Genetic Investigation of Irish Ancestry and Surname History

Brian Patrick McEvoy

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

Department of Genetics
University of Dublin
Trinity College

September 2004

DECLARATION

I hereby certify that this thesis, submitted to the University of Dublin, Trinity College, for the degree of Doctor of Philosophy, has not been presented for examination at this or any other university. The research described herein is entirely my own unless otherwise stated and acknowledged.

This thesis may be made available for consultation within the university library and may be copied or loaned to other libraries for the purposes of consultation.

_____

Brian McEvoy
September 2004

ACKNOWLEDGEMENTS

"No man is an island entire of itself; every man is a piece of the Continent, a part of the main". Although John Donne may not have been thinking about the PhD process when he wrote this in 1624, his words sum up my feelings on the great group of people whose support and encouragement make this work possible.

First and foremost to Dan Bradley, for giving me the opportunity to work in this field and, what's more, making sure I ploughed it thoroughly. His patient direction and creativity were crucial and this thesis would simply not have been possible without his input. A big thank you to each and every member of the lab, past and present: Caitriona, Elaine, Emma, Dave, David, Jill, Laoise, Stephen and Valeria, for their never-ending willingness to help, and for their friendship, which always made the lab such a great place to be. It was a privilege and pleasure to work with you all. A special word of gratitude to Ruth for all her help over the past four years, and to Ceiridwen for always offering the right advice at the right time. In addition, thanks to all the temporary members of the lab over the years, particularly Claire and Mindy for their part in the Viking work. My appreciation also to the members of the wider TCD Genetics Department for their help; I am particularly grateful to Professors Ken Wolfe and David McConnell, Sue, Dave, Brenda, Paul and everyone in the Gill and Wolfe labs, too numerous to mention everyone but your individual help was always appreciated.

The National Millennium Committee, through the Royal Irish Academy, and the generosity of Patrick Guinness and Joseph A. Donohoe, funded this research. The interest, enthusiasm and help of Patrick and Joseph were appreciated and it was always a great encouragement to know there was a willingness to support this field of research, in every respect. I pay particular tribute to Patrick Guinness in this regard.

I was very fortunate over the past few years to interact with the wider population genetics community and never ceased to be amazed by their willingness to help. The input of Peter Forster, and especially Martin Richards, in the mtDNA work is gratefully acknowledged and appreciated. Thanks also to all my family and friends for their support over the past few years, especially my parents, for keeping a long-term student, and Deirdre and Mike, for feeding and "watering" me!

Last and by no means least, a huge word of appreciation to all the volunteers who donated DNA samples and without whom this research really would not have been possible.

Go raibh míle maith agaibh go léir !

CONTENTS

FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

ABBREVIATIONS

°C	Degrees Celsius
µM	Micromolar
µg	Microgram
µl	Microlitre
6'FAM	6-Carboxy fluorescein
AMOVA	Analysis of molecular variance
AD	*Anno domini*
AMH	Atlantic Modal Haplotype
ASD_0	Average square distance
BC	Before Christ
bp	Base pairs
ca.	*circa*
CEPH	Centre d'Etude du Polymorphisme Humain
CGF	Classical gene frequencies
CI	Confidence/Credible Interval(s)
cM	Centimorgan
CRS	Cambridge Reference Sequence
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribose nucleic acid
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
EDTA	Ethylene-diamino-tetra-acetic acid
HEX	4,7,2',4',5',7' – Hexachloro-6-carboxyfluorescein
Hg	Haplogroup
HVS-1	Hyper-Variable Segment 1
HVS-2	Hyper-Variable Segment 2
HWE	Hardy-Weinberg equilibrium
IMH	Irish Modal Haplotype
INDEL	Insertion/Deletion
kb	Kilobase(s)
KCl	Pottasium Choloride
Km	Kilometre
LGM	Last Glacial Maximum
M	Molar
mb	Megabase(s)
MCMC	Markov chain Monte Carlo
MDS	Multidimensional Scaling
$MgCl_2$	Magnesium Chloride

MJ	Median-joining
m	Admixture proportion
ml	Millilitre
mM	Millimolar
mtDNA	Mitochondrial DNA
n	Sample population size
ng	nanogram
NA	Not applicable
NaCl	Sodium Chloride
NSG	Norse Surname Group
p	Probability
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
pg	Page
RE	Restriction Enzyme
RFLP	Restriction Fragement Length Polymorphism
ROX	6-Carboxy-X-rhodamine
RPM -	Revolutions per minute
SD	Standard Deviation
SDS	Sodium dodecyl sulphate
SMM	Stepwise Mutation Model
SNP	Single Nucleotide Polymorphism(s)
STR	Short Tandem Repeat
TAMRA	N',N',N',N',- Tetramethyl-6-carboxyrhodamine
TCD	Trinity College Dublin
TE	Tris EDTA
TMRCA	Time to the Most Recent Common Ancestor
TRIS	Tris (hydroxymethyl) aminomethane
Tris-Hcl	Tris – Hydrochroride
U	Units
U/C	Unclassified
UEP	Unique Event Polymorphism(s)
UK	United Kingdom
UV	Ultra Violet
V/V	Volume by Volume
VNTR	Variable Number Tandem Repeat
W/V	Weight by Volume
WTTI	Weighty Transitions to Transversions/Indels
YAP	Y *Alu* Polymorphism
YBP	Years Before Present

