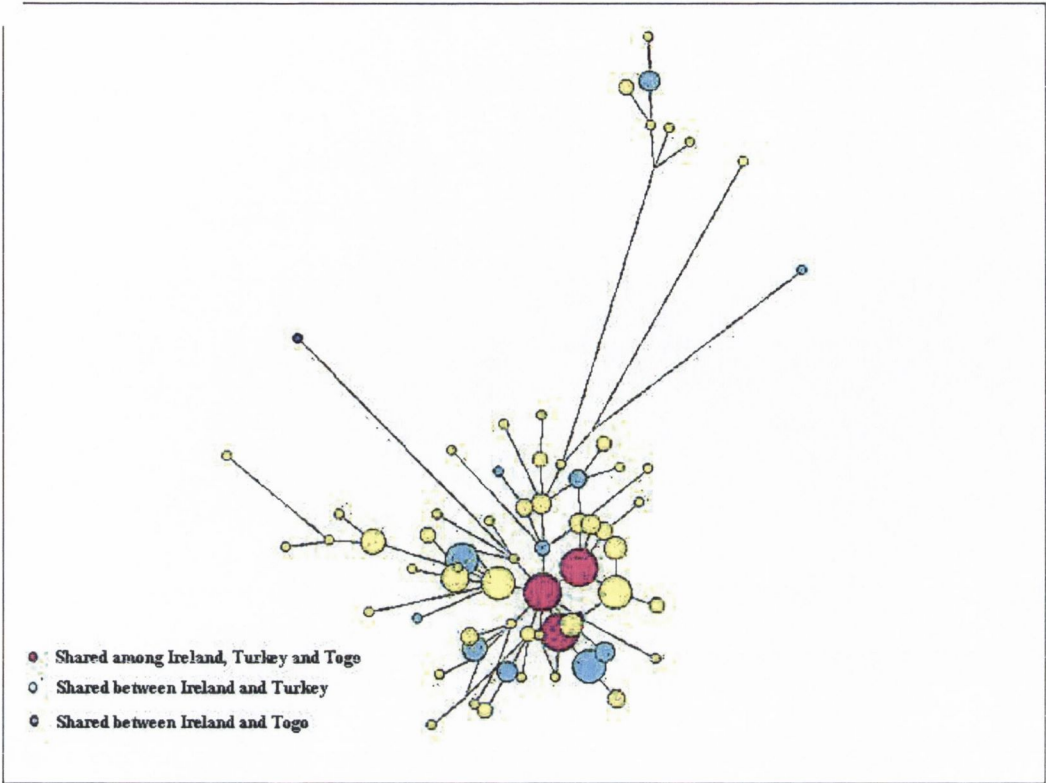


Figure 4.3.4.1: Median-joining network of Irish X chromosomes.

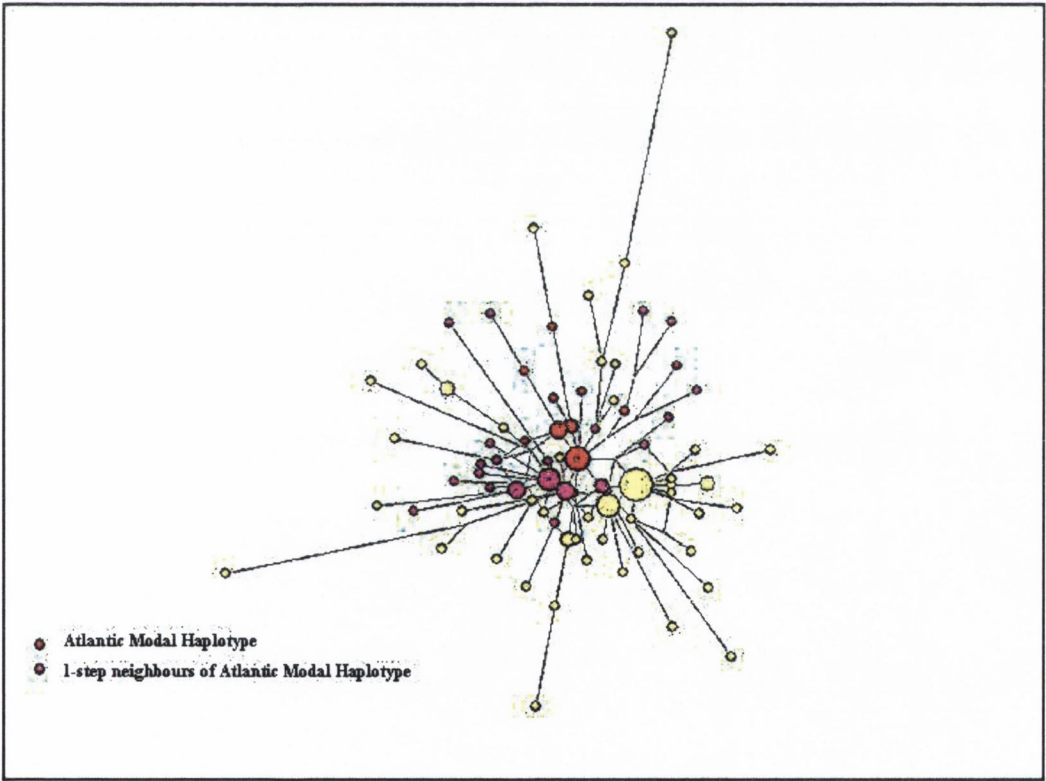


Figure 4.3.4.2: Median-joining network of Irish R1b3 Y chromosomes

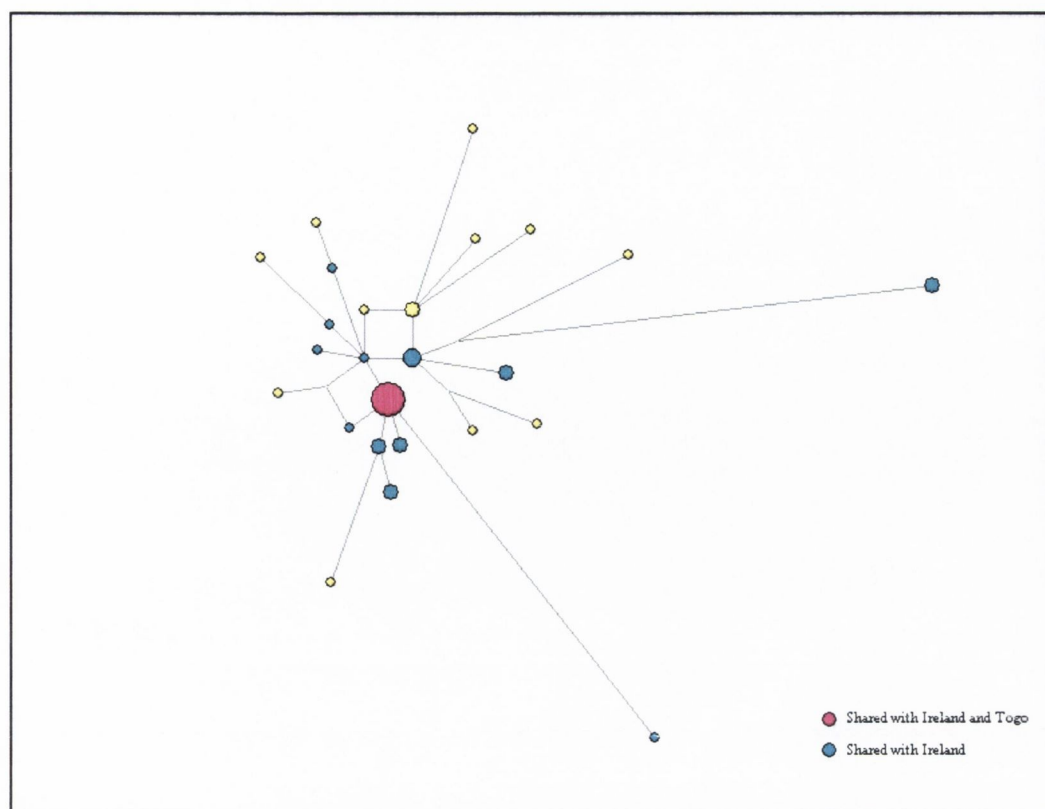


Figure 4.3.4.3: Median-joining network of Turkish X Chromosomes

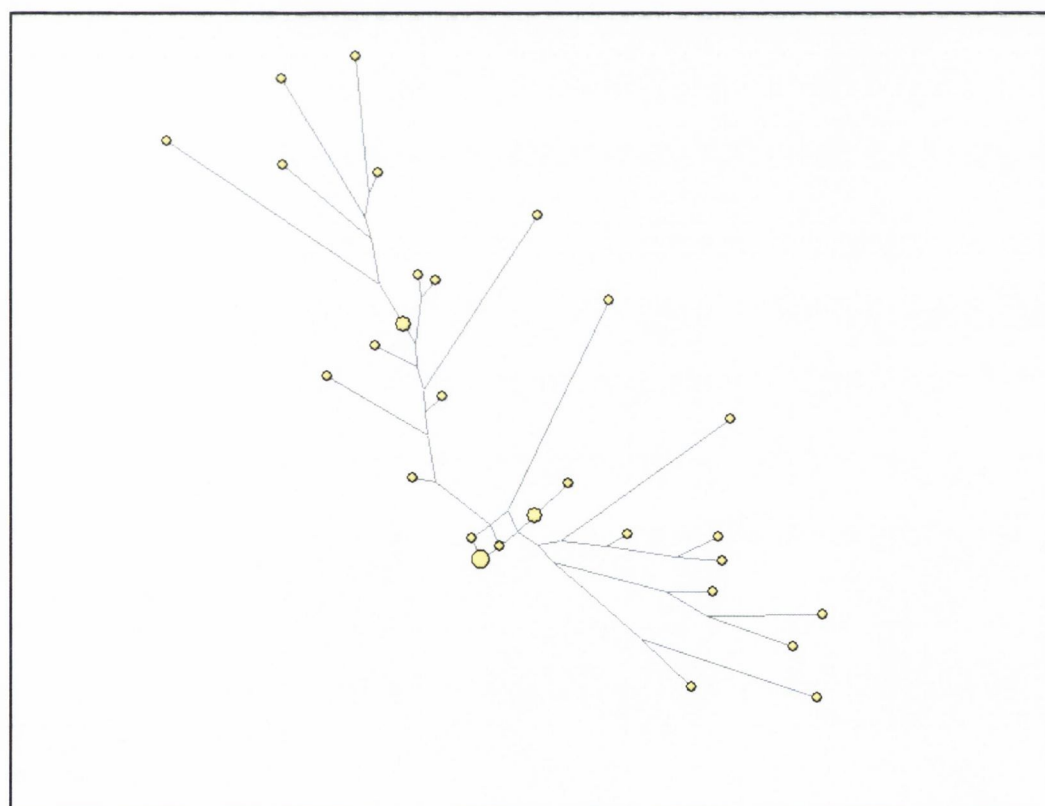


Figure 4.3.4.4: Median-joining network of Turkish Y chromosomes (all lineages).

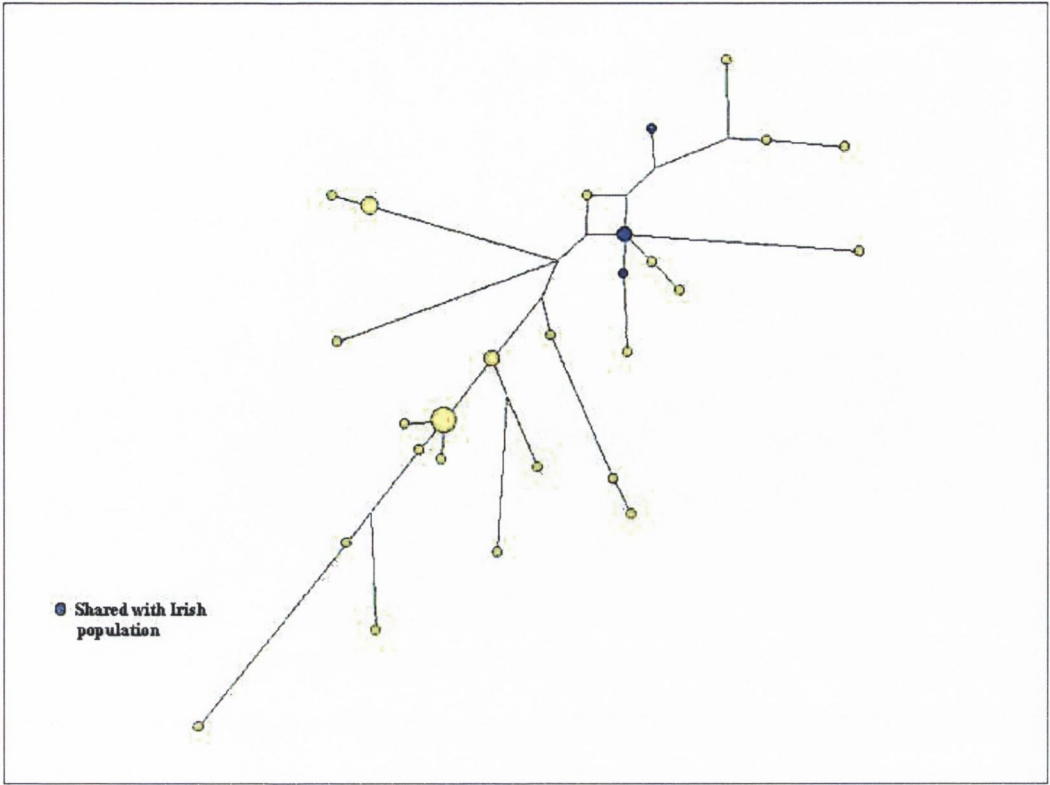


Figure 4.3.4.5: Median-joining network of Togolese X chromosomes

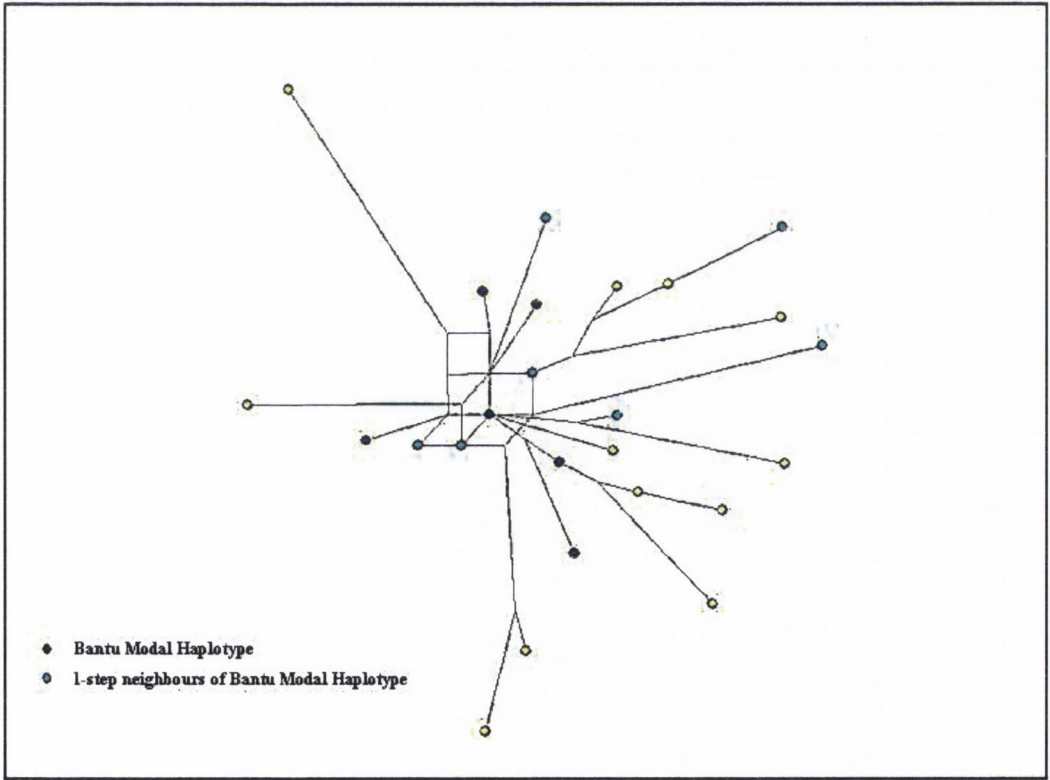


Figure 4.3.4.6: Median-joining network of Togolese Y chromosomes. Bantu Modal Haplotype chromosomes (Thomas et al. 2000, Periera et al. 2002) are indicated, discussed in Section 4.4.4

4.3.5 Irish X and Y Networks

Median-joining networks of the Irish, Turkish and Togolese X and Y samples are shown in Figures 4.3.4.1 - 4.3.4.6. In the Irish X network (Figure 4.3.4.1) the most common haplotypes occupy a central position. The less common haplotypes radiate from the centre in a star-like pattern, which is consistent with population expansion. The most common central haplotypes are also those that are shared with Turkey and Togo. Those shared between Ireland and Turkey and between Ireland and Togo occupy more peripheral positions on the network.

SNP typing indicates that there is one predominant lineage in the Irish sample, R1b3 (frequency 83.7%). A median-Joining network of the samples belonging to the Irish R1b3 lineage is shown in Figure 4.3.4.2. The topology of this network is similar to the X network, showing star-like expansion from a common central node.

4.3.6 Turkish X and Y median-joining networks

The Turkish X and Y networks are shown in Figures 4.3.4.3 and 4.3.4.4. The X network, shows similar levels of diversity to the Irish X network, though the haplotypes are somewhat less densely packed. The Turkish X network shows a common central node, which is shared with the Irish and Togolese populations. The Turkish Y lineage is substantially more diverse than the Irish R1b3 network. This is likely to be due to the greater number of haplogroups represented in the Turkish Y sample (the Irish sample is composed predominantly of one haplogroup, R1b3). Again this is consistent with the Turkish Y mismatch distribution picture of an older and more diverse Turkish Y population.

4.3.7 Togolese X and Y median-joining networks

The Togolese X network topology (Figure 4.3.4.5) is quite different to the Irish and the Turkish. The haplotypes are less densely packed and there is no obvious central node. It is substantially more diverse than either of the other X networks. This is consistent with the Togolese X population being the more ancient and diverse of the three. However the network of the predominant Togolese Y lineage, E3a (frequency ~89%), is more star-like than both the Togolese X network and the Turkish Y network, suggesting a Togolese Y population that is younger than both the Togolese X population and the Turkish Y population. It also suggests that the Togolese Y population has recently undergone expansion (Figure 4.3.4.6).

4.3.8 Dating expansion in the Irish and Togolese populations

Population	BATWING beta	95% C.I.
Irish X	880 YBP	100 – 1,700 YBP
Irish R1b3	5,300 YBP	2,300 – 10,200 YBP
Togolese X	2,000 YBP	200 – 8,000 YBP
Togolese E3a	8,300 YBP	4,300 – 12,200 YBP

Table 4.3.8.1 BATWING expansion date estimates for Irish and Togolese X and Y samples

BATWING (Wilson and Balding 2002) was used to estimate expansion dates in the Irish and Togolese X and Y populations (Table 4.3.8.1 above). Expansion in the Irish X population was estimated at 880 years ago (95% Credible Intervals 100 - 1,700 years ago). Expansion in the Irish R1b3 lineage estimated at 5,300 years ago (95% credible intervals 2,300 – 10,200 years ago). This is in good agreement with the expansion date obtained for the large R1b3 dataset in Chapter 2 (~5,100 YBP).

Using the same method expansion in the Togolese X sample was estimated at 2,000 years ago (95% Credible Intervals of 200 – 8,000 years ago). Expansion in the

Togolese Y E3a lineage was estimated at 8,300 years ago (2.5% and 97.5% credible intervals 4,300 – 12,200 years ago).

Population	Method	TMRCa	95% C.I.
Ireland R1b3	ASD	3,850	1,650 – 13,450
Ireland R1b3	Goldstein's Variance	3,900	1,500 – 19,300
Ireland R1b3	Bertranpetit and Calafell	3,250	1,400 – 11,400
Togo E3a	ASD	5,400	2,300 – 18,900
Togo E3a	Goldstein's Variance	4,600	2,750 – 23,400
Togo E3a	Bertranpetit and Calafell	5,350	2,300 – 18,750

Table 4.3.8.2 **TMRCa estimates for Irish R1b3 and Togolese E3a samples using various methods**

Other methods used to estimate the ages of the Irish and Togolese Y populations assess the amount of microsatellite diversity in the samples (Table 4.3.8.2 above). R1b3 coalescence in the Irish sample was estimated at 3,850 years ago, (95% confidence limits 1,650 -13,450 years ago) according to the ASD method, 3,900 years (95% confidence limits 1,500 – 19,300 years ago) according to Goldstein's Variance Method and 3,250 years ago (95% confidence limits 1,400 – 11,400 years ago) according to the Bertranpetit and Calafell method.

The same methods were used to date E3a coalescence in the Togolese sample at 5,400 years ago (95% limits 2,300 – 18,900 years ago), 4,600 years ago (95% limits 2,750 – 23,400 years ago) and 5,350 years ago (95% limits 2,300 – 18,750 years ago) respectively.

4.4 Discussion

4.4.1 Comparison of X haplotypes among the populations

The Togolese X sample shows the greatest diversity of the populations examined, consistent with a long human history in sub-Saharan Africa. The Togolese X sample is also significantly different from the Irish and Turkish X populations, whereas the Irish and Turkish X populations are not significantly different from one another. The Irish and Turkish X mismatch distributions show similar levels of diversity. These findings suggest an Out of Africa bottleneck effect, whereby the Togolese X population is the oldest and most diverse and there is greatly reduced genetic diversity in all non-African populations. This is likely due to the founding of all non-African populations by a small, genetically homogeneous group, as has been indicated by other genetic systems, including Y chromosome and mtDNA haplotypes.

The relative lack of diversity found in the Irish X sample is consistent with the late settlement of this peripheral region of Europe, however the star-like topology of Irish X network is consistent with recent expansion in this population.

BATWING expansion date estimates for the Irish and Togolese X populations suggest recent expansions in these populations, ~880 YBP in Ireland and ~2,000 YBP in Togo.

4.4.2 Y haplotype comparisons among the populations

Comparison of Y haplotype diversity levels among the 3 populations indicates that it is Turkey which has the oldest and most diverse Y population. This is suggested by a comparison of the Y mismatch distributions and networks, and by the greater diversity of haplogroups in the Turkish population. Both the Irish and the Togolese Y samples were dominated by single Haplogroups. Networks of the predominant E3a Hg in the Togolese sample and the predominant R1b3 Hg in Ireland have a star-like topology, consistent with population expansion. This suggests that the relative homogeneity in the Y chromosomes of these samples is due to demic expansion of these predominant Y lineages in the Irish and Togolese populations.

4.4.3 Bantu dispersals and the spread of agriculture

The time scales for the estimates of expansion in the Irish and Togolese Y populations suggest that expansion in these populations may have been due to the onset of agriculture. Linguistic, archaeological and genetic information from other African populations suggest that this reduction in diversity in the Togolese Y sample may be due to Bantu dispersals. The great majority of sub-Saharan Africans (about 200 million people) speak one of about 400 Bantu languages (Salas *et al.* 2002). This family of languages belongs to the Niger-Congo phylum and is young from an evolutionary point of view, its origins dating to between 3,000 – 5,000 years ago. The highest level of diversity of the Bantu languages is found in eastern Nigeria and western Cameroon. The nearest neighbours of Bantu languages are also from this region of West Africa (Greenberg 1963). Linguists have argued that the most likely explanation for the widespread distribution of such a young language family is through enormous expansion of its speakers. Archaeological evidence suggests that

such migrations may have been linked with the origin of agriculture, and later with the discovery of iron smelting technology. The linguistic evidence for expansions from a homeland in West Africa is in good geographical and temporal agreement with a farming culture dated to between 1600 and 700 BC and with an iron smelting culture at Nok, near the confluence of the Niger and Benue rivers, dated to about 500 BC (Cavalli-Sforza *et al.* 1994).

4.4.4 The Bantu Modal Haplotype

Togo is located near the proposed Bantu homeland and the predominant lineage in the Togolese sample, E3a, has been proposed as a marker of Bantu expansions (Thomas *et al.* 2000). Thomas *et al.* (2000) proposed a Bantu 5-microsatellite haplotype and this was found by Pereira *et al.* (2002) to be the main haplotype in regions that were in the path of the Bantu expansions. This haplotype (15-21-10-11-14 at loci DYS19-DYS390-DYS391-DYS392-DYS393) was found to be the most common in the Togolese sample (18.75%) and it and its one-step neighbours (~50% of the Togolese sample) occupy central positions on the Togolese E3a network (Figure 4.3.4.4), with two- and three- step neighbours occupying more peripheral positions. The frequency of this haplotype and its 5-marker one step neighbours in the Togolese sample is comparable to that found in other populations in the path of the Bantu dispersal (Pereira *et al.* 2002). The BATWING estimate of expansion in the Togolese E3a sample 8,130 years ago predates the Bantu expansion, however the 95% credible intervals from 4,335 to 12,205 encompass the time range for the Bantu dispersal. The other methods used to estimate Togolese E3a coalescence times were more consistent with one another (between 4,624 and 5,393 years ago, although confidence limits are large) and in good agreement with the time frame for the Bantu expansions (3,000 –

5,000 years ago). A substantial genetic contribution of Bantu farmers to the present Togolese population may explain the signs of recent expansion and relative lack of diversity in the Togolese Y sample.

4.4.5 The onset of agriculture in Ireland

The main lineage in the Irish Y sample, R1b3, has been present in Europe since the Palaeolithic and its frequency declines from east to west, being highest in the Atlantic populations. The present distribution of R1b3 in Europe has been interpreted as a signature of this postglacial re-expansion (Semino *et al.* 2000). The Irish R1b3 sample shows signs of recent expansion. The central haplotype in the R1b3 network occurs at high frequency and is equivalent to the Atlantic modal haplotype proposed by Wilson *et al.* (2001b) for 5 of the 16 microsatellite loci typed. This haplotype was found at high frequency in R1b3 samples in the British Isles and is also modal in Basques. Similarity of Basque and 'Celtic' Y chromosomes has been used to argue for genetic continuity along the Atlantic coast of Europe since the Upper Palaeolithic (Wilson *et al.* 2001b). R1b3 itself was more common in the west of Ireland sample (91%) than in the east (72%), consistent with the findings of Chapter 2.

The BATWING median output estimate for expansion in the Irish R1b3 sample is 5,300 years ago, consistent with the dates obtained for R1b3 expansion in the larger Irish sample in Chapter 2 (5,100 YBP). Other methods used gave dates from 3,850 to 5,350 years ago, with large confidence limits. This expansion date for the Y chromosome is consistent with archaeological evidence for the intensification of agriculture and an increasingly complex society. There is archaeological evidence for significant forest clearance, domestic animal remains, seeds, pottery, more permanent

dwellings, increased territoriality and the appearance of megalithic tombs during this period (Waddell 2000).

The large uncertainty associated with the dating estimates means that, as with the larger Irish R1b3 sample, it is possible that the signature of expansion seen in the Irish Y sample may have been due to numerous other population genetic events, including the original colonisation of the island. Likewise the large uncertainty associated with the Togolese dating estimates indicate that other scenarios besides the Bantu dispersals may have been responsible for this expansion signal.

4.4.6 Discrepancy between the X and Y haplotypes

The comparison of X and Y haplotypes in the same populations indicates discrepancies in the diversity patterns of these loci. In comparison to the X chromosome, Y shows greater between-population differences. There is haplotype sharing among the Irish, Turkish and Togolese X populations, with some haplotypes being found in all populations. No Y haplotypes were shared among Ireland, Turkey and Togo. The number of unique haplotypes was much greater for Y (90%) than for X (60%).

Fst pairwise genetic distance measures indicate that while significant distance can be detected between the Togolese X chromosomes and the other populations, the Irish and Turkish X populations were not significantly different from one another. In contrast there was significant genetic distance between all of the Y populations examined, and genetic structure was detected between the east and west Y sub-populations of Ireland. AMOVA analysis reinforced this picture, with a greater

portion of the genetic diversity being due to between-population differences on the Y than on the X. These findings are consistent with a stronger geographic signal associated with the Y chromosome.

Estimates of expansion times were substantially younger for the X samples than for Y samples in the same populations. This may be due to the confounding effects of recombination on the X chromosome, or to a higher mutation rate in microsatellites on the X chromosome.

This study of X and Y diversity levels in three populations has indicated that X and Y haplotypes record population history in different ways. This may be partly due to the larger effective population size of the X chromosome than the Y, which serves to accentuate the effects of drift on the Y, and to the effects of recombination on the X chromosome. However it is also likely to be due to the way behavioural differences in the sexes affect demographic history and genetic diversity levels. In particular it seems that population expansion events, such as those postulated to accompany the spread of agriculture, may affect X and Y chromosome diversity differently. These events appear to wipe out pre-existing variation on the Y chromosome, but not on the X chromosome. Such patterns are probably due to the higher levels of variance in male than female reproductive success, higher rates of polygamy than polyandry and possibly higher rates of patrilocality, all of which act to decrease the effective population size of the Y chromosome population.

This effect is most clearly suggested in the Togolese population. While Togolese X chromosomes are the most ancient and diverse of the three populations, there is a

strong suggestion of a recent population expansion that has reduced diversity dramatically on the Y chromosome. This suggests that the Bantu migrations and the demic effects of the spread of agriculture affected males and females differently in the Togloese population.

Chapter 5: A Genetic Investigation of Newfoundland Origins

63 male samples from Newfoundland were genotyped for 3 Y-linked SNP and 6 Y-linked microsatellite loci in order to examine the genetic ancestry of this population. Since historical evidence suggests the main colonizing populations were from Ireland and England, the resulting Newfoundland haplotypes were compared with data from Irish and English populations in order to determine the relative contributions of these source populations to the Newfoundland Y chromosome population. The similarity of Y haplotypes among the three populations suggests that Y chromosome variation in Newfoundland is due to admixture of colonizing populations, which came mainly from southeast Ireland and England. However the possibility of historical admixture between the southeast Irish and English source populations, attested to by the similarity of the Y chromosomes of these populations, makes it difficult to unequivocally distinguish the contributions of the source Irish and English populations to the present-day Newfoundland Y chromosome population.

5.1 Introduction

5.1.1 The history of Newfoundland

Newfoundland was one of the first areas of North America to be exploited regularly by Europeans. Shortly after Cabot's rediscovery in 1497, Portugese, Basques, Atlantic French, Bretons and Normans arrived annually to fish for cod. After 1575 they were joined by the English. By 1630 the English had displaced the continental Europeans and had begun to establish a year round settlement. Beginning around 1675, ship owners and ship masters from England's West Country began to recruit servants in southwest Ireland. Irish migration was closely linked to the provisions trade. English ships departed from Devon and called to Waterford and other southern Irish ports to stock up on salt provisions for the summer's fishery, and began to recruit Irish servants also. When later Irishmen became established as planters, they recruited servants almost exclusively from their homeland in southeast Ireland. Irish migrants to Newfoundland originated from a small geographic region. Waterford was the primary port of embarkation and the migrants came from places within a day's journey of the city. Migrants came mainly from 5 counties: Waterford, Wexford, Carlow, Kilkenny and parts of Tipperary. A chain of migration initiated by the English and continued by Irish planters, ship owners and merchants ensured this pattern of ethnic Irish succession. No other province in Canada or state in America drew such an overwhelming proportion of their immigrants from so geographically compact an area in Ireland for so prolonged a period in time (Mannion 1980, 1999).

5.1.2 The genetic history of Newfoundland

The historical information regarding a large Irish exodus to the Newfoundland combined with the distinctive Irish accent indigenous to some regions of Newfoundland suggests a strong Irish influence on Newfoundland ancestry. This study investigates the relative Irish and English genetic contribution to the Newfoundland population by comparing Newfoundland Y chromosome haplotypes to those from England and Ireland, the presumed main source populations.

5.2. Materials and Methods

5.2.1 Genotyping

DNA samples were obtained from 63 Newfoundland males with appropriate informed consent. They were collected by Dr. Jane Green and Ms. Sandra Adams of the Patient Research Centre, Health Sciences Centre, St. John's, Newfoundland. Samples were obtained from volunteers of well-established Irish, English or French ethnic origin. Since samples were chosen for their well-documented ethnicity and established history in Newfoundland, they do not represent a random sample of the Newfoundland population. For some analysis all Newfoundland Y chromosomes were categorized together to give an approximation of the total Newfoundland Y population, but results from such analysis must be interpreted with caution. 29 samples were of English ethnic origin, 26 of Irish ethnic origin and 8 of French ethnic origin. Samples were assayed for 6 Y-linked microsatellite loci: DYS19, DYS388, DYS390, DYS391, DYS392 and DYS393 by Shane Cronin using the MS1 multiplex outlined in Chapter 2. Samples were also assayed for the Y-linked SNPs indicated in Chapter 2.

5.2.2 Admixture Estimates

The proportions of English and Irish ancestry, respectively, in the Newfoundland population were estimated using the m_R estimator according to Helgason *et al.* (2000). This is based on Bernstein's model (1931) where allele frequencies in an admixed population are presumed to be a linear combination of the allele frequencies in the parental populations, and the effects of drift and mutation are assumed to be negligible.

These conditions should hold given the recent time frame of the colonization of Newfoundland. Allele frequencies in the admixed population are given by $P_h = P_1m + P_2(1 - m)$, where P_h is the admixed population, P_1 and P_2 are the source populations and m is the proportion of P_1 ancestry in the admixed population. This admixture estimator, m_r , was approximated by fitting a regression line through the points defined by $P_1 - P_2$ and $P_h - P_2$. The slope of this line approximates the ancestry of P_1 in the admixed population (Roberts and Hiorns 1965, see Helgason *et al.* 2000).

5.2.3 Irish and English source populations

The Irish Y chromosome data from Hill *et al.* (2000), typed for the additional marker DYS388, was used as the Irish source population. The Y chromosome data from the English towns presented in Weale *et al.* (2002) was used as the English source population. Data were classified according to haplogroup. Samples from the R1b3 haplogroup were further subdivided into 3 classes based on microsatellite haplotype: those possessing the Atlantic Modal Haplotype, 14-12-24-11-13-13 (Wilson *et al.* 2001b), those possessing the Irish modal haplotype, 14-12-25-11-14-13 (Chapter 2) and all other R1b3 samples.

Since historical records suggest that Irish Newfoundland settlers had a mainly southeast Irish origin, the same admixture estimates were performed using source populations of samples from this region. 71 samples from southeast Ireland (from counties Wexford, Carlow, Kilkenny and Waterford, as presented in Chapter 2) were used as the southeast Irish regional source population.

The relationship of the Newfoundland samples to populations from regions of Ireland, Britain and Europe were also examined by principal components analysis, plotted using POPSTR, as outlined in Chapter 2.

5.3 Results

5.3.1 Haplotype frequency and sharing

Newfoundland haplotypes are listed in Appendix 5.1. 79.4% of Newfoundland samples belonged to haplogroup R1b3. 20.6% belonged to Haplogroup I. Of these, one chromosome, accounting for 1.6% of the sample, belonged to the subgroup I1b2. When the Newfoundland samples were split into their ethnic subpopulations, 82.6% of the Newfoundland-English samples belonged to haplogroup R1b3 and 17.4% belonged to group IxI1b2. 73.1% of the Newfoundland-Irish origin samples belonged to group R1b3, 23.1% belonged to IxI1b2 and 3.8% belonged to I1b2. Of the Newfoundland-French population, 87.5% belonged to R1b3 and 12.5% belonged to IxI1b2.

The most common microsatellite haplotype in the Newfoundland population is 14-12-24-10-13-13, accounting for 17.4% of the sample. This haplotype is also found in the English, Irish and Southeast Irish populations (at frequencies of 9%, 11.9% and 10.8% respectively). It is the most common haplotype in the Newfoundland-Irish (24.1%) and Newfoundland English (13%) populations. It occurs in the Newfoundland-French population at a frequency of 13%. The second most common Newfoundland haplotype is Atlantic Modal Haplotype, (8.7%).

5.3.2 Admixture Estimates

The proportion of Irish ancestry in the Newfoundland population, given source populations from Ireland (Hill *et al.* 2000) and England (Weale *et al.* 2002) is estimated at 0.63 (Figure 5.3.2.1). When the admixed population includes data from the

Newfoundland Irish population only, this estimate drops to 0.54 (Figure 5.3.2.2). The proportion of Irish ancestry in the Newfoundland English sample, estimated by the same method, is 0.55 (Figure 5.3.2.3). When the source population used for Ireland comes from the southeast of Ireland only (using the southeast population of Chapter 2) the proportion of Irish ancestry estimated in the Newfoundland Irish sample is 0.41 (Figure 5.3.2.4). The proportion of southeast Irish ancestry estimated in the Newfoundland English sample is 0.44 (Figure 5.3.2.5).