SUMMARY

Several genetic systems were employed in this study to investigate Ireland's pre-history and history. Initially Irish origins were examined using mtDNA sequence variation in 200 new samples coupled with an extensive novel analysis of pre-existing European mtDNA, Y chromosome and autosomal diversity. The same prominent trend is recorded in loci with different inheritance modes and indicates a shared genetic ancestry among the people of the European Atlantic Façade. The origins of this probably lie in the post glacial recolonisation of Western Europe and Ireland from an Iberian Ice age refugium and its preservation precludes complete or substantial population replacement by central European Celtic invaders during the Iron Age. Additional examination of longitudinal east-west Irish autosomal diversity using classical genes and 380 STRs detected little evidence of the strong intra-Ireland genetic differentiation observed in the Y chromosome. This is in agreement with mtDNA and potentially points to the importance of male-specific behaviours, rather than preferential East-coast migration, in generating the Y chromosome pattern. Next, the extent of patrilineal coinheritance of Y chromosomes and Irish surnames was explored and then used to investigate the specific origins and histories of several surnames. Analysis of Y chromosome diversity in 1105 men using a combination of fast evolving STR and stable UEP markers, in both local and general Irish contexts, demonstrates that Irish surnames collectively reflect real and recent shared paternal ancestry. The extent of this varies between surnames reflecting basic differences in the number of early founders. Furthermore, modern surname populations are invariably a mixture of numerous paternal ancestries of varying legacy, indicating complex and continuously evolving post-foundation histories. While these results are in broad agreement with historical accounts, there is nonetheless a valuable role for a molecular Y chromosome approach in the elucidation of individual surname history. In a variation on this approach, Y chromosome diversity in 47 Irish men with a putative Viking/Norse surname origin was used to explore the nature and extent of Viking settlement in Ireland. Extensive admixture analysis detected little or no Scandinavian ancestry indicating a substantial Norse cultural impact in the absence of a large number of Scandinavian migrants; an observation perhaps best explained by an elite dominance model of Viking settlement in Ireland.

CHAPTER 1

GENERAL INTRODUCTION

1.0 GENERAL INTRODUCTION

1.1 DNA and the Human Past

Genetic variation in modern humans populations has been influenced and shaped by the events of our past. Ultimate origins, later migrations as well as population growth and contraction, are just some of the events that may be reconstructed from an exploration of current genetic diversity. Furthermore, by virtue of different inheritance modes, the human genome combines multiple sources, each of which provides a unique and potentially different perspective on the human past. Biparentally inherited autosomal DNA, comprising 22 chromosomes of the cell nucleus, forms the vast majority of the ~3.2 billion base pairs (bp) of the human genome. However, these chromosomes are constantly reshuffled by recombination during meiosis making it difficult to clearly trace specific segments (haplotypes). For this reason, inference from autosomal DNA has tended to be limited to population level comparisons through allele frequencies, but which nonetheless provide a valuable overview of human relationships.

Although also found in the nucleus, the distinct inheritance of the sex, X and Y, chromosomes have ensured important roles, particularly for the latter, in the reconstruction of population history (Jobling and Tyler-Smith 2003; Schaffner 2004). The Y chromosome is only ~60 megabases (mb) in size and mostly composed of 'junk' or non-coding DNA, yet its presence is crucial in male sex determination; the alternative of two X chromosomes results in female development. Consequently, the X chromosome spends most (two thirds) of its history in females while the Y chromosome is necessarily only found in men. As most of the Y chromosome is also recombination free, it passes intact from father to son through the generations giving a solely male perspective to human population history. Lack of recombination also means that Y chromosome diversity can be compounded into single highly informative haplotypes

allowing direct reconstruction of relationships between individuals. Furthermore, uniparental inheritance results in an effective population size one quarter that of autosomal loci, leading to greater divergence between populations from the increased impact of random frequency changes.

In contrast to the Y chromosome, the mitochondrial DNA (mtDNA) of any individual is derived from their direct female ancestor. Matrilineal inheritance allows mtDNA to be used as an important complement to the Y chromosome in an alternative and/or parallel exploration of human population history. Apart from inheritance mode and extra-nuclear location, it shares many features of the Y chromosome including lack of recombination and a relatively small effective population size. In addition, mtDNA shows a high mutation rate compared to the nucleus, which is especially concentrated in the non-coding 'control region'. Despite comprising just 16.5 kilobases (kb) of DNA, these features ensure high diversity and a role in the population genetics out of all proportion to its meagre size.