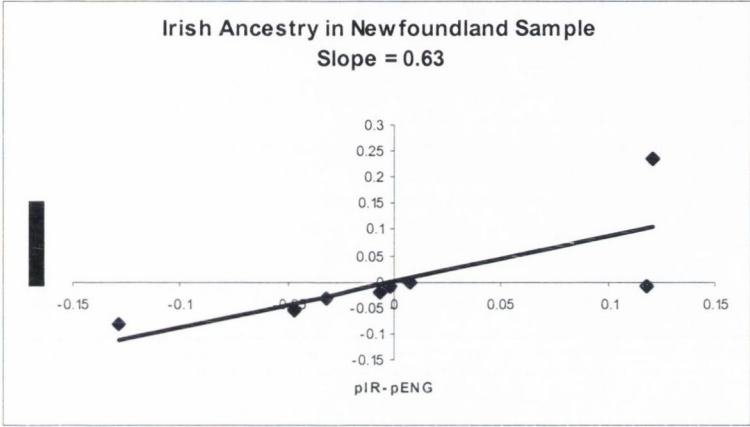


Figure 5.3.2.1: The relationship between p_{Ir} , p_{Eng} and p_{Nf} , the haplotype frequencies for Ireland, England and Newfoundland respectively. mR , the estimate of Irish ancestry in the Newfoundland sample, is approximated by the slope of the least-squares regression line.

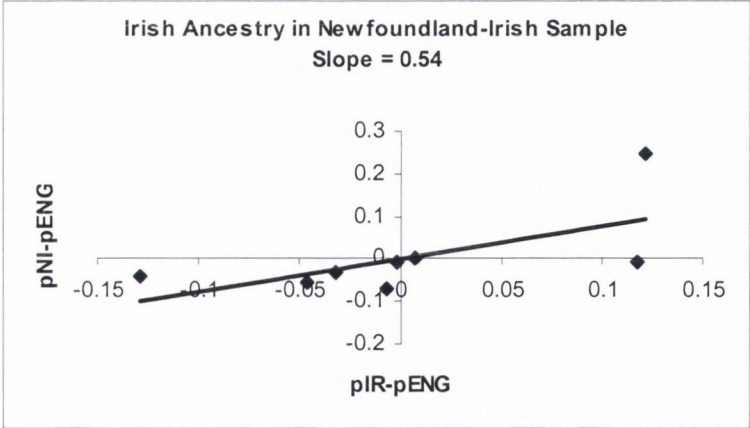


Figure 5.3.2.2: The slope of the least-squares regression line approximates Irish ancestry in the Newfoundland-Irish sample.

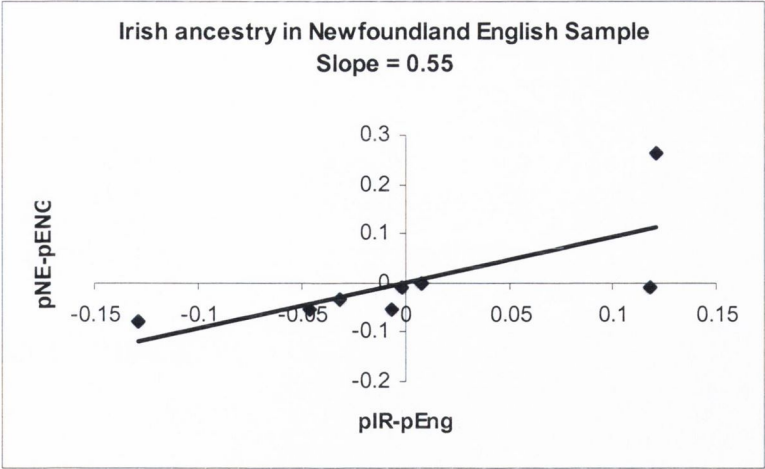


Figure 5.3.2.3. The slope of the least squares regression line approximates the proportion of Irish ancestry in the Newfoundland-English sample.

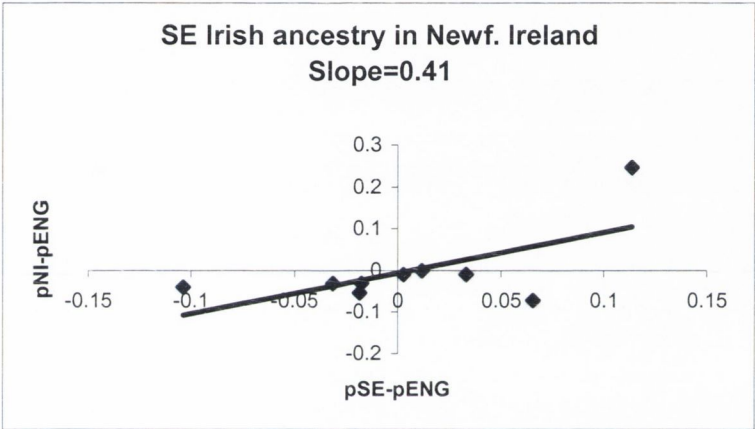


Figure 5.3.2.4. The slope of the least squares regression line approximates the proportion of Southeast Irish ancestry in the Newfoundland-Irish sample.

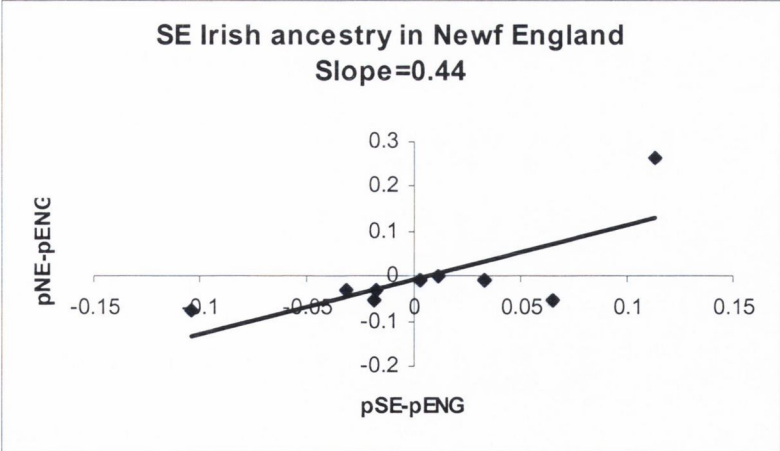


Figure 5.3.2.5 The slope of the least squares regression line approximates the proportion of Southeast Irish ancestry in the Newfoundland-English sample.

Admixed Popn.	Source Popn. 1	Source Popn. 2	Admixture Estimate (proportion. of Popn. 1)
Newfoundland	Ireland	England	0.63
Newfoundland Ireland	Ireland	England	0.54
Newfoundland England	Ireland	England	0.55
Newfoundland Ireland	Southeast Ireland	England	0.41
Newfoundland England	Southeast Ireland	England	0.44

Table 5.3.2.1 Admixture Estimates for Newfoundland Samples

5.3.3 Principal Components Analysis

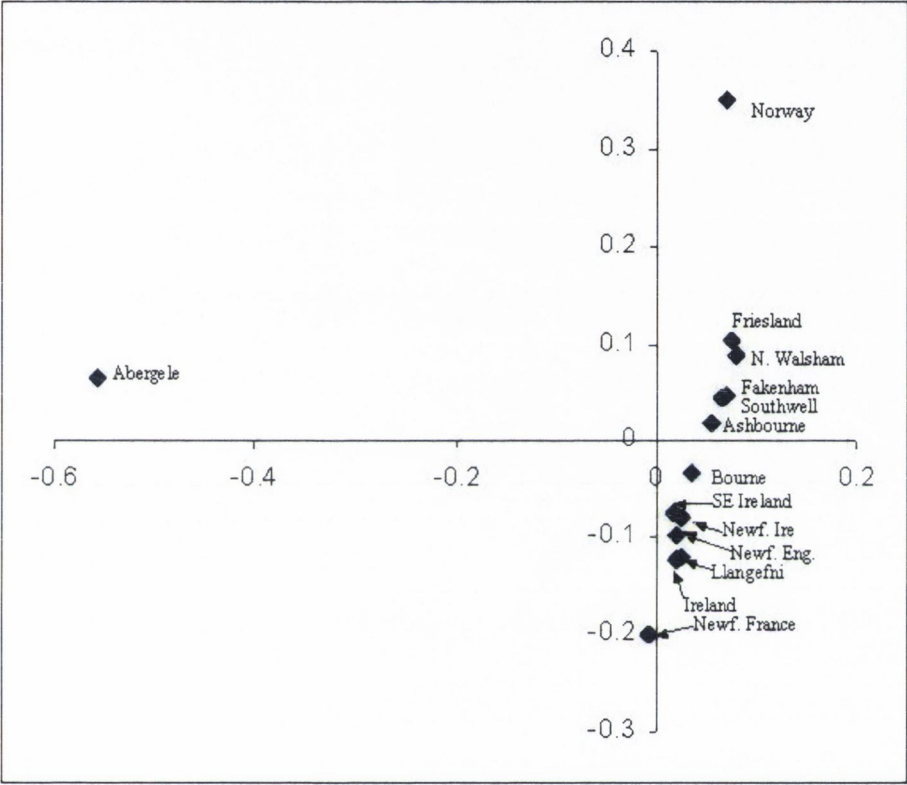


Figure 5.3.3.1: Principal Components plot of PC1 vs. PC2. PC1 explains 42.3% of the variance and PC2 explains 29.6% of the variance. Samples were classified according to haplogroup, except R1b3 samples which were classified into 3 groups, as before.

Principal Components Analysis (Figure 5.3.3.1) summarises the variation in Y chromosome frequencies and illustrates the relationships of the Newfoundland subpopulations to one another, to two Irish populations (the southeast as described above, and the general Irish sample), and to the British and European data of Weale *et al.* (2002).

5.4 Discussion

R1b3 accounts for ~84% of Irish Y chromosomes (Hill *et al.* 2000), and ~60% of the English sample of Weale *et al.* (2002). The intermediate frequency of R1b3 in the total Newfoundland sample, ~79%, is consistent with admixture from Irish and English source populations, but with perhaps a greater Irish contribution.

The two most common R1b3 haplotypes in the Newfoundland sample, the Atlantic Modal Haplotype and 14-12-24-10-13-13, are found at high frequency in the Irish and English sample, consistent with these being the main source populations for Newfoundland Y chromosomes. Surprisingly, R1b3 is found at lower frequency in the Newfoundland-Irish sample than in the total Newfoundland sample (~73%, versus ~79%). This may be because the geographic region of the Irish source population, the southeast, is also the region of Ireland with the lowest frequency of R1b3 chromosomes (~75%, compared with ~84% in the island as a whole), likely a consequence of substantial colonization of the southeast by Vikings, Anglo- and Cambro-Normans in the late middle ages, and British settlement in the 16th and 17th centuries as discussed in Chapter 2.

The m_R estimator provides a quantitative estimate of Irish ancestry in the Newfoundland sample at 63%, but when samples from the Newfoundland-Irish population only are used this proportion drops to 54%. This is likely to be caused by the relatively low R1b3

frequency in the Newfoundland Irish population. The low frequency of R1b3 in the Newfoundland-Irish population is mirrored in the relatively low frequency of R1b3 in the southeast of Ireland, the geographic area from where most Irish migrants are reported to have come. However, when a sample from the Southeast of Ireland is used as the Irish source population, the estimated proportion of Southeast Irish ancestry in the Newfoundland Irish sample drops to 41%, indicating that the Newfoundland sample more closely resembles the general Irish sample than the Southeast sample. The estimated Irish sample in the Newfoundland English sample when the southeast Irish source population is used is 44%.

In the plot of PC1 vs. PC2 (Figure 5.3.3.1) the Irish sub-populations are skewed towards the bottom right of the plot whereas most British populations occupy a more central position, between the Irish and Mainland European populations. The intermediate position of the Newfoundland populations is illustrated in this plot: they occupy a position between the Irish and British populations, consistent with their being admixed from Irish and British source populations. However, they are positioned more closely to the Irish populations, consistent with a closer genetic relationship to these populations. In particular, the Southeast Ireland population clusters with the Newfoundland-Irish and Newfoundland-English populations, suggesting a closer genetic relationship among these populations.

The PC plot and admixture analyses are consistent with Irish populations contributing substantial genetic input to both the Newfoundland-Irish and Newfoundland-English

populations. Samples were assigned ethnicity on the basis of family history and also surname. It is possible that some samples were assigned mistakenly to the English subpopulation. The east of Ireland has a history of colonization by Vikings, Normans and British and many non-indigenous Irish surnames occur in the region. Also, many Newfoundland surnames occur both in Southeast Ireland and Southwest Britain, the proposed homeland of the Newfoundland-English population (Mannion 1999). Supporting a non-indigenous element in the southeast Irish Y chromosome population, the plot suggests that the Southeast Irish population is closer to the British populations than the other Irish subpopulation from the west, thought to have remained largely free of Viking, Norman and British genetic introgression.

5.5 Conclusion

This investigation of Newfoundland genetic history shows that at the Y haplotype level the Newfoundland-Irish and Newfoundland-English populations are very similar. This is likely due to genetic admixture between these populations. Historical and genetic evidence suggests that may have been 'admixture at source'. In the southeast of Ireland, the region from where most Newfoundland-Irish immigrants supposedly came, there is evidence for British genetic influence (see Chapter 2). Alternatively, it may be that the Anglicisation of Gaelic surnames and the relatively common presence of non-Gaelic surnames in the Irish population has meant that the ethnicity of some Newfoundland samples was incorrectly assigned. However this is unlikely as samples were chosen for their well-documented ethnicity and established history in Newfoundland.

The Principal Components plot suggests that the sample from the southeast of Ireland may be genetically close to the Newfoundland samples, in agreement with historical evidence for substantial levels of migration to Newfoundland from this region. This may also be partly due to a substantial British genetic contribution to the southeast of Ireland, effectively making this regional sample an admixture of Irish and English Y chromosomes also. It is also possible that the Newfoundland populations were closest to the southeast Irish population because the appropriate English regional subpopulation was not sampled.

Chapter 6 – Autosomal microsatellite analysis of Irish, Danish and Basque samples.

78 Samples with ancestry from the east and the west of Ireland were analysed for 8 autosomal microsatellite loci in order to examine whether there is detectable regional genetic differences in the Irish population for these loci. The Irish data were compared with samples of English, Danish and Basque ancestry. These were analysed using the Bayesian clustering algorithm implemented in Structure 2.1 (Pritchard et al. 2000), which indicated no detectable genetic structure within these populations.

6.1 Introduction

6.1.1 Sex-specific loci and Irish population genetic history

Y chromosome variation in Europe is non-random. The most obvious pattern is a continent-wide southeast – northwest cline, with Hg R1b3 frequency reaching a maximum in populations on the Atlantic coast of Europe. Populations from the Basque country and the west of Ireland show exceptionally high frequency of Hg R1b3.

Within Ireland Y chromosome Hg frequency is non-random, with the north and west possessing higher Hg R1b3 frequency than the south and east. In contrast, there does not appear to be significant differences among regional Irish mtDNA populations (McEvoy *et al.* 2004). This discrepancy between male- and female-transmitted loci has been noted in other human populations and illustrates how different loci record population history in different ways. Nevertheless the usefulness of the Y chromosome and mtDNA in population genetic studies has been precisely due to the fact that they are single loci, from which complex haplotypes can be constructed to infer evolutionary history because these molecules do not undergo recombination.

6.1.2 Analysis of autosomal loci

A more complete picture of the genetic history of populations may be obtained by examination of autosomal loci, since these do not have the sex-specific inheritance mechanisms of the Y and mtDNA chromosomes. Autosomal DNA undergoes

recombination, meaning extended haplotypes cannot usually be reconstructed for this type of data; but it is possible to use allele frequencies at autosomal loci to infer genetic ancestry of individuals and populations. Gene frequencies of protein polymorphisms have described patterns in global genetic variation that which have been interpreted as the result of past population movement (Cavalli-Sforza *et al.* 1994).

Within Ireland, blood group and disease gene frequencies have described regional patterns, which are often a continuation of a European-wide cline. For example, Blood Group O frequency peaks in the west of Ireland (Hackett and Dawson 1958), as does the relative frequency of the R408W phenylketonuria mutation (Tighe *et al.* 2003). X chromosome analysis has not indicated significant differences between east and west Irish populations (Chapter 4).

6.1.3 Atlantic and Continental admixture?

The peak of Y chromosome Hg R1b3 frequency in western Ireland and the Basque country suggests a common Atlantic zone Y chromosome heritage. Y chromosome analysis in the British Isles has suggested that non-R1b3 Y chromosomes in this region may be mainly of a continental or Germanic source (Wilson *et al.* 2001, Weale *et al.* 2002, Capelli *et al.* 2003). These findings suggest that at the Y chromosome level at least, the east – west European cline may approximate to an Atlantic versus continental, or ‘Celtic’ versus ‘Germanic’ cline of genetic admixture.

6.1.4 Linkage disequilibrium, admixture and ethnic ancestry

Recently, approaches have been developed which use population-specific alleles to estimate the relative parental genetic contributions to an admixed population. Such underlying genetic structure may have important epidemiological consequences (Wilson *et al.* 2001). A similar approach can be used to estimate the proportionate ethnic ancestry of a particular individual within an admixed population.

When genetically distinct populations merge, linkage disequilibrium is created among non-linked loci in the admixed population for all loci that have large allele frequency differences in the parental populations. This has been exploited in the mapping of disease genes in admixed populations and in the detection of genetic ancestry in populations created when two ethnically distinct populations hybridise. For example Parra *et al.* (1998) used this approach to estimate the European genetic contribution to ten populations of African American descent in the U.S. and Jamaica. Brutsaert *et al.* (2003) estimated the proportion of Spanish ancestry in a group of Peruvian Quechua and Collins-Schramm *et al.* (2002) estimated the proportion of European and African ancestry in African-Americans using this approach.

Loci with large frequency differences among ethnically distinct populations are known as Ancestry Informative Markers (AIMs). They can be used to estimate ancestry at the level of the population, subgroup, and individual. (Kittles and Weiss 2003).

6.1.5 Recent approaches to the analysis of autosomal loci

Recently, model-based approaches to the analysis of autosomal data have been developed which can be used to detect admixture within populations. Rosenberg *et al.* (2002) examined allele frequencies at 377 unlinked loci in 52 global populations using a Bayesian clustering approach developed by Pritchard *et al.* (2000) which infers population structure on the basis that any given genetic dataset contains *K* subpopulations, each of which is characterised by a set of allele frequencies at each locus. The programme Structure (Pritchard *et al.* 2000) identifies sub-groups within a population which have distinctive allele frequencies and therefore presumably distinct genetic ancestry. Individuals can be assigned probabilistically to one or more of the sub-populations on the basis of their genotype, and their proportionate genetic ancestry can be estimated. Rosenberg *et al.* (2002) identified six main genetic clusters in their global dataset, five of which corresponded to major geographic regions.

6.1.6 Detecting genetic structure within the Irish population

Data from various genetic systems, notably Y chromosomes, ABO blood group data and disease gene frequency data, have indicated east – west clines within Ireland, which may reflect Atlantic versus Continental genetic ancestry. In order to examine whether Irish autosomal genetic variation may be described in terms of Atlantic and Germanic admixture, samples from east and west Ireland were compared with samples from the Basque population (representing the presumed Atlantic source population) and the Danish population (representing the presumed Continental source population). Samples with British ancestry were also examined. Samples were typed

for autosomal microsatellite loci indicated to have high frequency differences among Basque versus Continental populations, and admixture was assessed at the population and the individual level.

6.2 Materials and Methods

6.2.1 Selection of AIMs

The selection of suitable AIMs is required in order to test for genetic structure within the Irish population. The European data of Rosenberg *et al.* (2002) was analysed. Their Basque sample was used to represent the Atlantic source population. All analyses were repeated for two samples used to represent Continental source population. Firstly the French sample of Rosenberg *et al.* (2002) was used to represent the Continental source population, and subsequently all calculations were repeated with the Continental sample represented by the French, Italian and Orcadian samples of Rosenberg *et al.* combined.

Since the desired characteristic for AIMs are large genetic distances between parental populations, we computed Nei's Standard Genetic Distance for a single locus with N alleles (D_s) between the Basque and the European populations, and selected those loci showing the highest genetic distance between the two, on the basis that these loci would be most likely to reflect Atlantic versus Continental genetic ancestry.

D_s for each locus was calculated as follows:

First, the genetic identity, I , was calculated for each locus between the European and Basque samples:

$$I = J_{xy}/(J_x J_y)^{1/2}$$

Where:

$$J_{xy} = \sum_{i=1}^n P_{ix} P_{iy}$$

$$J_x = \sum_{i=1}^n P_{ix}^2$$

$$J_y = \sum_{i=1}^n P_{iy}^2$$

X and Y are the two populations being compared

P_{ix} is the frequency of the allele in population X, summed for the frequencies of each allele

P_{iy} is the frequency of the allele in population Y, summed for the frequencies of each allele

$$J_{xy} = \sum (\text{summed over all loci})(\text{freq. of allele } x \times \text{freq allele } y)$$

Then $D_s = -\ln I$

Ds distances were calculated for the 377 loci presented in Rosenberg *et al.* and the 20 loci with the highest Ds values were chosen. For these loci, Ds distances were calculated between the Basques and each of the European sub-populations (France, Italy, Orkney) separately. 10 loci which indicated consistently highest Ds distances between Basques and each of the individual European populations were chosen for amplification. These loci are listed in Table 6.2.3.1.

6.2.2 Choosing samples for amplification

Samples from east and west Ireland were chosen for amplification. In order to maximise the chance of finding genetic structure between the east and west Irish sub-populations, all samples chosen from the west of Ireland possessed Hg R1b3 Y chromosomes, whereas all samples chosen from the east of Ireland possessed non-Hg

R1b3 Y chromosomes. 39 samples from the east and 39 from the west were amplified along with 22 samples with English ancestry, 33 Danish samples and 40 Basque samples.

6.2.3 PCR amplification of loci

Primers for the loci were obtained from the Genome Database (www.gdb.org) and are listed below.

Locus	Ds* France/Europe		Chr.	Dye	Primers
GATA73D10	0.21	0.29	7	Tamra	FOR-tca tga caa aac aga aat taa gtg REV-agt aaa tgg aat tgc ttg tta cc
GATA72E05	0.35	0.26	23	Hex	FOR-agc cca ttt tca taa taa atc c REV-aat cag tgc ttt ctg tac tat tgg
GATA43H03	0.25	0.11	13	Hex	FOR-ggc ata aaa ata gta cag caa gc REV-att tgg aca gag gca tgg ac
GATA42A03	0.24	0.09	20	Fam	FOR-cca aag aca agt ttg ttt atc tg REV-tgt tat tgc agc agc aat aga aaa cta a
GATA31E10	0.36	0.29	Y	Hex	FOR-tat att tta cac att ttt ggg cc REV-tga cag taa aat gaa cac att gc
GATA23F06	0.22	0.16	11	Fam	FOR-tac atg gca gca ggc ata ta REV-gag taa aca aga ttg cta gat agg c
GATA164B08	0.53	0.15	3	Hex	FOR-ctg tga tca cac cac tgc ag REV-ttc ggt att ctg tgt cag agc
GATA144D04	0.36	0.31	23	Tamra	FOR-agc cag aag gag gag gat ta REV-atc ttt ctg tac ttg cag tga gg
ATA28G05	0.29	0.14	15	Fam	FOR-cac att gga ggg gtg att ta REV-ccc tca gaa aag cct tta gc
273YF9	0.22	0.29	15	Fam	FOR-gct gtg tgt aag tgt gtt tta tat c REV-gca agc cag tgg aga g

Table 6.2.3.1: Chromosomal location and primers for 10 loci selected for amplification.
*Indicates average Ds genetic distance of Basques vs. French and Basques vs. Europe (France, Itlay, Orkney combined).

A 10ul final volume PCR mix for each amplification reaction contained:

15mM PCR buffer (as Chapter 2)

1.5μM FOR and REV primer

2uM dNTPs,

0.1μl enzyme

3.9μl H₂O

PCR cycling conditions for all loci were:

35 cycles of:

94°C 30 sec

55°C 30 sec

72°C 60 sec

Followed by 5 minutes at 72°C

Loci were amplified and electrophoresed under the conditions outlined in Chapter 2.

Amplification was successful for all loci except GATA144D04 and GATA42A03.

6.2.4 Analysis of genotype data

The genotypes for the east and west Irish, English, Danish and Basque samples were analysed using the programme Structure 2.1. (Pritchard *et al.* 2000).

The admixture model was used and the MCMC scheme was varied for values for K (the number of sub-populations) from $K=1$ to $K=5$. Three runs for each K value were performed to check that the MCMC scheme was converging. No information about the geographic origins of the samples was used in implementing the programme. 100,000 burnin steps were used, followed by 100,000 simulation steps.

D_s values were also computed at the population level for Basques versus west Irish, east Irish, British and Danish populations.

To further assess admixture at the individual level, an allele-sharing matrix was constructed which compared all individuals to one another and this was visualised using a Multidimensional Scaling Plot implemented in SPSS.

6.3 Results

6.3.1 Selection of loci with large Ds values

A graph showing the Ds values obtained by comparing Basque versus French data for Rosenberg’s 377 loci is shown below (Figure 6.3.1.1). Those loci with a Ds value greater than 0.2 were chosen as potential AIMs.

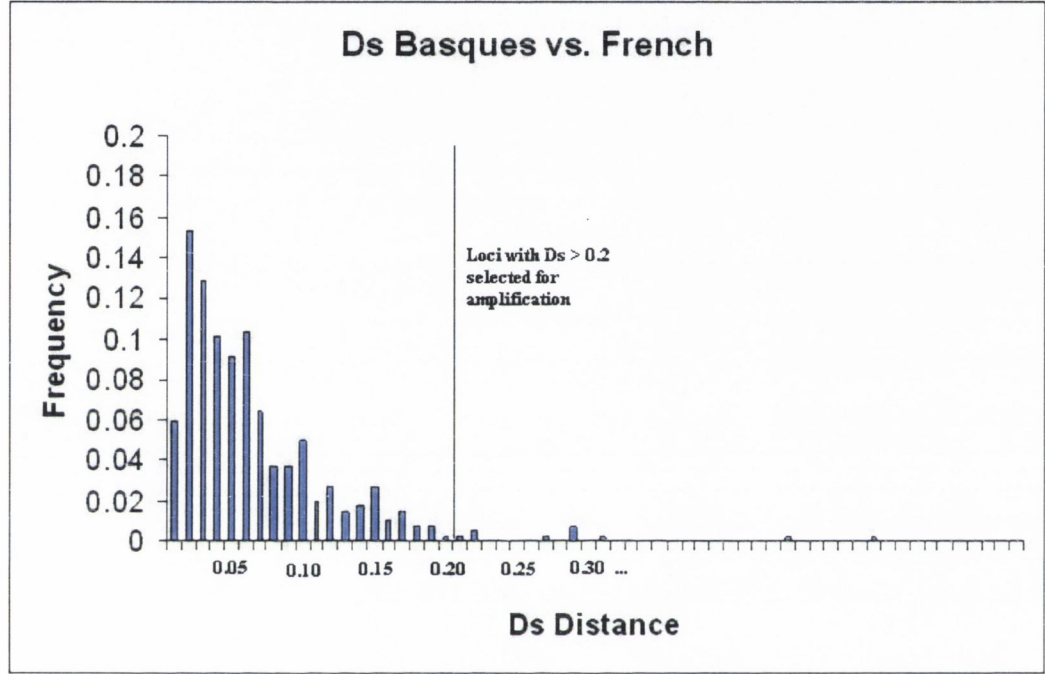


Figure 6.3.1.1: The frequency distribution of Ds values obtained for the Basque versus the French samples of Rosenberg et al. (2002). Loci with a Ds value of greater than 0.2 were selected as possible AIMs.

A graph comparing Ds values for Basque versus French and Basque versus Europeans for each locus is shown below (Figure 6.3.1.2). Before loci were chosen for amplification it was ensured that it scored a high Ds value for Basques versus all European populations.

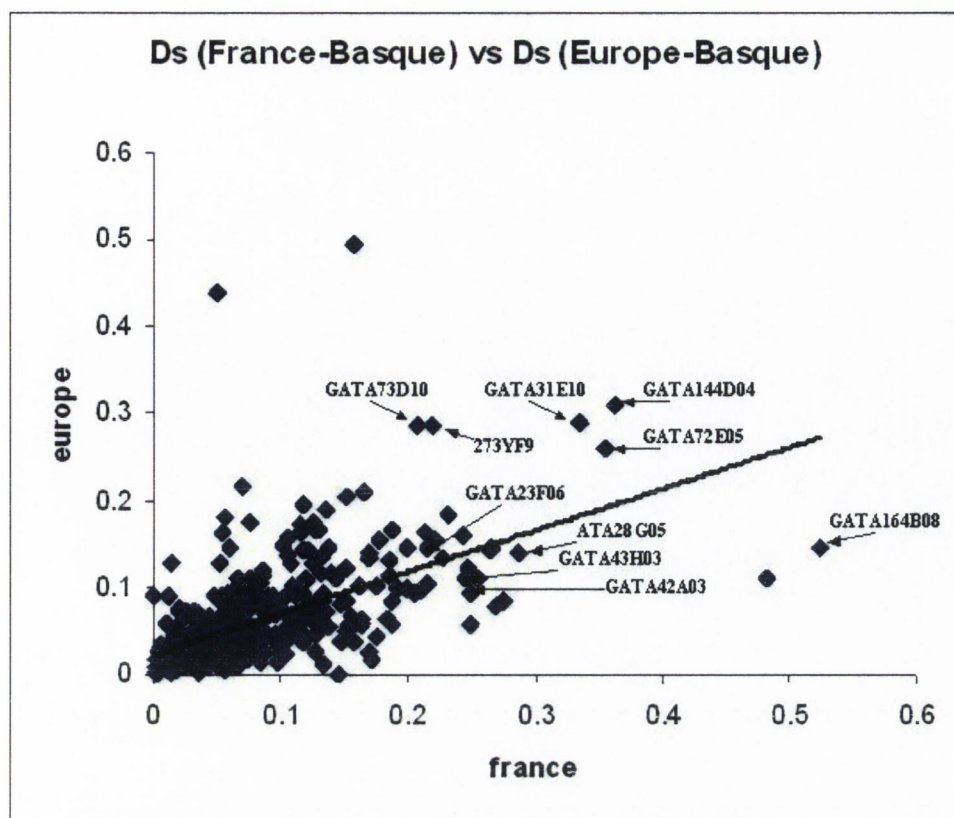


Figure 6.3.1.2: Scatter plot comparing Ds values obtained from Basque vs. French and Basque vs. Europeans (French, Italian and Orcadian combined). Loci with high Ds values in both groups were chosen for amplification. These loci are indicated on the plot.

6.3.2 Genotype data

Genotypes for the Irish, British, Danish and Basque samples are given in Appendix 6.1.

6.3.3 Structure Results

Figure 6.3.3.1 below shows a plot of K versus Estimated Log Probability, varying K from 1 to 5. The plot shows the general trend that as K (the number of sub-populations) decreases, the estimated log probability of the data increases. This is suggestive of a general lack of genetic structure within the data.

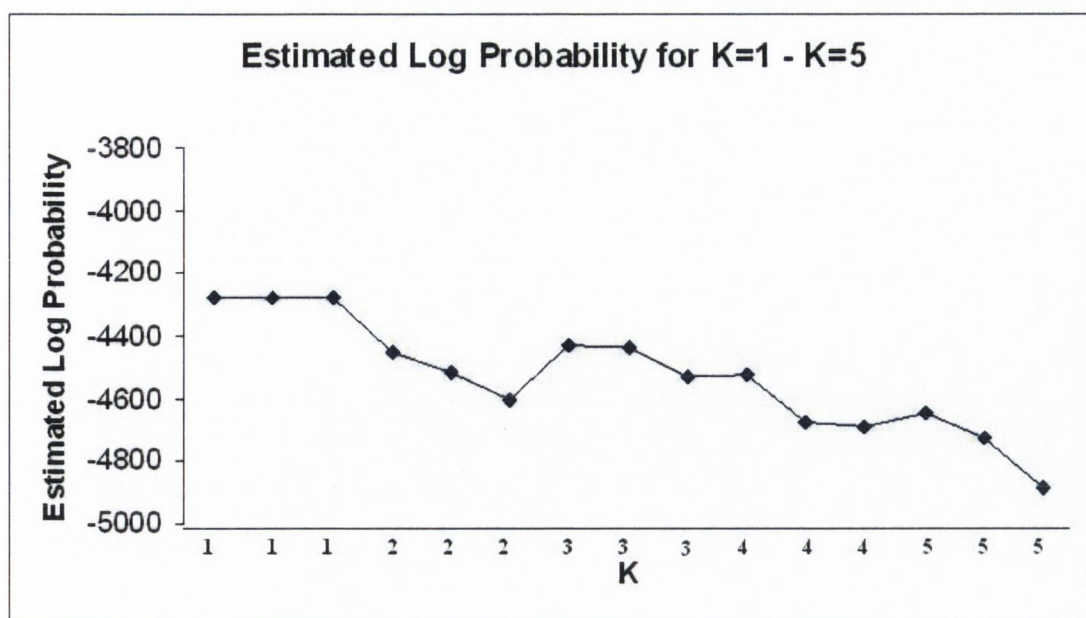


Figure 6.3.3.1: variation of estimated log probability varies for different values of K, the number of inferred clusters in the data.

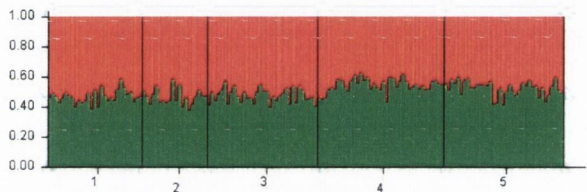
When Structure is run for different values of K, the estimated log probability of the data for each run can be compared. The value of K for which the probability of the data is highest represents the best estimate for K, the number of sub-populations. Multiple runs at each K ensure that the estimates are reasonably consistent across runs. Examination of the posterior probabilities of K shows that K=1 consistently gives the highest estimated log probability of the data (-4276.6, -4274.1,-4277.7), suggesting that there is not significant genetic structure among the Irish, English, Danish and Basque populations for these loci.

Plots of structure outputs varying K from 2 to 5 are shown in Figure 6.3.3.2. The plots represent the estimated membership coefficients for each individual in each cluster / sub-population. Every individual is represented by a vertical line divided into K coloured segments, each of which represent the individual's membership fraction of

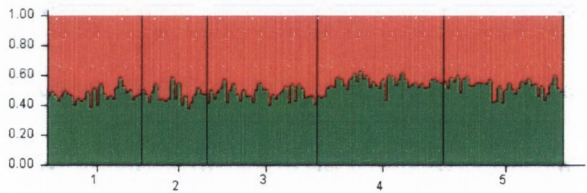
that cluster. Each plot represents a single run of the Structure programme for a given value of K, and the estimated Log probability of the data for that run is given below each plot.

Plots for values of $K > 1$ show patterns whereby the proportions of each sample assigned to each cluster are roughly symmetric, with $\sim 1/k$ assigned to each cluster and most individuals being fairly evenly admixed. According to Pritchard et al. (2000) this is indicative of lack of structure within the data.