1.2 Exploring Genetic Diversity

The role of genetics in elucidating the human past is inherently based on the presence and detection of diversity between individuals or, more broadly, between groups of people (populations). Early attempts to draw such inference focused on gross phenotypic difference such as skin colour. However, many of these traits are influenced by extrinsic environmental factors. The discovery of the first protein polymorphisms *circa* 1900 (ABO blood group system) provided an important new source of variation, which was more closely linked to underlying genetic diversity, for investigating relationships. These polymorphisms, typically encoded by the autosomal DNA, were plentiful and easy to detect and consequently became the workhorse of population

genetics for much of the 20th century (Cavalli-Sforza and Feldman 2003). However as these 'classical genes' are based on phenotypic difference, they are potentially subject to selective pressure. Ideally, the investigation of the human past should be conducted using markers whose diversity is purely a product of the events, for example time since divergence or migration, which human evolutionary genetics seeks to reconstruct.

The call for this 'neutral' variation, which is invisible to selection, was answered by the discovery of direct DNA diversity in non-coding regions of the genome, coupled with parallel advances in the technology, such as the polymerase chain reaction (PCR), needed to routinely type them (Cavalli-Sforza 1998). These DNA sequence polymorphisms can generally be divided into those that differ in character and those that differ in size. Character diversity or single nucleotide polymorphism (SNP) results from the mutation of individual base pairs. Due to a generally low mutation rate, each such event is only likely to have occurred once and they are therefore often described as unique event polymorphism(s) (UEP). Such differences are not uncommon and it is estimated that one in every 1250 bases (about 2.6 million sites) differ between any two humans (Sachidanandam 2001; Venter 2001). However, in practice most of these are rare and/or dispersed widely across the genome. Therefore, direct and continuous sequencing is only efficient in areas of concentrated mutation, such as the control region of mtDNA, and elsewhere sites are best examined individually using various PCR based methods (reviewed by Syvänen 2001).

The second major group of DNA polymorphism are size differences resulting from the relative deletion or insertion (Indels) of one or more base pairs. The major class of these are simple DNA sequences repeated in a variable number series, termed variable number tandem repeats (VNTR). Microsatellites or short tandem repeats (STR)

composed of 1-6 base pair subunits are perhaps the most widely employed in population genetics. These tend to display higher mutation rates than SNPs and therefore often have a wide range of alleles (multiallelic states) providing increased discriminatory power in the reconstruction of relationships. The exact mutational process underlying these markers is uncertain but is often approximated by the stepwise mutation model (SMM) (reviewed by Ellegren 2004). Under this model, each mutation event is equally likely to result in the gain or loss of a single repeat unit at a rate independent of allele size. Other VNTR markers are composed of larger repeat units, which is expressed in their designations as minisatellites and macrosatellites. Finally, size differences can also results from the presence or absence of transposable element insertions. These are repetitive but not arranged in a continuous series. Examples include the *Alu* element family. Since each insertion event occurred once in any location they represent a class of stable UEP markers.

The goal of human evolutionary genetics is the identification of events from our past using DNA diversity. However, a full comprehension of these often requires at least some knowledge of their chronology. The accumulation of neutral polymorphisms is a function of time. Therefore, information on mutation rate in combination with observed diversity can be used to estimate the time to the most recent common ancestor (TMRCA) of non-recombining lineages like the Y chromosome and mtDNA. However, the 'molecular clock' can only provide approximate dates, which should be interpreted cautiously owning to the high uncertainty in its initial calibration.

1.3 Ancient DNA

The direct examination of genetic diversity from the time and place of interest has an obvious attraction in exploring the past. Despite initial high hopes, the use of 'ancient DNA' presents a number of practical problems. The most obvious of these is the post-mortem decay of DNA, which places an inherent time limitation on this avenue, theoretically no greater than a million years and practically much less (reviewed in Hofreiter et al. 2001; Pääbo et al. 2004). However, substantially shorter survival times would still encompass most of the anatomically modern human past. Unfortunately, even when DNA can be retrieved from ancient human biological material, it is often degraded and consequently acutely sensitive to contamination with external DNA.

Nonetheless, some ancient DNA studies have made important contributions to resolving questions from the human past. One of these concerns the fate of pre-*Homo sapiens* hominid species. mtDNA from seven individuals of one such group, the Neanderthals of Europe, has been successfully retrieved and found to be quite distinct from those of modern humans (Krings et al. 1997; Ovchinnikov 2000; Schmitz et al. 2002; Serre et al. 2004). These findings seem to indicate that *Homo sapiens* did not substantially interbreed with these species, but rather out-competed and replaced them. However, the results also illustrate some limitations of ancient DNA analysis. Multicopy mtDNA stands the best chance of preservation but this can only provide a female population perspective. Secondly, the small sample sizes, imposed by stringent technical demands to ensure authenticity, make broad based general conclusions difficult. It is also impossible to rule out some gene flow between Neanderthals and humans as genetic drift may have randomly erased the known Neanderthal mtDNA sequences from the modern *Homo sapiens* population.