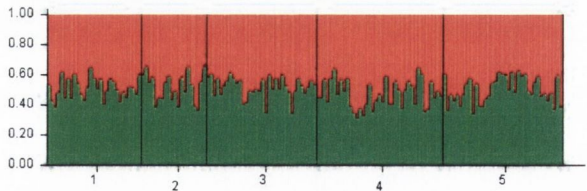
K = 2



Run 1: Estimated Log Probability of Data = -4601.6

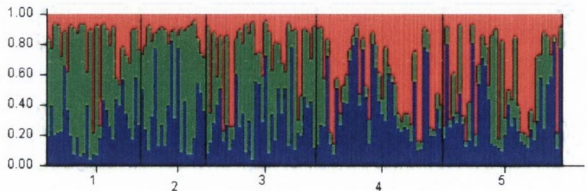


Run 2: Estimated Log Probability of Data = -4516.0

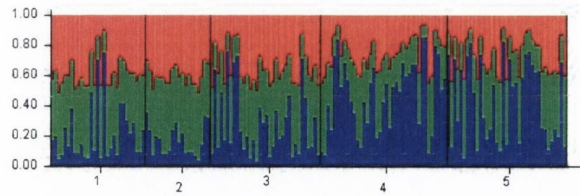


Run 3: Estimated Log Probability of Data = -4448.3

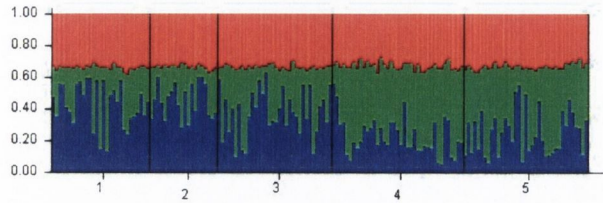
K = 3



Run 1: Estimated Log Probability of Data = -4430.2

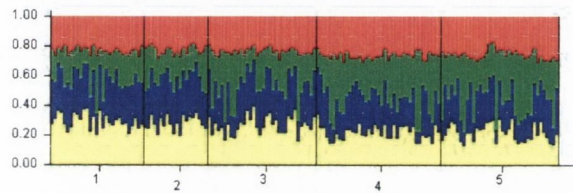


Run 2: Estimated Log Probability of Data = -4526.9

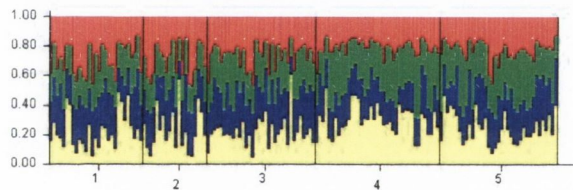


Run 3: Estimated Log Probability of Data = -4437.7

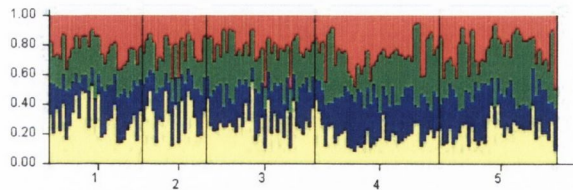
K = 4



Run 1: Estimated Log Probability of Data = -4677.1

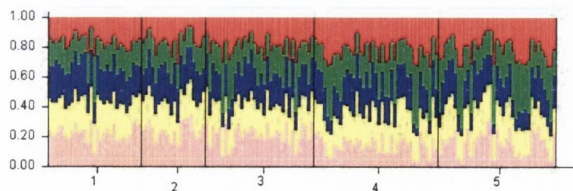


Run 2: Estimated Log Probability of Data = -4687.4



Run 3: Estimated Log Probability of Data = -4522.7

K = 5



Run 1: Estimated Log Probability of Data = -4728.0

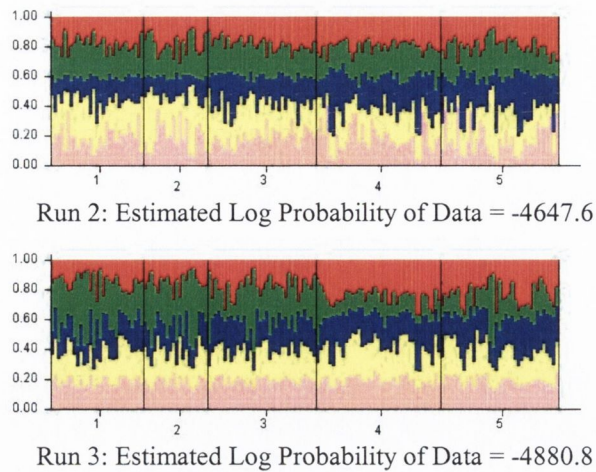


Figure 6.3.3.2: Structure outputs for K=1 to K=5. Each individual is represented by a single line, divided into K segments, representing the membership coefficient of the individual to each inferred cluster. Individuals are grouped according to population as follows: Population 1 = Danes, Population 2 = English, Population 3 = East Ireland, Population 4 = West Ireland, Population 5 = Basques.

The plots for K=3 shows some evidence of asymmetric proportions assigned to individuals, which may indicate population structure. Populations 4 (West Ireland) and 5 (Basques) appear to have similar ancestry proportions to one another, whereas the population from East Ireland (3) appears to have proportions similar to the English (2) and Danes (1). However the log probability of the data for each run at K=3 (-4430.2, -4526.9, -4437.7) is consistently substantially lower than for each run at K=1 (-4276.6, -4274.1, -4277.7), indicating that K=1, or no underlying population structure, constitutes the best fit for the data.

6.3.4 Genetic distance by population

Ds values for each calculated for Basques versus each of the other groups are shown below (Figure 6.3.4.1). They show that for 7 of the 8 loci analysed there does not appear to be a detectable closer affinity among the Basque and the Irish populations

than among the Basque and the British or Danish populations, again suggesting lack of detectable genetic structure within these populations. Indeed the only obvious exception to this pattern, locus GATA73E10, is located on the Y chromosome, which is already known to be structured within these populations from the choice of individuals used to represent the west and east sub-populations. Genetic distances between the Basque and other populations for this locus follow an east – west geographic pattern, with the Basque – Danish distance being greatest, followed by the Basque – British distance, then the Basque – east Irish and finally the Basques and west Irish.

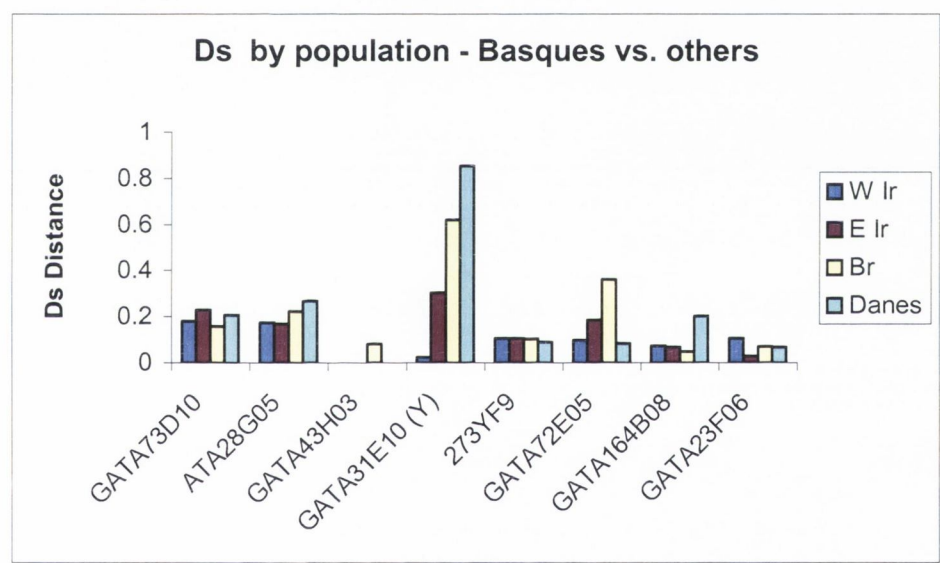


Figure 6.3.4.1: Ds distances for each locus calculated by population. Only the Y locus shows a gradient of increased Ds from ‘Atlantic’ through to ‘Continental’ populations.

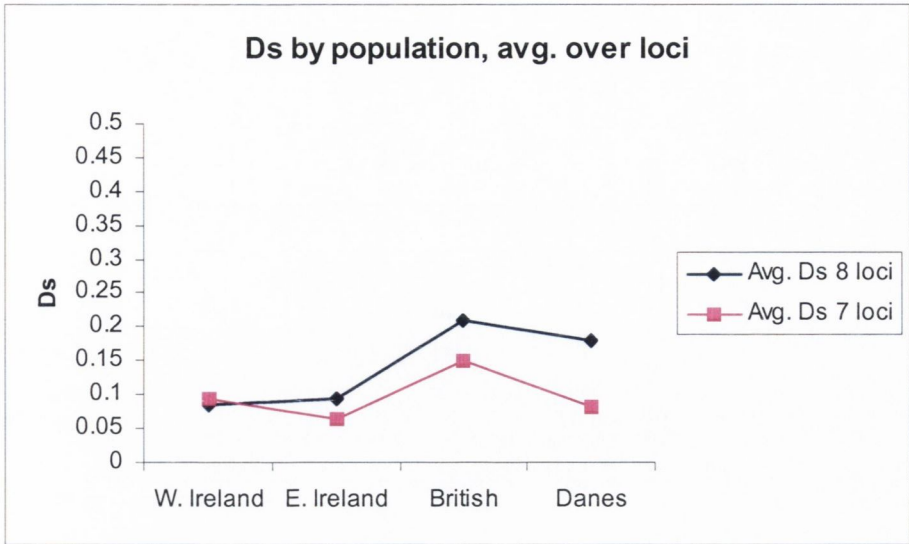


Figure 6.3.4.2: Average Ds for Basques vs. other populations. The navy line indicates Ds when the Y locus GATA73D10 is included in the analysis, the pink line indicates Ds when GATA73D10 is omitted.

Figure 6.3.4.2 shows that when Ds distances by population are averaged over all 8 loci there is a general increase in genetic distance from ‘Atlantic’ to ‘Continental’ populations. However, when the Y locus, GATA73E10, is removed from this calculation, this trend disappears.

6.3.5 MDS analysis

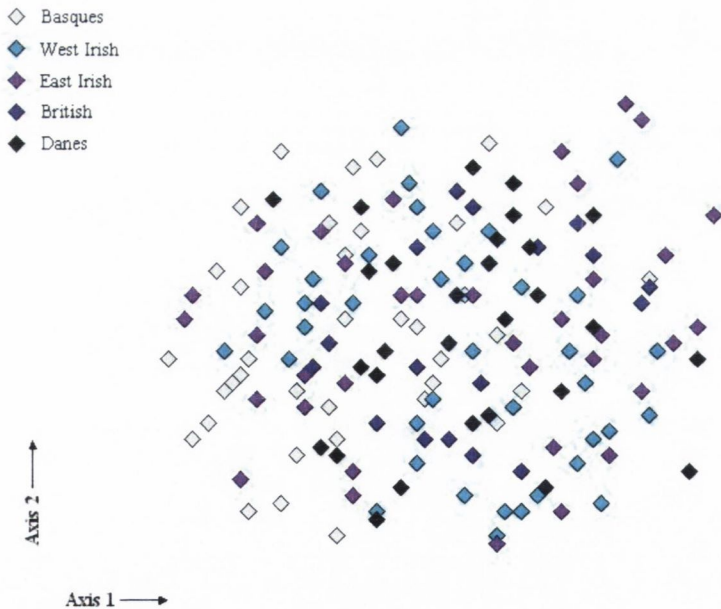


Figure 6.3.5.1: MDS plot constructed from an allele sharing matrix of 8 unlinked microsatellite loci showing the relationship of individuals, whose origin is indicated in the legend.

The Multidimensional Scaling Plot shown in Figure 6.3.5.1 was constructed from an allele-sharing matrix of 8 microsatellite loci. Individuals are represented by diamonds and their source population is indicated by colour. It appears that individuals from the various populations are randomly distributed and there is no obvious clustering of individuals according to population of origin, or any clustering of populations. This is in agreement with the Structure outputs which do not suggest strong genetic structure within these populations.

6.4 Discussion

This study suggests that there is no detectable genetic structure between sub-populations from the east and west of Ireland for the loci examined here. This contrasts sharply with the findings of regional differences in Y chromosome diversity in previous chapters (and with the Y microsatellite locus typed in this study). Genetic studies have indicated that there is a stronger geographic signal attached to the Y chromosome than to other loci. This may be due to the smaller effective population size of the Y, $\frac{1}{4}$ that of autosomes, which accentuates the effects of genetic drift, or to the comparatively lower mobility of males throughout history (*Seielstad et al.* 1998).

It is possible that the lack of genetic structure in Ireland is a reflection of general European genetic homogeneity. In this study genetic structure was not detected among the various populations examined for these loci. In their examination of genetic structure, Rosenberg *et al.* (2002) found that Europe was the most difficult region in which to detect significant population structure, and that Europe had the smallest among-population variance component for the 377 loci examined (0.7%).

It may be that significant genetic structure within the Irish population will be detectable with a larger sample size and number of loci. However, the 8 loci in question were selected on the basis that they were likely to be ancestry-informative and the findings of the lack of genetic structure are in agreement with Irish mtDNA

variation (McEvoy *et al.* 2004), Irish X chromosome variation (Chapter 4) and with a study of 380 autosomal loci in 194 individuals from east Ireland, west Ireland and Britain, which failed to detect significant genetic structure (Brian McEvoy, personal communication).

Appendix Section 1: Conversion from base pairs to repeats

MS1

DYS19	CTAT/C	DYS388	ATA	DYS390	CTG/AT	DYS391	CTAT	DYS392	ATT	DYS393	GATA
174	10	123	10	192	18	157	8	155	7	107	9
178	11	126	11	196	19	161	9	158	8	111	10
182	12	129	12	200	20	165	10	161	9	115	11
186	13	132	13	204	21	169	11	164	10	119	12
190	14	135	14	208	22	173	12	167	11	123	13
194	15	138	15	212	23	177	13	170	12	127	14
198	16	141	16	216	24			173	13	131	15
202	17			220	25			176	14		
206	18			224	26			179	15		
210	19			228	27			182	16		

CTS

DYS434	CTAT	DYS435	TGGA	DYS436	GTT	DYS437	TCAT	DYS438	TTTTTC	DYS439	GATA
110	8	220	11	128	10	186	14	203	6	238	9
114	9	224	12	131	11	190	15	208	7	242	10
118	10	228	13	134	12	194	16	213	8	246	11
122	11			137	13	198	17	218	9	250	12
				140	14	202	18	223	10	254	13
				143	15			228	11	258	14
								233	12		

EBF

DYS389I	CTG/AT	DYS389II	CTG/AT	YA7.1	GATA	YA7.2	GATA	GO4911	TGTA
142	10	258	26	170	9	174	8	180	10
146	11	262	27	174	10	178	9	184	11
150	12	266	28	178	11	182	10	188	12
154	13	270	29	182	12	186	11	192	13
158	14	274	30	186	13	190	12	196	14
162	15	278	31						
166	16	282	32						
		286	33						
		290	34						

Appendix Section 2: Chapter 2 Genotypes

HG	DYS 19	388	DYS 390	391	DYS 392	393	434	435	DYS 436	437	438	DYS 439	389I	DYS	G0	Frequency										Total	
																N. west	N. east	N. Conn.	Midl. N	CGT	Midl. S	Dub	S. East	S. West			
R1b3xR1b3f	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	10	11	11	3	12	0	0	2	1	0	1	30
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	12	13	16	11	10	11	1	1	0	1	5	2	3	4	1	18
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	11	13	16	11	10	11	1	1	1	1	1	3	1	0	3	12
R1b3xR1b3f	14	12	24	10	13	13	9	11	12	15	12	12	13	16	11	10	11	2	1	0	0	0	2	1	0	3	9
R1b3xR1b3f	14	12	24	10	13	13	9	11	12	15	12	11	13	16	11	10	11	0	0	1	1	2	0	0	2	1	7
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	13	13	16	11	10	11	0	2	0	1	0	0	1	2	0	6
R1b3xR1b3f	14	12	25	10	14	13	9	11	12	15	12	12	13	16	11	10	11	0	1	1	3	0	0	0	0	0	5
R1b3xR1b3f	14	12	24	11	14	13	9	11	12	15	12	12	13	16	11	10	11	0	0	2	0	0	3	0	0	0	5
R1b3xR1b3f	14	12	24	10	14	13	9	11	12	15	12	12	13	16	11	10	11	1	1	0	1	0	0	0	2	0	5
R1b3xR1b3f	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	10	11	0	0	2	0	0	1	1	1	0	5
R1b3xR1b3f	14	12	25	11	13	13	9	11	12	15	12	12	13	16	11	10	11	0	0	0	0	0	1	0	1	3	5
R1b3xR1b3f	14	12	24	10	13	13	9	11	12	15	12	13	13	16	11	10	11	0	0	1	1	0	0	1	0	0	3
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	14	13	16	11	10	11	0	1	0	0	0	1	1	0	0	3
R1b3xR1b3f	14	12	24	10	13	13	9	11	12	15	12	12	13	16	11	9	11	0	0	0	0	1	0	2	0	0	3
R1b3xR1b3f	14	12	23	11	13	13	9	11	12	15	12	12	13	16	10	10	11	0	0	1	0	1	0	1	0	0	3
R1b3xR1b3f	14	12	25	10	14	13	9	11	12	15	12	13	13	16	11	10	11	1	1	0	1	0	0	0	0	0	3
R1b3xR1b3f	14	12	25	10	14	13	9	11	12	15	12	13	13	16	11	10	11	1	0	1	0	0	0	0	0	0	3
ix1b2	15	13	24	10	12	15	9	11	12	15	10	12	13	16	11	10	12	0	1	0	1	0	0	1	0	0	3
R1b3xR1b3f	14	12	24	10	14	13	9	11	12	15	12	13	13	16	11	10	11	0	0	0	1	0	2	0	0	0	3
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	12	13	16	11	11	11	0	0	0	2	0	1	0	0	0	3
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	12	14	16	11	10	11	0	0	0	1	0	0	2	0	0	3
R1b3xR1b3f	14	12	23	11	13	13	9	11	12	15	12	12	13	16	10	10	11	0	0	1	0	0	1	0	1	0	3
R1b3xR1b3f	14	12	23	11	13	13	9	11	12	15	12	11	13	16	11	10	11	0	0	0	0	0	1	0	1	0	3
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	12	13	16	11	10	11	1	0	1	0	0	0	0	1	2	3
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	11	13	16	10	10	11	0	0	0	0	0	0	0	2	1	3
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	11	13	16	10	10	11	0	0	1	1	0	0	0	0	1	3
R1b3xR1b3f	15	12	25	11	14	13	9	11	12	15	12	13	13	16	11	10	11	1	0	0	0	1	0	0	0	0	2
R1b3xR1b3f	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	11	11	1	0	0	0	0	1	0	0	0	2
R1b3xR1b3f	14	12	25	10	13	13	9	11	12	15	12	13	13	16	11	10	11	0	0	0	1	0	0	1	0	0	2
R1b3xR1b3f	14	12	26	10	14	13	9	11	12	15	12	12	13	16	11	10	11	0	0	0	1	0	0	1	0	0	2
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	11	14	16	11	10	11	0	1	0	1	0	0	0	0	0	2
R1b3xR1b3f	14	12	24	10	14	13	9	11	12	15	12	12	13	16	11	11	11	0	1	0	1	0	0	0	0	0	2
R1b3xR1b3f	14	12	25	11	14	13	9	11	12	15	12	13	13	16	11	10	12	0	0	1	1	0	0	0	0	0	2
R1b3xR1b3f	14	12	25	11	13	13	9	11	12	15	12	12	13	16	11	10	11	0	0	0	1	0	0	0	0	0	2
R1b3xR1b3f	14	12	25	11	13	13	9	11	12	15	12	13	13	16	11	10	11	0	0	0	1	0	0	0	0	0	2
R1b3xR1b3f	14	12	23	12	13	13	9	11	12	15	12	11	13	16	11	10	11	0	0	1	1	0	0	0	0	0	2
R1b3xR1b3f	14	12	24	11	13	13	9	12	12	15	12	12	14	16	11	10	11	0	0	0	1	0	1	0	0	0	2
R1b3xR1b3f	15	12	24	11	13	13	9	11	12	15	12	13	13	16	11	10	11	0	0	1	0	0	1	0	0	0	2
R1b3xR1b3f	15	12	24	11	13	13	9	11	12	15	12	13	13	16	11	10	11	0	0	1	0	0	1	0	0	0	2
R1b3xR1b3f	14	13	25	11	14	13	9	11	12	15	12	13	13	16	11	10	11	1	0	1	0	0	0	0	0	0	2
R1b3xR1b3f	14	13	25	11	14	13	9	11	12	15	12	13	13	16	11	10	11	0	0	1	0	0	0	0	0	0	2
R1b3xR1b3f	14	12	23	10	13	13	9	11	12	15	12	11	13	16	11	11	11	0	0	0	0	0	0	1	1	0	2
R1b3xR1b3f	14	12	24	11	13	12	9	11	12	15	12	12	13	16	11	10	11	0	0	0	0	0	0	2	0	0	2
R1b3xR1b3f	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	10	11	0	0	0	0	0	0	1	1	0	2
R1b3xR1b3f	14	12	23	10	13	13	9	11	12	15	12	11	13	16	11	10	11	0	0	0	0	0	0	1	1	1	2
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	12	14	16	10	9	11	0	0	0	0	0	0	0	1	1	2
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	12	13	16	11	10	11	0	0	0	1	0	0	0	1	1	2
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	12	13	16	11	10	11	0	0	0	1	0	0	0	1	1	2
R1b3xR1b3f	14	12	25	12	13	13	9	11	12	15	12	12	13	16	11	10	11	0	0	0	0	0	0	0	2	2	2
R1b3xR1b3f	14	12	24	10	13	13	9	11	12	15	12	11	14	16	11	10	11	0	0	0	0	1	0	0	1	1	2
R1b3xR1b3f	14	12	25	11	13	13	9	11	12	15	12	13	13	16	11	10	11	0	0	0	1	0	0	0	0	1	2
R1b3xR1b3f	14	12	23	11	13	13	9	11	12	15	12	13	13	16	11	10	11	0	0	0	1	0	0	0	0	1	2
R1b3xR1b3f	15	12	23	11	13	13	9	11	12	15	12	12	13	16	11	10	11	0	0	0	0	1	0	0	0	1	2
R1b3xR1b3f	15	12	23	11	13	13	9	11	12	15	12	12	13	16	11	10	11	0	0	0	0	1	0	0	0	1	2
R1b3xR1b3f	14	12	24	10	13	13	9	11	12	15	12	13	16	11	11	11	11	0	0	0	0	0	1	0	0	1	2
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	13	16	11	10	11	11	0	0	0	0	0	0	1	0	1	2
R1b3xR1b3f	15	12	23	11	13	13	9	11	12	15	12	12	13	16	11	10	11	0	0	0	0	1	0	0	0	1	2
R1b3xR1b3f	14	12	24	10	13	13	9	11	12	15	12	13	16	11	11	11	11	0	0	0	0	0	0	1	0	1	2
R1b3xR1b3f	14	12	24	11	13	13	9	11	12	15	12	13	16	11	10	11	11	0	0	0	0	0	0	0	1	2	2
R1b3xR1b3f	14	12	24	10	13	13	9	11	12	15	12	13	16	11	10	11	11	0	0	0	0	0	0	0	1	2	2
R1b3xR1b3f	14	12	25	11	13	13	9	11	12	15	12	13	13	16	11	10	11	0	0	0	1	0	0	0	0	1	2
R1b3xR1b3f	14	12	23	11	13	13	9	11	12	15	12	13	13	16	11	10	11	0	0								