However, where the historical question is specific and biological remains more recent, forensic 'ancient' DNA can be useful. For example, it successfully confirmed the discovery in 1991 of the remains of the last Russian Tsar, Nicholas II (1868-1918) and his family (Gill et al. 1994; Ivanov et al. 1996).

1.4 Reconstructing the Past from the Present

The vast majority of insights into the human past from genetics have come from analysis of current populations. The most fundamental concerns the earliest origin of *Homo sapiens*, a question that has been a source of considerable debate. The 'multiregional hypothesis' proposes that modern humans evolved, with gene flow, from several older regional *Homo* species. The extreme alternative 'Out of Africa' theory proposes a single, more recent, origin for *Homo sapiens* in Africa. The phylogeny or relationship of modern humans from both mtDNA and Y chromosomes show a stark divide between African and non-African lineages, with the deepest branches and highest diversity found in Africa. Furthermore, the TMRCA estimates for worldwide human mtDNA and Y chromosome lineages are relatively recent, within the past 200,000 years (Ingman et al. 2000; Thomson et al. 2000). These observations are most consistent with the 'Out of Africa' model and suggest that modern humans arose here and migrated across the rest of the world to substantially replace earlier species; a conclusion in some agreement with ancient DNA analysis of Neanderthals remains, but far broader based.

Similar approaches have also had an important role in exploring more recent and/or regionally specific aspects of the human past and history. Europe is perhaps the best-surveyed continent and illustrates the full range and versatility of multiple genetic systems and their various marker types in addressing these questions. It seems clear that *Homo sapiens* first populated the continent about 40,000 years before present (YBP) but it is less certain what proportion of modern Europeans descend from these Palaeolithic

settlers. Early work using classical gene frequencies detected a prominent genetic gradient from Southeast to Northwest Europe. Although consistent with the arrival of people in Europe from the Middle East, it does not in itself indicate when this occurred. Nonetheless, the effect was attributed to migrants bringing agriculture to Europe (Neolithic transition) starting 10,000 YBP; a proposal termed the 'demic diffusion model' of agricultural spread (Ammerman and Cavalli-Sforza 1984). Others describe the Neolithic transition as a culturally mediated process that occurred without large-scale migration or population replacement (Zvelebil 2000). Under this scenario, the most prominent trend in European genetic variation was related to the earlier peopling of the continent. Later surveys of maternal mtDNA and paternal Y chromosomes supported that conclusion, with about 80% of modern Europeans descended from the continent's early Palaeolithic inhabitants (Richards et al. 2000; Semino et al. 2000). The debate continues and while it is an important aspect of the European past, it is clear from the many other genetic trends and patterns (Cavalli-Sforza et al. 1994; Rosser et al. 2000) that current populations are a composite of multiple events and processes, each of varying significance in different local or regional areas of Europe.

1.5 The Male and Female Perspective

The short history of Iceland illustrates both the local importance of specific events and the possibility for a divergence or asymmetry in the male and female genetic legacies of a single population. Iceland was uninhabited prior to the settlement of ScandinavianVikings in the 9th century AD. However, historical sources suggest that some of the early female settlers originated in Britain and Ireland. In agreement, most maternally inherited Icelandic mtDNA lineages appear to have British Isles ancestry with the remainder atrributed to Scandinavia. (Helgason et al. 2000a; Helgason et al. 2001). In contrast, a large majoirty of Icelandic Y chromosomes are descended from

Norse/Scandinavian settlers (Helgason et al. 2000b). Similar but more extreme examples are found in South America where the Y chromosome pool is often highly influenced by Europeans while maternal lineages still retain a substantial legacy from the pre-Columbus (1492) indigenous population (for an example see Carvajal-Carmona 2000 et al.). The asymmetry proabably results from the dramatic reproductive variance between European and Amerindian males during the conquests of South America in the 16^{th} and 17^{th} centuries. Conversely, female sex-specific behaviours may also shape differential genetic legacies. A greater ability of women to cross physical or cultural distances may be behind the smaller inter-population genetic distances that are sometimes observed in mtDNA relative to autosomes and particularly Y chromosomes (Bamshad et al. 1998; Seielstad et al. 1998; Pérez-Lezaun et al. 1999). Although the global extent of this pattern has been questioned (Wilder et al. 2004), at least on a local scale the practice of patrilocality in many societies, where women move to the natal home of their partners, may result in relative homogenity of mtDNA. Direct studies of genetic distances in patrilocal versus matrilocal cultures in Thailand seems to confirm the impact of this social phenomenon, and the differential migration rates it engenders, on mtDNA and Y chromosome diversity (Oota et a. 2001).