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Appendix Section 3: Chapter 3 Ui Neill Genotypes

Surname	R1b3?	MS1							CTS							EBF							No. diffs from IMH
		14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	10	11					
Donnelly	no	14	14	22	10	11	13	9	11	12	16	10	?	13	16	12	10	11	13				
Flynn	no	15	12	24	10	10	13	9	11	12	15	12	12	14	14	9	9	12	14				
Hynes	no	16	13	23	11	12	16	9	11	12	15	12	12	12	?	11	11	12	13				
O'Connor	no	13	14	22	10	11	13	9	11	12	15	12	12	13	16	11	10	11	10				
Boyle	yes	15	12	25	11	14	13	9	11	12	15	12	12	13	16	11	9	11	2				
Boyle	yes	14	12	24	11	13	13	9	11	12	15	12	12	13	16	11	10	11	2				
Bradley	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	11	11	1				
Bradley	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	10	11	0				
Cannon	yes	15	12	24	11	13	13	9	11	12	15	12	11	13	16	11	10	11	4				
Devlin	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	10	11	0				
Doherty	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	10	11	0				
Donnelly	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	10	10	11	1				
Egan	yes	14	12	24	11	13	13	9	11	12	15	12	13	13	17	10	10	11	5				
Flynn	yes	14	12	24	11	14	13	9	11	12	15	12	12	13	16	11	10	11	1				
Gormley	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	11	11	1				
Hynes	yes	14	12	24	10	13	13	9	11	12	14	10	10	13	16	11	11	11	9				
Kee	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	10	11	0				
McCaul	yes	14	12	25	10	13	13	9	11	12	15	12	12	13	16	11	10	11	2				
McGovern	yes	14	12	25	10	14	13	9	11	12	15	12	12	13	16	11	10	11	1				
McKee	yes	14	12	24	11	14	13	9	11	12	15	12	12	13	16	11	10	11	1				
McLoughlin	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	9	11	1				
McManus	yes	14	12	25	11	13	13	9	11	12	15	12	12	13	16	11	11	12	3				
McMenamin	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	10	11	0				
Molloy	yes	14	12	24	11	13	12	9	11	12	15	12	12	13	16	10	10	12	5				
O'Boyle	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	10	10	12	2				
O'Boyle	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	10	11	0				
O'Connor	yes	14	12	24	11	13	13	9	11	12	15	12	12	13	16	11	10	12	3				
O'Connor	yes	14	12	24	10	13	13	9	11	12	16	10	10	13	16	11	10	11	8				
O'Kane	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	10	11	0				
O'Reilly	yes	14	12	25	10	14	13	9	11	12	15	12	12	13	16	11	10	11	1				
O'Reilly	yes	14	13	24	11	13	13	9	11	12	14	12	12	13	16	10	10	11	5				
O'Rourke	yes	14	12	24	11	14	13	9	11	12	15	12	12	13	16	11	10	11	1				
Quinn	yes	14	12	25	11	14	13	9	11	12	15	12	12	13	16	11	10	11	0				

Appendix Section 4: Chapter 4 Y Genotypes

Hg	DYS 19	DYS 390	DYS 391	DYS 392	DYS 393	DYS 434	DYS 435	DYS 436	DYS 437	DYS 438	DYS 439	DYS 389I	DYS 389b	YA 7.1	YA 7.2	GO 4911
East Ireland																
R1b3xR1b3	14	24	10	13	13	10	11	12	15	12	12	13	17	10	11	11
1x1b2	15	23	10	11	13	10	11	12	16	10	12	12	16	9	10	13
R1b3xR1b3	14	23	11	13	13	10	11	12	15	12	12	13	15	11	10	11
1x1b2	14	23	10	11	13	10	11	12	16	10		13	16	12	10	11
R1b3xR1b3	14	25	10	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	12	13	16	11	11	11
R1b3xR1b3	14	24	10	13	13	10	11	12	16	12	12	13	18	11	11	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	12	12	16	11	10	11
R1b3xR1b3	14	24	10	14	13	10	11	12	15	12	12	14	16	12	11	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	12	12	16	10	10	13
R1b3xR1b3	14	23	11	13	13	11	11	12	15	9	13	13	16	12	9	11
R1b3xR1b3	13	23	11	13	13	10	11	12	15	12	12	14	16	11	10	12
R1b3xR1b3	14	23	11	13	13	10	11	12	14	11		12	15	10	10	11
E3b	14	24	10	11	13	10	11	12	14		12	13	18	11	10	12
R1b3xR1b3	14	24	10	14	13	10	11	12	15	12	12	13	16	11	11	11
R1b3xR1b3	15	25	11	14	13	10	11	12	15	12	12	13	16	12	11	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	12	13	16	11	11	11
R1b3xR1b3	14	24	11	13	13	10	11	12	14	12	12	13	16	10	10	11
R1b3xR1b3	14	21	11	13	13	10	11	12	15	12	12	13	16	11	11	11
R1b3xR1b3	14	25	10	14	13	10	11	12	15	12	12	15	16	11	10	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	11	13	16	11	10	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	24	10	13	13	10	11	12	15	12	12	13	16	11	9	11
R1b3xR1b3	14	23	11	13	13	10	11	12	15	12	11	13	16	11	10	11
1x1b2	14	22	10	11	13	10	11	12	16	10	11	15	15	12	10	12
R1b3xR1b3	14	25	10	15	13	10	11	12	15	12	12	12	18	11	10	11
R1b3xR1b3	14	22	10	13	13	10	11	12	15	13	13	13	16	11	10	12
R1b3xR1b3	14	24	10	13	13	10	11	12	14	11	12	13	16	11	9	12
1x1b2	15	23	11	11	13	10	11	12	16	12	13	14	16	10	10	12
R1b3xR1b3	14	23	11	13	13	10	11	12	15	12	13	13	16	11	10	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12		13	16	12	10	11
1x1b2	15	23	10	11	13	10	9	12	16	10	11	12	16	10	10	13
1x1b2	14	21	11	11	13	10	11	12	15	10	13	13	17	10	11	11
R1b3xR1b3	15	24	11	13	14	10	11	12	15	12	12	14	16	11	10	11
R1b3xR1b3	14	24	11	13	13	10	12	12	15	12	12	14	16	11	10	11
R1b3xR1b3	14	23	11	13	13	11	11	12	15	12	12	13	16	11	10	11
1x1b2	15	23	10	12	14	10	10	12	14	10	11	14	18	11	10	12
1x1b2	14	22	10	11	13	10	11	12	16	10	13	12	16	11	10	12
1x1b2	15	23	10	11	13	10	11	12	16	10	12	13	16	11	10	12
R1b3xR1b3	14	23	11	13	13	10	11	12	14	12	12	12	17	10	10	11
1x1b2	15	23	10	12	15	10	11	12	14	10	11	13	16	10	10	11
R1b3xR1b3	14	26	11	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	24	10	13	13	10	11	12	15	12	11	13	16	11	11	11
1x1b2	15	24	10	11	13	10	11	12	15	12	12	13	17	10	11	11
N3	15	23	11	14	14	10	11	12	15	12	12	13	16	11	10	11
West Ireland																
R1b3xR1b3	14	25	11	13	13	10	11	12	15	12	12	13	13	10	10	11
R1b3xR1b3	14	23	10	13	13	10	11	12	15	12	12	12	16	10	9	10
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	25	11	13	13	10	11	12	15	12	12	13	16	11	11	11
R1b3xR1b3	14	23	11	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	24	10	13	13	10	11	12	15	12	13	13	17	11	9	11
R1b3xR1b3	14	25	11	12	13	10	11	12	15	12	12	13	17	11	11	11
R1b3xR1b3	14	22	11	13	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	11	13	17	11	10	11
R1b3xR1b3	14	24	10	13	13	10	11	12	15	12	12	13	16	11	10	11
1x1b2	14	22	10	11	12	10	11	12	16	10	10	12	16	11	10	12
R1b3xR1b3	14	24	11	13	13	10	11	10	14	11	12	13	16	11	10	11
R1b3xR1b3	14	25	11	13	13	10	11	12	16	12	11	13	16	11	10	11
KxPN3	12	23	10	14	13	10	11	12	14	9	12	13	16	11	10	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	12	14	17	10	10	11
R1b3xR1b3	14	24	10	13	14	10	11	12	15	14	12	11	17	11	10	11
R1b3xR1b3	14	24	9	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	22	11	13	13	10	11	12	15	12	13	14		11	10	11
R1b3xR1b3	14	24	11	14	13	10	11	12	14	12	12	14	17	10	10	11
R1b3xR1b3	14	24	11	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	15	23	10	13	13	10	11	12	15	11	13	14	17	11	10	11
R1b3xR1b3	14	25	12	14	13	10	11	12	15	12	13	13	16	11	10	11
R1b3xR1b3	14	24	10	13	13	10	11	12	15	12	11	13	16	11	10	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	25	10	14	13	10	11	12	15	12	12	13	16	10	10	11
R1b3xR1b3	15	23	11	13	13	10	11	12	15	12	12	13	16	11	9	11

R1b3xR1b3	14	24	10	13	13	10	11	12	15	12	11	13	16	11	10	11
R1b3xR1b3	16	24	11	13	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	24	10	13	13	10	13	12	15	12	11	13	16	11	10	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	15	23	10	13	13	10	11	12	15	12	13	14	17	11	10	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	24	11	13	12	10	11	12	15	12	12	13	16	10	11	11
R1b3xR1b3	14	24	11	12	14	10	11	12	15	12	12	14	16	11	10	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	11	12	16	12	10	11
R1b3xR1b3	14	23	10	13	13	8	11	12	15	12	12	13	17	11	9	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	25	11	14	13	10	11	12	16	13	14	13	16	11	10	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	25	11	14	14	10	11	12	15	12	12	13	16	13	12	12
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	24	10	13	13	10	11	12	15	12	13	13	16	11	10	11
Ix1b2	15	23	10	12	15	10	11	12	14	10	11	13	16	11	10	11
R1b3xR1b3	14	24	10	13	13	10	11	12	15	12	11	13	16	11	10	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	13	13	17	11	10	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	25	11	13	13	10	11	12	15	12	13	13	16	9	10	11
R1b3xR1b3	14	24	10	13	13	10	9	12	15	12	13	13	16	11	10	11
R1b3xR1b3	15	23	10	13	13	10	11	12	15	12	13	14	17	11	10	11
R1b3xR1b3	14	24	10	14	13	10	11	12	15	12	13	13	16	11	11	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	15	25	11	14	13	10	11	12	15	11	12	13	17	11	10	11
R1b3xR1b3	14	26	11	14	13	10	11	12	15	12	12	13	16	12	10	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	12	13	17	11	10	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	13	12	16	11	10	11
R1b3xR1b3	14	25	10	14	13	10	11	12	15	12	13	13	16	11	10	11
R1b3xR1b3	14	25	10	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	11	13	16	11	10	11
R1b3xR1b3	15	23	11	13	15	10	11	12	15	12	12	14	17	11	10	11
Ix1b2	16	24	10	11	13	10	11	12	16	9	14	12	16	11	8	11
R1b3xR1b3	15	24	10	13	13	10	11	12	15	12	12	14	16	12	11	12
R1b3xR1b3	14	23	10	13	13	10	11	12	15	12	12	13	17	11	10	11
R1b3xR1b3	14	25	10	14	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	23	11	13	12	10	11	12	15	12	11	13	16	11	10	11
R1b3xR1b3	15	24	11	13	13	10	11	12	15	12	13	13	16	11	10	11
R1b3xR1b3	14	26	11	14	13	10	11	12	15	12	12	13	16	12	10	11
Ix1b2	15	23	10	12	15	10	11	12	14	10	11	14	18	11	10	12
R1b3xR1b3	14	25	11	14	13	10	12	14	16	12	11	13	16	11	10	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	13	12	16	11	10	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	12	13	17	11	9	11
Ix1b2	14	22	10	11	14	10	11	12	16	10	11	13	17	10	9	12
Ix1b2	14	23	10	11	13	10	11	12	16	10	12	12	17	10	10	13
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	12	12	15	10	10	13
R1b3xR1b3	14	24	10	13	13	10	11	12	15	12	11	13	16	11	9	11
R1b3xR1b3	14	24	11	13	13	10	11	12	15	12	12	13	16	11	10	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	11	12	14	17	11	11	11
R1b3xR1b3	14	25	11	14	13	10	11	12	15	12	12	13	16	11	10	11

Togo																
E3a	15	21	10	11	13	8	11	12	14	11	11	13	19	12	11	12
E3a	15	22	10	11	14	8	11	12	14	10	13	12	17	10	12	12
ExE3	15	22	11	11	12	8	11	12	17	10	12	12	17	11	10	13
E3a	15	21	11	11	13	8	11	12	15	12	12	13	17	10	12	12
E3a	16	21	10	11	15	?	11	12		11	11	13	17	11	12	12
E3a	15	21	10	13	13	8	11	12	14	11	11	13	17	11	11	12
E3a	16	21	10	11	14	8	11	12	14	12	11	13	18	10	11	12
E3a	15	22	10	11	13	8	11	12	14	11	11	13	17	10	11	12
E3a	15	21	10	11	13	8	11	11	14	11	12	14	17	10	11	12
E3a	15	21	10	11	13	8	11	12	14	11	11	13	17	10	11	12
E3a	15	21	10	11	13	8	12	12	14	11	11	13	18	10	11	13
E3a	15	21	10	11	14	8	11	12	14	11	11	14	18	11	10	12
E3a	16	21	10	11	13	8	11	12	14	11	12	13	17	10	11	12
ExE3	15	21	10	11	13	8	11	12	17	10	12	12	18	10	10	13
ExE3	15	22	10	11	12	8	11	12	14	10	12	13	17	9	10	12
E3a	13	22	9	11	12	8	11	13	14	11	12	13	17	10	11	12
E3a	17	21	10	11	14	8	11	12	14	12	11	13	18	10	10	12
ExE3	15	22	10	11	12	8	10	11	17	9	10	12	17	10	10	13
E3a	15	21	10	11	13	8	11	12	14	12	11	13	17	10	11	12
E3a	16	21	10	11	14	8	11	12	14	11	12	14	18	10	10	12
E3a	15	21	10	11	13	8	11	12	14	11	12	13	17	10	9	12
E3a	17	22	10	11	15	8	11	12	14	11	12	14	17	11	11	11
E3a	15	22	10	11	13	8	11	12	14	11	11	13	17	10	11	13
E3a	16	21	10	11	14	8	13	12	14	11	12	12	17	9	10	12
E3a	16	21	10	11	13	8	11	12	17	10	11	13	17	11	10	11
E3a	16	21	10	12	14	8	11	12	14	11	12	13	18	10	11	12
E3a	15	21	11	11	14	8	11	11	14	12	11	13	17	10	10	12
E3a	16	21	10	11	13	8	11	12	14	11	11	12	17	10	12	12
E3a	16	21	10	12	15			9	13	11	12	13	17	10	11	12
E3a	15	21	11	11	14	7	11	11	14	10	12	13	17	12	12	13
E3a	17	22	10	11	15	8	11	12	14	11	11	13	16	11	11	12