1.6 Genetics and Hereditary Social Traditions

In addition to addressing broad questions of population level origin and affinity, it has proved possible to use a genetic approach with greater finesse to investigate within population social or cultural hereditary traditions. Hindu Indian society, for example, is divided into a hierarchy of upper, middle and lower ranking castes, with movement between ranks restricted. Genetic evidence supports some differentiation between these caste groupings in mtDNA, Y chromosome and autosomal DNA. The difference is greatest in the Y chromosome, possibly reflecting the greater upward mobility of

females through marriage. Furthermore, the Y chromosomes of caste members particularly upper echelons, are more similar to Central Asians than 'tribal' Indian populations, which are thought to represent the aboriginal inhabitants of the Indian sub-continent prior to the arrival of Indo-European speakers about 3,500 YBP. In line with the dominance suggested by language replacement, migrants associated with these languages appear to have anointed themselves to the highest castes or perhaps were responsible for establishing the entire system (Bamshad 1998; Bamshad 2001; Cordaux 2004; Wooding 2004).

The Jewish priesthood or *Cohanim* is traditionally restricted to those with direct patrilineal descent from Aaron, brother of Moses, who putatively lived *ca*. 3000 YBP. As in many societies, Jewish surnames are patrilineally inherited and consequently membership of the *Cohanim* is often indicated by the 'Cohen' surname or related variants like 'Kane' or 'Kahn'. The Y chromosomes of this group show the striking frequency of one diversified microsatellite haplotype with a TMRCA of 2,650 YBP; both observations consistent with the historical account (Thomas et al. 1998). The patrilineal coinheritance of Y chromosomes and surnames has also been used to resolve the minor historical question of whether Thomas Jefferson (1743-1826), 3rd President of the United States, fathered any children by his slave Sally Hemings. Y chromosomes from current patrilineal relatives, indicated by the carriage of the Jefferson surname, were found to match those found in descendents of one of Hemings' sons (Eston Hemings Jefferson) supporting the historical oral tradition (Foster et al. 1998).

These examples together with those from previous sections illustrate both the depth and width to the role of modern population genetic diversity in addressing questions about our past. These applications range from our ultimate species origin perhaps 200,000 years ago to individual historical curiosities from little more than two centuries ago.

1.7 A Brief History of Ireland

The following summary of Ireland's past is based on the account given by Duffy (2000) in his Concise History of Ireland.

During the last glacial maximum (LGM), about 20,000 YBP, most of Ireland was covered by ice and uninhabited. The first known settlers arrived during the Mesolithic period about 9,000 YBP, following climatic warming and glacial retreat. They possibly came from Britain *via* a land bridge, that was later inundated by rising sea levels, and seem to have lived as small bands of hunter-gathers. Evidence from pollen records of forest clearance and the appearance of crops about 6,000 YBP heralds the arrival of the Neolithic period and agriculture in Ireland. It necessarily brought a sedentary way of life and probably facilitated a substantial population increase, both features indicated by extensive archaeological excavations in the 'Céide fields' of County Mayo. This period also witnessed the construction of Megalithic monuments of various types but perhaps most impressively represented by Newgrange and the other passage tombs of the Boyne Valley.

In the traditional archaeological sequence, the Bronze Age begins about 4,400 YBP with the arrival of metalworking and Beaker pottery. The latter shows similarities with other areas of the Atlantic seaboard such as Brittany suggesting some contact. However, as with the Neolithic revolution, the appearance of new and/or exogenous material cultures does not inherently indicate matching substantial immigration. Metal technology advanced further approximately 2,500 YBP with the appearance of Iron. The occurrence of a distinctive central European 'La Téne' art style in Ireland about this time has been posited to reflect the invasion of the island by the Celts of central Europe. The introduction of the Indo-European Irish language and related Celtic languages to other parts of the British Isles is often attributed to this event. However, there is little

archaeological evidence to support such a dramatic invasion and substantive population replacement scenario.

The earliest historical accounts of Ireland come from Greek and Roman writers in the early centuries of the first millennium AD. However, the true Irish historical era begins with the arrival of Christianity and writing around the 5th century AD. Ireland at this time, and for some time after, was not a unified state but rather a patchwork of local populations groups and kingdoms arranged in a shifting hierarchy. Some of these are reputed to have established colonies in Scotland, Wales and Cornwall in the middle of the first millennium AD following the collapse of the Roman Empire in Britain. In the late 8th century AD, the first attacks by Scandinavian Viking raiders are recorded. In succeeding centuries, they became settlers and appear to have integrated well with the native Irish, establishing the first true towns including Dublin, Cork and Waterford.

The Anglo-Norman invasion of Ireland, beginning in 1169 AD, was to be of more lasting and dramatic impact. Large parts of the island came under English control leading to the arrival of British settlers and radical changes in social structure. However, from about 1300 AD, the area under direct control begin to contract in the face of Irish attacks and a strong tendency of earlier colonists, now referred to as 'Anglo-Irish', to adopt Irish culture and language. Direct English administration effectively became restricted to an area around Dublin known as 'The Pale'. Renewed interest in Ireland under the Tudors led to the gradual and eventual reestablishment of English rule over the entire island. There were organised parallel attempts to introduce British settlers through the 16th century plantations of Laois/Offaly in Leinster and other areas in Munster (**Figure 1.1**). The total collapse of the 'Gaelic order', or native society and government, in the early 17th century resulted in further substantial plantations,

particularly in Ulster. Later Cromwellian and Williamite campaigns cemented English control. Thereafter followed strong and concerted pressure to emulate the colonial power through cultural anglicisation manifest, for example, by the abandonment of the Irish language. The 19th century 'Great Famine' hastened the language's decline as the Western areas hardest hit by mortality and emigration were also coincident with Irish language strength. A dramatic population decrease, estimated at 8.2 million to 5.95 million, in the six-year period between 1845-1851 was the harbinger of long-term population decline that was only recently reversed. The later decades of the 19th century witnessed resurgent nationalism, expressed first in demands for local autonomy ('Home Rule') and then full independence, which was partially realised in 1922.