Turkey																
I	16	22	10	10	14	8	11	11	14	10	13	12	17	10	10	11
R1b3xR1b3	15	23	10	10	14	8	11	11	14		10	14	15	10	10	13

I	15	24	10	11	12	8	11	12	16	9	13	12	16	12	8	11
E3a	13	25	10	11	13	8	11	12	14	10	12	13	17	9	10	13
E3a	13	25	10	11	13	8	11	12	14	10	12	13	17	9	11	12
I	14	22	9	11	12	10	11	12	15	9	11	14	14	10	10	11
KxPN3	16	23	10	13	13	8	10	12	14	9	11	13	16	11	9	12
I	15	22	10	11	13	8	11	11	16	10	10	12	16	11	9	12
I	14	25	10	11	12	6	10	12	14	9	12	14	17	11	10	11
E3a	13	24	10	11	13	8	11	12	14	10	11	13	17	9	10	12
KxPN3	14	23	10	14	13	7	11	12	14	10	10	13	16	12	11	12
E3a	13	24	11	11	13	7	12	11	15	11	13	13	17	9	10	12
E3a	13	24	11	11	13	11	11	12	14	10	12	13	17	9	10	12
E3a	13	24	11	11	13	8	11	12	14	10	12	13	17	12	10	12
I	14	23	10	11	12	8	11	12	15	9	10	14	14	11	10	11
E3a	13	23	11	11	13	8	11	12	14	10	12	13	17	11	11	11
I	14	24	11	11	13	8	11	12	15	9	11	13	16	11	11	11
I	14	23	10	11	13	8	12	12	16	10	12	13	16	10	10	13
R1a	16	26	10	11	13	8	11	11	14	11	10	13	19	11	9	12
I	14	23	11	11	13	8	11		14	10	11	13	17	10	9	11
I	15	23	10	12	13	8	11	12	14	10	11	14	18	10	10	12
R1b3xR1b3	15	23	10	13	12	8	11	12	15	12	12	13	16	11	9	11
I	14	23	11	11	12	8	11	12	14	10	11	13	17	10	9	11
E3a	13	24	10	11	13	8	11	12	14	10	11	13	17	9	10	12
R1b3xR1b3	13	23	12	13	13	8	11	12	15	12	11	12	17	10	10	11
I	14	23	10	11	12	8	11	12	14	10	11	13	18	12	9	11
I	14	23	10	11	12	?	11	?	?	10	11	13	17	12	9	11
I	14	22	10	11	12	8	11	12	14	10	11	13	17	12	9	11
I	14	22	10	11	12	8	10	12	14	9	11	14	15	10	10	11
I	14	23	10	11	12	8	11	12	14	10	11	13	17	12	9	11
R1b3xR1b3	14	22	12	13	12	8	11	12	15	12	12	14	15	11	9	11
I	14	23	11	11	12	8	11	12	14	10	11	13	17	11	9	11
KxPN3	16	24	10	15	11	?	11	12	15	10	11	13	17	11	10	12
I	14	23	11	11	12	8	11	12	14	10	12	13	17	10	9	11
I	14	23	11	11	12	8	11	12	14	10	11	13	17	10	9	11
I	14	23	11	11	12	8	11	12	14	10	11	13	17	10	9	11

Appendix Section 5: Newfoundland Genotypes

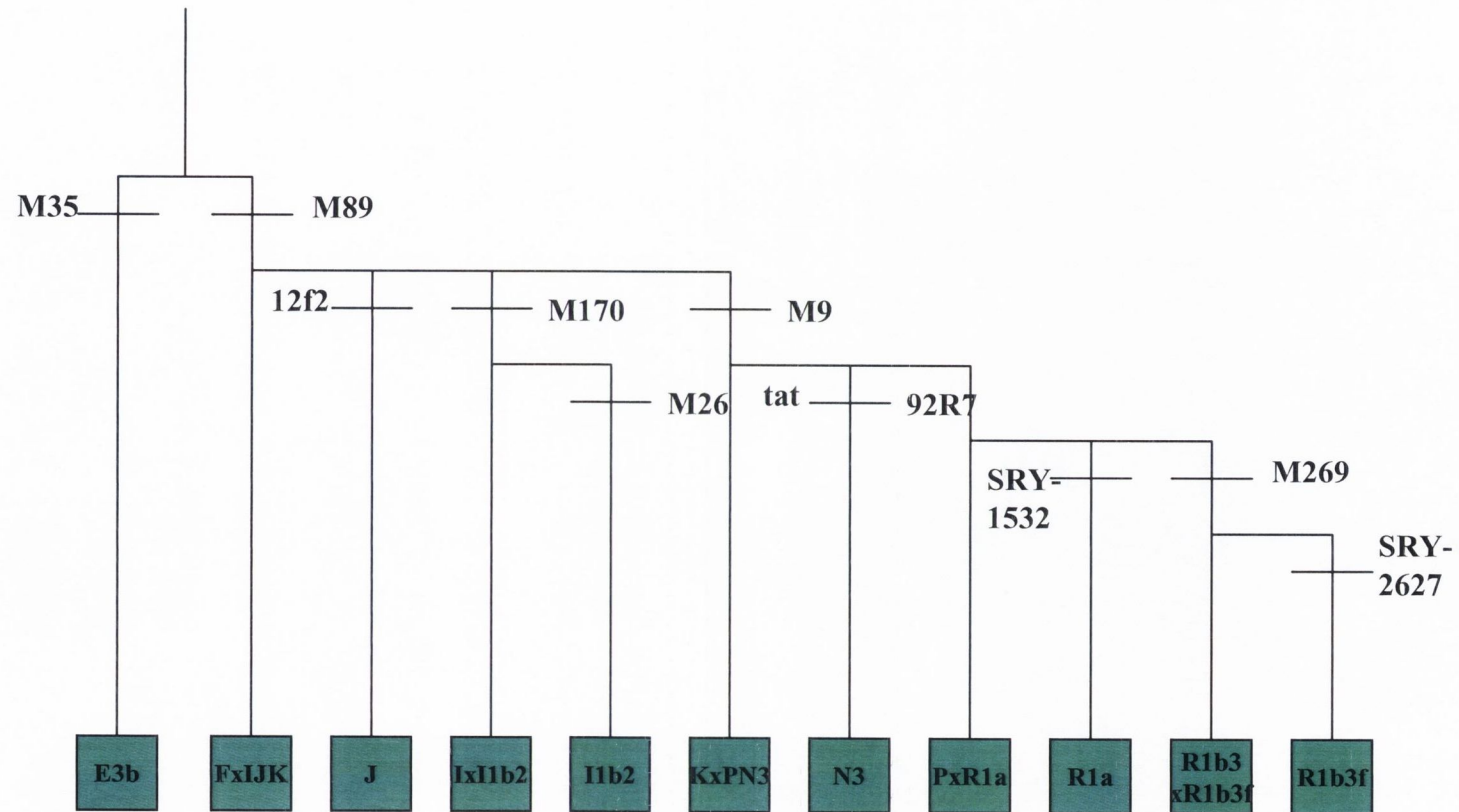
Sample No.	ETHNICITY	HG	DYS 19	DYS 388	DYS 390	DYS 391	DYS 392	DYS 393
10	English	R1b3xR1b3f	16	14	25	12	15	15
11	English	R1b3xR1b3f	15	14	24	12	15	15
17	English	R1b3xR1b3f	14	12	24	10	13	13
18	English	R1b3xR1b3f	14	12	24	10	13	13
19	English	R1b3xR1b3f	14	12	23	10	13	13
20	English	R1b3xR1b3f	15	14	24	12	15	16
21	English	R1b3xR1b3f	14	12	24	10	13	13
22	English	R1b3xR1b3f	14	13	24	10	13	13
23	English	R1b3xR1b3f	14	12	24	10	13	13
24	English	R1b3xR1b3f	14	12	24	11	13	13
26	English	R1b3xR1b3f		12	23	11	13	13
29	English	R1b3xR1b3f	15	12	24	10	11	13
40	English	R1b3xR1b3f	14	12	24	11	13	15
41	English	R1b3xR1b3f	14	12	24	11	14	13
46	English	R1b3xR1b3f	12	9	23	8	10	11
47	English	R1b3xR1b3f	16	10	24	10	12	12
48	English	R1b3xR1b3f	14	11	24	11	13	12
53	English	R1b3xR1b3f	14	12	26	11	13	13
54	English	R1b3xR1b3f	14	12	24	11	13	13
58	English	R1b3xR1b3f	15	12	23	11	13	13
65	English	R1b3xR1b3f	14	14	22	10	11	13
73	English	R1b3xR1b3f	14	12	25	10	13	13
76	English	R1b3xR1b3f	14	12	24	11	13	13
27	English	R1b3xR1b3f	15	12	24	11	13	13
7	English	lxl1b2	15	13	23	10	12	15
14	English	lxl1b2	14	14	22	10	12	13
16	English	lxl1b2	14	14	22	10	11	13
45	English	lxl1b2	16	12		12	15	13
69	English	lxl1b2	15	13	23	10	12	15
4	French	R1b3xR1b3f	14	12	24	11	13	13
13	French	R1b3xR1b3f	14	12	24	10	13	13
49	French	R1b3xR1b3f	14	11	25	11	13	11
50	French	R1b3xR1b3f	16	15	24	12	14	16
77	French	R1b3xR1b3f	14	12	23	10	13	13
79	French	R1b3xR1b3f	14	10	23	11	13	13
5	French	R1b3xR1b3f	14	12	24	11	13	13
51	French	R1b3xR1b3f	14	12	24	11	13	13
8	French	lxl1b2	14	14	23	10	11	13
1	Irish	R1b3xR1b3f	14	12	24	10	13	13
2	Irish	R1b3xR1b3f	14	12	23	11	13	13
15	Irish	R1b3xR1b3f	14	10	23	8	12	12
28	Irish	R1b3xR1b3f	14	12	24	10	13	13
31	Irish	R1b3xR1b3f	14	12	24	11	13	13
38	Irish	R1b3xR1b3f	14	12	24	10	13	13
39	Irish	R1b3xR1b3f	14	12	24	10	13	13
56	Irish	R1b3xR1b3f	13	12	24	11	13	13
63	Irish	R1b3xR1b3f	14	12	24	10	13	13
64	Irish	R1b3xR1b3f	13	13	23	9	11	13
66	Irish	R1b3xR1b3f	14	12	25	11	13	13
66	Irish	R1b3xR1b3f	15	12	24	11	13	13
67	Irish	R1b3xR1b3f	14	12	24	10	13	13
70	Irish	R1b3xR1b3f	15	14	22	10	11	13
74	Irish	R1b3xR1b3f	14	12	23	10	14	13
80	Irish	R1b3xR1b3f	14	12	25	11	13	13
25	Irish	R1b3xR1b3f	14	12	24	10	13	13
62	Irish	R1b3xR1b3f	14	12	25	10	13	13
72	Irish	R1b3xR1b3f	13	12	24	11	13	13
59	Irish	R1b3xR1b3f	14	12	24	11	13	13
68	Irish	R1b3xR1b3f	14	12	23	11	13	13
30	Irish	lxl1b2	14	12	25	11	13	13
37	Irish	lxl1b2	14	12	24	10	13	13
55	Irish	lxl1b2	15	13	24	11	11	13
60	Irish	lxl1b2	14	12	24	11	13	13
71	Irish	lxl1b2		12	23	11	13	13
78	Irish	lxl1b2	14	12	24	11	13	13
61	Irish	l1b2	14	12	25	10	13	13

Appendix Section 6: Chapter 6 Genotypes

Locus name	ATA28G05		273YF9		GATA43H03		GATA31E10		GATA73D10		GATA72E05		GAT164B08		GAT23F06	
Chromosome	15		15		13		y		7		23		3		11	
Repeat Type	3bp		2bp		4bp		4bp		4bp		4bp		4bp		4bp	
sample	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2
D1	240	240	197	199	261	261	209	-	229	233	-	-	204	224	113	137
D2	237	240	201	205	249	265	217	-	229	233	287	287	220	220	117	125
D3	240	249	201	211	249	-	213	-	225	253	283	287	212	216	109	129
D4	240	240	193	203	253	-	205	-	229	229	287	287	224	204	113	121
D5	249	237	199	199	253	261	217	-	225	253	279	283	204	204	117	125
D6	237	249	199	199	253	265	209	-	233	265	287	291	232	236	113	121
D7	237	249	201	207	257	249	209	-	217	229	279	291	208	220	113	121
D8	240	243	199	199	265	253	213	-	225	229	283	287	212	224	109	129
D9	240	240	199	201	253	257	213	-	253	257	-	-	-	-	113	125
D11	237	240	199	199	261	265	213	-	233	233	287	291	208	224	121	125
D12	-	-	201	201	261	265	213	-	233	257	283	287	212	228	121	133
D13	240	240	199	201	261	249	213	-	233	233	-	-	220	216	125	133
D15	237	237	195	201	249	261	217	-	229	233	283	287	204	208	113	117
D17	240	240	201	201	253	261	209	-	253	253	279	283	208	228	117	113
D18	234	234	201	201	253	-	-	-	253	253	287	283	224	220	109	125
D19	240	249	201	201	249	253	213	-	233	233	287	291	204	224	117	121
D20	234	246	199	207	253	253	221	-	229	229	287	287	204	224	109	113
D22	240	249	193	205	249	261	205	-	229	265	287	287	204	224	125	117
D23	240	240	193	201	253	265	205	-	217	217	283	295	204	228	109	117
D24	240	240	201	205	253	265	209	-	217	257	291	291	200	224	113	121
D25	240	249	193	199	253	265	209	-	217	229	291	291	224	228	113	117
D26	237	246	199	201	253	-	205	-	229	257	287	303	208	220	113	121
D27	246	246	201	203	249	261	205	-	237	257	283	283	224	228	105	113
D28	246	249	199	199	257	253	213	-	229	229	283	287	196	224	109	125
D29	243	249	197	201	257	273	209	-	233	233	279	283	220	232	113	129
D30	-	-	193	197	261	269	-	-	225	229	283	291	208	224	113	117
D31	249	249	201	203	261	269	213	-	229	261	283	283	208	224	117	125
D32	237	249	203	203	269	269	221	-	217	229	287	287	204	236	113	117
D33	249	249	199	207	253	-	209	-	217	257	283	283	208	228	109	125
B5	237	246	201	203	-	-	209	-	237	257	279	279	208	226	113	117
B11	240	249	199	199	253	265	209	-	237	257	279	283	208	208	113	117
B19	240	249	199	199	253	261	213	-	233	225	279	279	208	212	109	117
B29	237	240	199	199	253	269	217	-	225	237	279	291	204	224	117	125
B107	-	-	199	200	265	265	217	-	217	265	291	287	204	204	113	125
B156	240	249	201	201	249	261	217	-	241	253	287	287	208	224	125	121
B273	240	243	201	201	257	273	225	-	233	261	287	291	204	228	113	117
B295	237	249	201	201	249	253	209	-	229	257	283	283	204	224	117	133
B325	237	249	201	205	249	253	213	-	233	229	279	279	204	204	109	113
B369	237	237	203	205	265	253	213	-	229	249	275	279	204	196	109	113
B692	240	240	195	201	265	253	213	-	217	233	279	279	216	228	109	117
B711	237	237	191	201	265	265	213	-	225	257	283	287	208	232	109	129
B724	240	240	201	201	269	273	209	-	229	261	283	283	204	204	113	109
B739	237	249	203	203	269	269	221	-	233	257	-	-	204	224	117	125
B767	240	240	199	199	253	253	213	-	217	233	-	-	204	228	113	113
B884	240	240	198	198	253	253	213	-	-	-	-	-	208	228	117	125
B950	240	249	193	199	-	-	213	-	-	-	287	299	212	224	117	129
B979	237	249	201	205	-	-	213	-	265	265	287	287	220	224	129	129
B1003	237	237	193	205	253	261	213	-	257	257	287	287	208	208	113	125
B1046	237	237	199	199	-	-	209	-	257	257	-	-	204	204	113	117
E23	240	249	193	199	253	261	213	-	233	257	283	283	208	228	113	125
E25	234	237	199	201	-	-	217	-	229	233	283	283	204	224	113	125
E77	249	243	201	203	261	261	213	-	233	261	287	287	204	208	113	125
E86	237	237	199	201	-	-	217	-	217	261	287	287	204	204	109	117
E104	237	249	199	199	257	261	217	-	245	245	283	283	204	228	117	125
E121	234	237	197	199	-	-	217	-	233	257	-	-	200	224	117	129
E135	237	240	199	203	257	261	209	-	233	233	287	287	200	208	113	117
E139	234	234	201	201	-	-	217	-	233	233	-	-	204	204	113	121
E179	237	252	197	199	253	261	213	-	-	-	283	283	200	204	113	117
E200	237	249	199	203	257	257	213	-	-	-	-	-	200	204	109	117
E248	237	240	199	203	257	261	209	-	-	-	-	-	224	228	117	121
E283	243	249	199	201	261	265	217	-	-	-	-	-	224	224	117	129
E305	240	249	201	203	261	261	217	-	-	-	-	-	212	228	125	129
E366	231	240	199	199	249	257	209	-	-	-	287	287	216	228	113	113
E384	240	240	201	201	253	269	213	-	229	233	291	291	204	212	117	133
E431	237	237	203	205	253	253	213	-	229	261	287	287	200	224	117	117
E434	237	249	199	199	253	261	217	-	225	249	-	-	208	224	117	137
E439	234	252	199	201	257	269	209	-	225	225	283	287	204	228	109	113
E455	240	249	203	205	249	249	213	-	229	229	303	303	224	240	125	125
E461	240	240	201	207	253	269	213	-	229	229	283	283	204	224	113	117
E497	237	240	201	203	253	261	221	-	229	229	-	-	200	224	121	129
E585	237	240	199	199	265	269	213	-	229	253	283	291	204	224	121	129
E619	240	243	203	205	253	253	217	-	217	217	-	-	-	-	113	117
E647	237	249	199	201	253	265	209	-	233	237	283	283	204	232	113	117
E685	237	237	199	197	-	-	213	-	237	-	287	303	204	208	117	117
E704	240	240	199	201	-	-	213	-	213	233	287	287	204	224	113	117
E765	246	249	211	211	-	-	-	-	257	257	303	303	200	224	121	129
E776	240	249	201	201	265	269	213	-	229	229	-	-	220	228	113	129
E830	234	237	201	201	-	265	217	-	217	237	-	-	208	208	121	125
E842	234	246	199	201	261	269	217	-	225	229	287	287	208	228	113	129
E1000	237	240	203	203	-	-	-	-	229	233	-	-	204	208	109	125
E1016	237	249	199	203	261	265	213	-	-	-	291	291	204	224	113	117
E1025	234	249	201	203	-	-	213	-	-	-	-	-	204	204	109	121