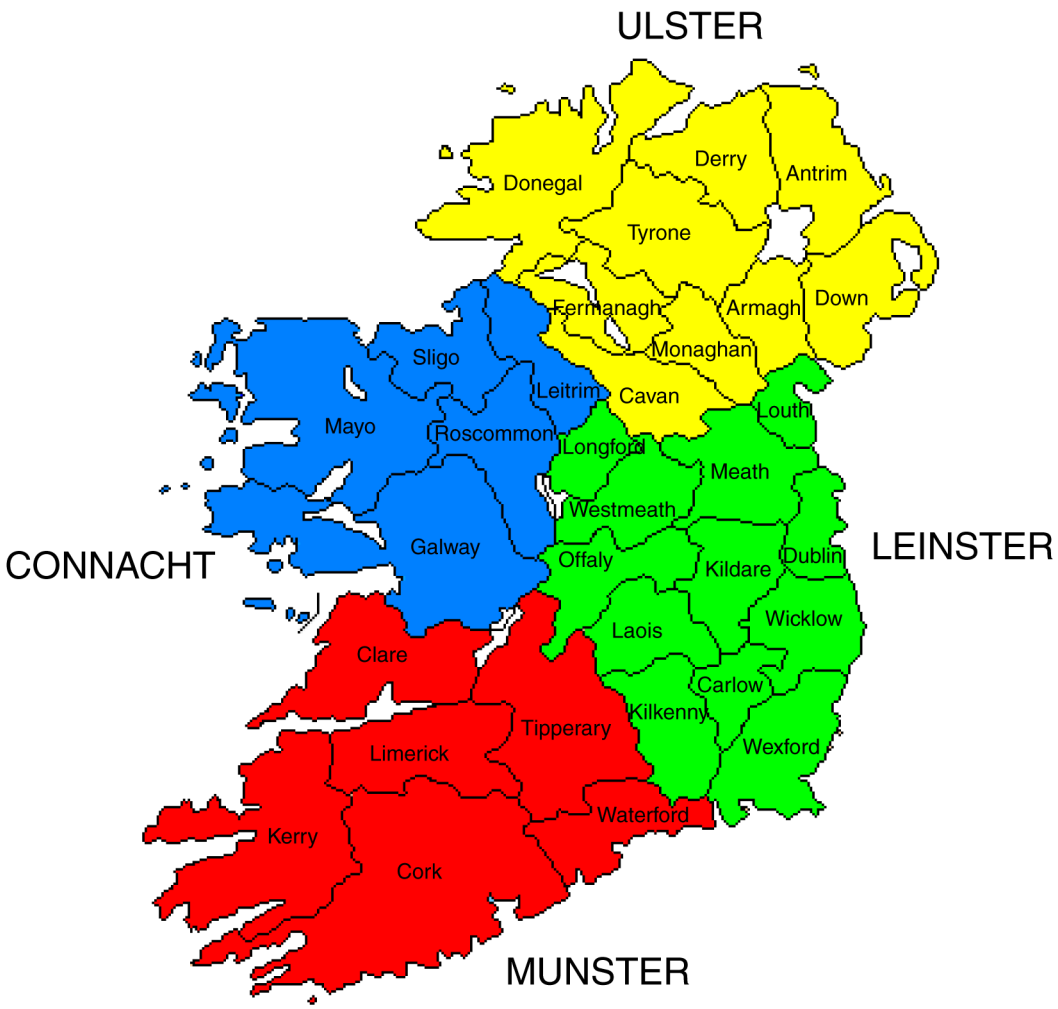


Figure 1.1 *Map of the Historical Counties and Provinces of Ireland.*

1.8 Genetic Variation and Irish Origins

Previous studies of genetic variation in the Irish population have made a significant contribution to the understanding of early Irish origins. Large-scale investigations of the ABO and other blood group (classical gene) frequencies were conducted during the 1950s and 1960s (Hacket and Dawson 1958; Dawson 1964). These revealed a high frequency of the blood group O allele, which fitted into a broader European gradient of increasing frequency of this allele from southeast to northwest. The trend continues within Ireland from east to west and is conversely reciprocated by decreasing frequency of the blood group A allele in the same direction. In addition, a similar longitudinal intra-Ireland pattern was observed in the distribution of rhesus positive/negative blood group alleles.

More recent molecular examination of Irish Y chromosomes has echoed these findings. As in many societies, Irish surnames are patrilineally inherited. They are also of early origin with some dating back to the 10^{th} century, while many also display a strong geographic specificity. Hill et al. (2000) took advantage of these features to assign each linked Y chromosome a thousand year old place of origin and in so doing demonstrated substantial differences between Y chromosomes from the eastern province of Leinster compared to the western Connacht region. The high frequency of haplogroup R1b Y chromosomes in Gaelic or indigenous Irish surnames and its phylogenetic coherence together suggests that they represent the paternal genetic legacy of early Irish inhabitants. Furthermore, their uneven distribution between East and West Ireland is the continuation of a wider cline in increasing R1b frequency from Southeast to Northwest Europe. This type also reaches very high frequencies in the Basque population of Northeast Spain and Southwest France (Rosser et al. 2000). The Basques are often thought to represent something of the Palaeolithic or pre-agricultural European population as they retain a non-Indo European language and an unusual genetics. The

findings from both the Y chromosome and autosomal classical genes like the ABO blood group, strongly suggested that the modern Irish population is predominantly descended from the earliest Palaeolithic settlers of the island. Furthermore, the preservation of this genetic signature indicates that Ireland has been relatively undisturbed by subsequent population movements in Europe. The east-west intra-Ireland variation may reflect the arrival and association of any later migrants with eastern regions of the island, which are more accessible to Britain and, *via* here, the continent.

1.9 Present Study

Modern genetic diversity has proved an important source in the elucidation of various aspects of the human past, ranging from ultimate origins to the socio-cultural processes of recent times. In this work, different genetic inheritance systems and marker types are used to explore facets of the pre-history and history of the people of Ireland. Initially, the earliest origins of the Irish are investigated using maternally inherited mtDNA placed in a broad European context and complemented by concurrent surveys of Y chromosome and autosomal diversity. The extent and impact of subsequent migrations to Ireland is then examined through east-west longitudinal variation in the autosomal genome using classical genes and several hundred STR loci. Later, the patrilineal coinheritance of surnames and Y chromosomes is extensively examined to learn about this relatively recent aspect of Irish cultural history. Both the general extent of common paternal ancestry underlying modern surname and the origins and histories of specific surnames are explored. Finally, surnames are used as markers of ethnic affinity and origin to assess the legacy of Scandinavian Viking settlement on the island. Collectively, it is hoped that the novel perspective afforded by this molecular genetic investigation will make a valuable contribution to our understanding of the Irish past.

CHAPTER 2

GENETIC DIVERSITY ON THE ATLANTIC FAÇADE OF EUROPE

2.1 INTRODUCTION

2.1.1 Atlantic Genetic Affinities

Previous studies have provided evidence for a shared ancestral heritage amongst the Atlantic peoples of Europe. Y chromosome haplotype analysis has highlighted some similarity between the Pyrenean populations of Northern Spain (Basques) and Western ('Celtic') population samples from the British Isles (Hill et al. 2000; Wilson et al. 2001). Specifically, a modal haplotype described by SNP and microsatellite markers (Atlantic Modal Haplotype or AMH) is unusually common in each and has been interpreted as a Palaeolithic legacy which has been relatively undisturbed at the edge of Europe. Some classical gene systems also hint at Atlantic affinities: the ABO blood group O allele and the Rhesus positive condition both show European frequency peaks in the Basque region and in areas with continuing or recent Celtic language use (Cavalli-Sforza et al. 1994).

Apart from human populations there is also a recognised biogeographical pattern (Lusitanian biogeography) seen in several species of Irish flora and fauna that suggest affinities with Southwest Europe. This has been augmented by recent matching molecular phylogeography results from several species including oak, crayfish, hedgehog, pine martens and pygmy shrews (Dumolin-Lapegue 1997; Davison et al. 2001; Seddon et al. 2001; Gouin et al. 2003; Mascheretti et al. 2003). It is also interesting to note that one of the earliest foundation myths for the Irish speaks of an origin in Northern Spain (Duffy 2000, pg 7) These accounts are first recorded with the arrival of Christianity and writing (after 500AD) but are perhaps based on an earlier prehistoric knowledge and tradition. Writing in the 1st century AD, the Roman historian Tacitus also speculated on links between the people of Western Britain and *Hispania* (Spain) (James 1999, pg 52).