E1040	240	249	201	203	257	261	213	-	-	-	-	-	204	224	113	113
W81	240	249	197	201	253	253	217	-	257	265	-	-	204	212	117	133
W131	249	249	201	201	-	-	217	-	217	241	283	283	204	236	113	133
W171	237	240	197	201	257	257	213	-	217	229	287	291	204	220	109	113
W300	234	237	199	201	253	253	217	-	229	241	279	287	236	224	101	113
W322	234	243	197	201	253	261	221	-	217	217	283	291	208	228	109	121
W409	234	234	199	201	253	253	221	-	217	261	283	287	208	208	113	121
W482	234	246	199	195	257	265	217	-	225	261	283	287	196	204	113	117
W567	234	246	199	199	253	265	209	-	237	261	283	283	224	232	109	117
W594	234	243	195	195	253	261	209	-	217	233	287	287	224	208	113	121
W603	237	246	199	199	253	261	213	-	217	229	283	275	204	228	113	121
W604	246	252	195	203	253	257	217	-	229	233	279	295	204	224	113	125
W627	246	252	201	205	261	265	217	-	217	257	291	275	232	236	117	125
W691	246	246	199	203	249	253	213	-	233	221	291	295	192	224	113	125
W692	246	252	201	205	261	265	217	-	225	229	291	291	204	220	105	105
W760	231	243	195	203	253	265	217	-	229	229	295	295	208	224	113	117
W783	243	231	197	199	253	265	217	-	237	261	287	291	216	224	109	113
W785	246	246	201	201	253	253	213	-	233	261	283	287	200	220	113	117
W787	234	243	203	205	253	261	217	-	221	233	275	287	200	208	109	133
W788	237	246	203	199	261	265	213	-	229	233	275	287	204	220	113	125
W791	237	237	199	193	257	261	217	-	233	261	279	287	220	224	109	129
W811	231	237	193	197	249	265	217	-	217	261	291	283	224	228	121	129
W864	234	237	199	199	253	273	213	-	229	261	279	287	204	228	113	133
W860	234	246	195	195	261	265	217	-	229	265	283	291	200	204	113	137
W867	237	246	195	197	-	-	217	-	233	237	279	295	200	216	121	125
W871	237	246	199	199	253	253	213	-	237	229	287	291	208	228	125	129
W872	234	240	195	197	265	265	213	-	237	261	283	287	208	220	113	125
W873	231	252	197	199	261	261	217	-	233	257	287	291	220	224	105	129
W874	237	231	199	199	253	257	217	-	229	257	283	287	204	220	105	113
W876	246	237	201	207	253	257	217	-	229	261	287	291	204	208	113	121
W878	246	246	201	201	253	253	221	-	229	261	283	287	224	232	113	121
W879	243	246	201	201	261	265	217	-	225	237	283	287	204	216	117	125
W881	234	234	197	199	253	257	213	-	229	257	279	283	200	204	113	121
W887	234	234	197	197	249	253	213	-	229	257	283	279	200	204	117	121
W890	243	243	203	203	257	257	217	-	225	253	279	287	232	236	117	125
W891	240	243	189	195	253	253	217	-	217	233	283	295	204	224	109	117
W892	237	246	197	199	253	261	221	-	229	237	279	291	200	224	109	117
W894	231	234	197	197	245	261	217	-	229	233	279	283	200	204	117	129
W895	237	237	201	205	253	253	217	-	229	233	279	291	204	220	105	125
W897	231	234	201	201	253	261	217	-	229	253	291	295	204	228	109	129
S1	234	234	199	205	265	261	213	-	237	257	283	287	204	208	113	125
S2	252	255	201	203	257	265	213	-	233	265	283	287	208	204	109	117
S4	252	246	197	207	257	265	217	-	229	261	279	287	224	208	109	113
S7	246	249	201	207	253	261	217	-	225	265	283	291	220	228	113	129
S9	252	246	207	207	261	253	217	-	237	257	283	287	204	220	117	125
S16	240	249	203	209	269	273	217	-	233	241	283	287	232	208	113	117
S17	231	237	195	197	249	261	209	-	249	261	283	283	208	224	109	121
S18	234	247	197	197	253	265	213	-	249	257	283	291	200	208	113	125
S20	-	-	199	199	253	265	217	-	253	217	283	291	220	224	113	117
S21	-	-	203	205	257	265	217	-	213	213	287	295	204	208	117	117
S23	234	246	201	201	265	261	217	-	217	217	283	283	200	224	117	129
S24	246	246	201	201	261	265	217	-	221	233	287	291	204	220	109	117
S27	240	249	205	207	265	269	217	-	229	237	295	295	208	224	117	109
S31	237	246	203	203	261	265	217	-	257	217	287	283	-	-	117	125
S32	237	246	207	207	261	265	217	-	229	229	-	-	-	-	117	125
S33	-	-	198	198	261	261	-	-	229	261	-	-	204	204	113	113
S34	240	249	198	198	261	253	-	-	225	257	279	283	200	204	113	117
S36	234	237	197	199	253	261	217	-	217	233	279	283	200	204	117	121
S37	240	240	201	201	257	257	217	-	253	257	283	287	204	208	117	121
S38	237	237	201	207	253	261	217	-	229	253	279	283	204	200	109	113
S39	234	252	193	201	265	257	217	-	-	-	287	287	200	204	113	117
S42	237	246	199	199	257	261	-	-	-	-	283	291	204	204	113	121
S48	240	246	199	203	253	253	217	-	-	-	287	287	224	228	117	125
S58	-	-	199	199	253	261	-	-	217	249	279	283	204	208	113	117
S60	234	234	205	205	261	257	-	-	249	249	283	283	208	204	113	117
S64	234	234	205	205	253	257	221	-	-	-	283	287	204	208	113	117
S67	234	247	199	201	253	253	217	-	217	253	283	287	220	232	129	129
S70	234	237	195	197	249	265	213	-	233	253	283	287	204	228	125	129
S73	234	249	197	197	269	257	205	-	237	253	279	283	204	204	125	129
S76	237	246	201	201	261	269	209	-	225	229	279	283	204	204	117	125
S94	237	237	205	209	261	269	217	-	217	217	283	287	224	224	-	-
S103	240	243	201	201	265	269	217	-	-	-	287	287	204	208	117	117
S104	240	243	199	201	265	265	217	-	221	233	283	287	204	208	117	121
S105	246	258	199	203	269	265	217	-	217	253	283	283	204	208	117	117
S109	258	258	203	209	265	265	217	-	217	249	279	283	200	204	-	-
S112	234	247	201	201	253	269	217	-	217	253	279	283	208	208	121	121
S123	237	237	203	203	257	253	213	-	241	261	295	299	208	208	121	121

Appendix Section 7: Partial Y Chromosome Phylogeny



Literature Cited

- Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C. (2000) Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res.* 2000 Jan 15;28(2):e8.
- Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37-48
- Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations. *Genetics* 141:743-753
- Bertranpetit J, Calafell F (1996) Genetic and geographical variability in cystic fibrosis: evolutionary considerations. *Ciba Found Symp* 197:97-114; discussion 114-8
- Bosch E, Lee AC, Calafell F, Arroyo E, Henneman P, de Knijff P, Jobling MA (2002) High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. *Forensic Sci Int* 125:42-51
- Brutsaert TD, Parra EJ, Shriver MD, Gamboa A, Palacios JA, Rivera M, Rodriguez I, Leon-Velarde F (2003) Spanish genetic admixture is associated with larger V(O₂) max decrement from sea level to 4338 m in Peruvian Quechua. *J Appl Physiol* 95:519-28
- Cann RL (2001) Genetic clues to dispersal in human populations: retracing the past from the present. *Science* 291:1742-8
- Capelli C, Redhead N, Abernethy JK, Gratrix F, Wilson JF, Moen T, Hervig T, Richards M, Stumpf MP, Underhill PA, Bradshaw P, Shaha A, Thomas MG, Bradman N,

- Goldstein DB (2003) A Y chromosome census of the British Isles. *Curr Biol* 13:979-84
- Cavalli-Sforza LL, Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. *Nat Genet* 33 Suppl:266-75
- Cavalli-Sforza LL, Minch E (1997) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 61:247-54
- Cavalli-Sforza LL, Menozzi P, Piazza A. (1994) *The History and Geography of Human Genes*, Princeton University Press, New Jersey.
- Clarke A (1988) The English. In *The People of Ireland*, ed. Loughrey P. Appletree Press, Dublin
- Collins-Schramm HE, Phillips CM, Operario DJ, Lee JS, Weber JL, Hanson RL, Knowler WC, Cooper R, Li H, Seldin MF (2002) Ethnic-difference markers for use in mapping by admixture linkage disequilibrium. *Am J Hum Genet* 70:737-50
- Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA (2002) A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* 70:1197-214
- Cunliffe B (2001) *Facing the Ocean: The Atlantic and its peoples 8000 BC - AD 1500*. Oxford University Press, Oxford.
- Davies N (1999) *The Isles: A History*. Macmillan, London
- Felsenstein J (1989) PHYLIP (Phylogeny Inference Package) Version 3.5c Distributed by the author, Department of Genome Sciences, University of Washington, Seattle.

- Flanagan M.T (1988) The Vikings. In *The People of Ireland*, ed. Loughrey P. Appletree Press, Dublin
- Foley R (1998) The context of human genetic evolution. *Genome Res* 8:339-47
- Forster P, Torroni A, Renfrew C, Rohl A (2001) Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol Biol Evol* 18:1864-81
- Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and Evolution of Native American mtDNA Variation: A reappraisal. *American Journal of Human Genetics* 59:935-945
- Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995a) An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463-71
- Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995b) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci U S A* 92:6723-7
- Goldstein DB, Zhivotovsky LA, Nayar K, Linares AR, Cavalli-Sforza LL, Feldman MW (1996) Statistical properties of the variation at linked microsatellite loci: implications for the history of human Y chromosomes. *Mol Biol Evol* 13:1213-8
- Greenberg JH (1963) *The Languages of Africa*. Indiana University Press, Bloomington IN.
- Hackett WE, Dawson GW (1958) The distribution of the ABO and simple Rhesus (D) blood groups in the Republic of Ireland from a sample of 1 in 37 of the adult population. *Ir. J. Med. Sci.* 16:99-109
- Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* 66:591-600

- Hedrick P (1999) *Genetics of Populations* pp.84. Jones and Bartlett Publishers, Sudbury Massachusetts.
- Helgason A, Sigurethardottir S, Nicholson J, Sykes B, Hill EW, Bradley DG, Bosnes V, Gulcher JR, Ward R, Stefansson K (2000) Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. *Am J Hum Genet* 67:697-717
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405:907-13
- Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Mol Genet* 6:799-803
- Hill EW, Jobling MA, Bradley DG (2000) Y-chromosome variation and Irish origins. *Nature* 404:351-2
- Hill EW (2000) *A Genetic Investigation of Irish Origins*. Ph.D. Thesis, University of Dublin
- Hurles ME, Nicholson J, Bosch E, Renfrew C, Sykes BC, Jobling MA (2002) Y chromosomal evidence for the origins of oceanic-speaking peoples. *Genetics* 160:289-303
- Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Perez-Lezaun A, Bosch E, Shlumukova M, Cambon-Thomsen A, McElreavey K, Lopez De Munain A, Rohl A, Wilson IJ, Singh L, Pandya A, Santos FR, Tyler-Smith C, Jobling MA (1999) Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. *Am J Hum Genet* 65:1437-48

- Isaac G (1976) Stages of cultural elaboration in the Pleistocene: Possible archaeological indicators of the development of language capabilities. In *Origins and Evolution of Language and Speech* Vol. 280: 275-288
- Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598-612
- Jobling MA, Tyler-Smith C (2000) New uses for new haplotypes the human Y chromosome, disease and selection. *Trends Genet* 16:356-62
- Kittles RA, Weiss KM (2003) Race, ancestry, and genes: implications for defining disease risk. *Annu Rev Genomics Hum Genet* 4:33-67
- Klein R (1999) *The Human Career*. University of Chicago Press, Chicago IL.
- Loughrey P (1988) *The People of Ireland*. Appletree Press, Dublin
- Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2:13
- MacLysaght E (1960) *More Irish Families*. O'Gorman, Galway, Dublin.
- MacNiocaill G (1972) *Ireland Before the Vikings*. Gill and Macmillan, Dublin
- Mannion J (1999) Tracing the Irish - A geographical guide. *Munster Express* December 1999
- Mannion J (1980) The Waterford merchants and the Irish-Newfoundland provisions trade 1770-1820. In *Négoce et Industrie en France et en Irlande aux XVII^e et XIX^e Siècles*. Centre National de la Recherche Scientifique, Paris
- Mathias N, Bayes M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. *Hum Mol Genet* 3:115-23

- McBrearty S, Brooks A (2000) The revolution that wasn't: a new interpretation of the origin of modern human behaviour. *Journal of Human Evolution* 39:453-563
- McEvoy B, Richards M, Forster P, Bradley DG (2004) The Long Duree of genetic ancestry: multiple genetic marker systems and the Celtic origins of the Atlantic Façade of Europe. *Am. J. Hum. Genet.* 75:693-702
- Mellars PA (1992) Archaeology and the population-dispersal hypothesis of modern human origins in Europe. *Philos Trans R Soc Lond B Biol Sci* 337:225-34
- Nei M, (1972) Genetic distance between populations. *Amer. Nat.* 106:263-282
- Otte, M. (1990) in *The World at 18,000 BP Vol. 1*, pp. 54-68, eds. Soffer, O. & Gamble, C. Unwin Hyman, London
- Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD (1998) Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 63:1839-51
- Pereira L, Gusmao L, Alves C, Amorim A, Prata MJ (2002) Bantu and European Y-lineages in Sub-Saharan Africa. *Ann Hum Genet* 66:369-78
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-59
- Quintana-Murci L et al. (2004) Where west meets east: the complex mtDNA landscape of the southwest and central Asian corridor. *Am. J. Hum. Genet.* 74:827-45
- Richards M, Macaulay V, Torroni A, Bandelt HJ (2002) In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet* 71:1168-74

- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67:1251-76
- Richards MB, Macaulay V, Sykes BC, Pettitt P, Forster P, Bandelt H-J (1997) Palaeolithic and neolithic lineages in the European mitochondrial gene pool: Reply to Cavalli-Sforza and Minch. *American Journal of Human Genetics* 61:247-255
- Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt HJ, Sykes B (1996) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59:185-203
- Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, et al. (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet* 75:128-37
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. *Science* 298:2381-5
- Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, et al. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 67:1526-43
- Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67:718-726

- Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, Macaulay V, Carracedo A (2002) The making of the African mtDNA landscape. *Am J Hum Genet* 71:1082-111
- Schneider, S., Roessli, D., and Excoffier, L. (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva
- Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. *Nat Genet* 20:278-80
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290:1155-9
- Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS (1996) A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am J Hum Genet* 59:964-8
- Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, Stoneking M (1994) Mismatch distributions of mtDNA reveal recent human population expansions. *Hum Biol* 66:761-75
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457-62
- Stringer C (2003) Out of Ethiopia. *Nature* 423:692-695
- Stringer C (2000) Coasting out of Africa. *Nature* 405:24-27

- Sullivan KM, Mannucci A, Kimpton CP, Gill P. (1993) A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *Biotechniques*. 15(4):636-8, 640-1.
- Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, le Roux M, Bradman N, Goldstein DB (2000) Y chromosomes traveling south: the cohen modal haplotype and the origins of the Lemba--the "Black Jews of Southern Africa". *Am J Hum Genet* 66:674-86
- Thomas MG, Bradman N, Flinn HM. (1999) High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome. *Hum Genet*. 105(6):577-81.
- Thorne A, Grun R, Mortimer G, Spooner NA, Simpson JJ, McCulloch M, Taylor L, Curnoe D. (1999) Australia's oldest human remains: age of the Lake Mungo 3 skeleton *J Hum Evol*. 36:591-612
- Tighe O, Dunican D, O'Neill C, Bertorelle G, Beattie D, Graham C, Zschocke J, et al. (2003) Genetic diversity within the R408W phenylketonuria mutation lineages in Europe. *Hum Mutat* 21:387-93
- Torrioni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, et al. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 69:844-52
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65:43-62

- Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7:996-1005
- Veitia R, Ion A, Barbaux S, Jobling MA, Souleyreau N, Ennis K, Ostrer H, Tosi M, Meo T, Chibani J, Fellous M, McElreavey K (1997) Mutations and sequence variants in the testis-determining region of the Y chromosome in individuals with a 46,XY female phenotype. *Hum Genet* 99:648-52
- Waddell J (2000) *The prehistoric archaeology of Ireland*. Galway University Press, Galway
- Warren L (1988) The Normans. In *The People of Ireland*, ed. Loughrey P. Appletree Press, Dublin
- Watson E, Forster P, Richards M, Bandelt HJ (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61:691-704
- Weale ME, Weiss DA, Jager RF, Bradman N, Thomas MG (2002) Y chromosome evidence for Anglo-Saxon mass migration. *Mol Biol Evol* 19:1008-21
- Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, et al. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci U S A* 98:10244-9
- White T.D (2003) Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* 423:742-747

- Whitfield LS, Hawkins TL, Goodfellow PN, Sulston J (1995) 41 kilobases of analyzed sequence from the pseudoautosomal and sex-determining regions of the short arm of the human Y chromosome. *Genomics* 27:306-11
- Wilder JP, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF (2004) Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nat Genet.* Sep 19 [Epub ahead of print]
- Wilson I, Weale M, Balding D (2003) Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 166: 155-188.
- Wilson I, Balding D (2002) Bayesian Analysis of Trees With Internal Node Generation
<http://www.maths.abdn.ac.uk/~ijw>
- Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MG, Bradman N, Goldstein DB (2001) Population genetic structure of variable drug response. *Nat Genet.* 29:265-9
- Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, Goldstein DB (2001) Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci U S A* 98:5078-83
- Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 12:339-48.
<http://ycc.biosci.arizona.edu>
- Ye S, Dhillon S, Ke X, Collins AR, Day IN (2001) An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res* 29:E88-8

- Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* 60:1174-83
- Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P, Yuldasheva N, Ruzibakiev R, Xu J, Shu Q, Du R, Yang H, Hurles ME, Robinson E, Gerelsaikhan T, Dashnyam B, Mehdi SQ, Tyler-Smith C (2003) The genetic legacy of the Mongols. *Am J Hum Genet* 72:717-21
- Zhivotovsky LA, et al. (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population divergence time. *Am. J. Hum. Genet.* 74:50-61