2.1.2 European mtDNA Diversity

The study of Mitochondrial DNA (mtDNA) diversity provides an important complement to Y chromosome population genetics. Both yield highly informative non-recombining haplotypes but the potential for different demography in male and female lineage histories makes extrapolation from either alone hazardous (Seielstad et al. 1998).

However, the promise of mtDNA studies in Europe was initially frustrated by the apparent lack of structure within the continent leading to considerable debate about the inferences on demographic history that could justifiably be drawn from European mtDNA diversity (Richards et al. 1996; Cavalli-Sforza and Minch 1997; Richards et al. 1997; Barbujani et al. 1998; Richards and Sykes 1998; Simoni et al. 2000a; Simoni et al. 2000b; Torroni et al. 2000). The apparent lack of differentiation and structure was in stark contrast to the Y chromosome where even neighbouring populations often show significantly different Y chromosome complements, which in turn resolve into numerous continental wide trends and patterns (Rosser et al. 2000). Variation in autosomal loci often exhibits similarly high levels of structure (Cavalli-Sforza et al. 1994).

More recently, principal component analysis (PCA) of mtDNA haplogroup frequencies has demonstrated the existence of the major southeast to northwest trend across Europe typically found in other markers (Richards et al. 2002). However, by considering each mtDNA haplogroup as a segregating allele it did not take advantage of the full range of European variation present at the individual haplotype level. The analysis was also conducted on a coarse geographical level (since samples sizes were still relatively small) and therefore could provide little direct information on finer-scale local variation within the continent.

Smaller scale local investigations of maternally inherited mtDNA and X chromosome variation (each X chromosome spends two-thirds of its history in females) have suggested that there is a substantial asymmetry in the male and female pasts of Ireland and other Celtic language speaking areas of the Britain Isles. Principal component analysis of these markers indicated a greater similarity to central European populations than for the Y chromosome, prompting the explanation that at least one of the "cultural transitions" in the British Isles history involved female immigration (Wilson et al. 2001)

2.1.3 Celtic Languages

Family resemblances amongst the Celtic languages were first identified by Edward Lhuyd at the end of the seventeenth century (Renfrew 1987; Cunliffe 1997; James 1999). The surviving insular Celtic (as opposed to the extinct continental Celtic) languages are usually sub-divided into two groups on the basis of a single consonantal shift. Welsh, Breton and the extinct Cornish are referred to as Brythonic, or P-Celtic, whereas Irish, Scottish Gaelic and the extinct Manx are thought to be more archaic and are referred to as Goidelic, or Q-Celtic (Trask 1996). The continental Gaulish language, formerly spoken in Northern and Central France, was more closely related to the insular Brythonic group, but has no surviving descendants. Traces also survive of Celtiberian, a Celtic language spoken in antiquity in Western and Central Iberia, alongside the non-Indo-European Basque language of North-eastern Spain and Aquitaine; and Lepontic, spoken in Northern Italy from the sixth century BC. Lhuyd named the languages 'Celtic' after one of the names applied to the people of Northern France in Greek and Roman ethnographic writings.

2.1.4 Celtic Ethnogenesis

The Celts or *Keltoi* are first described by the ancient Greeks around the 5th century BC as a people living in central Europe, north of the Alps. From this heartland they are traditionally supposed to have ranged over and conquered large parts of Europe in subsequent centuries. Although 'Celt' is a genuine historical term it is less clear what it meant in an ethnic or indeed any other sense. Notwithstanding this, Lhuyd conflated the historical ethnographic 'Celts' and modern 'Celtic speakers' and followed up his discovery by proposing that the two forms of Celtic language had been brought to Britain and Ireland by waves of invasion from the continent. Prior to this (*ca.* 1700) there was no specific Celtic ethnic identity in the British Isles nor did classical writers ever describe the inhabitants of Britain and Ireland as Celts. However, in the years that followed, archaeologists interpreted similarities in their insular material finds to those of the elite La Tène cultural package of the central European Iron Age as the physical manifestation of Lhyud's waves of Celtic invaders (James 1999).

2.1.5 Atlantic Celts

Despite the adoption of La Tène art styles, archaeological evidence for large-scale Iron-Age migrations into the British Isles has been singularly lacking. In Ireland, for example, La Tène artefacts are relatively rare and are almost always of indigenous manufacture rather than external origin (Raftery 1994), raising questions about the previously accepted (but still only 300 year old) idea of mass Celtic migration to Ireland (Ó Donnabháin 2000). More generally, Renfrew (1987), amongst others, proposed that the roots of insular Celtic identity lay within the region in which the Celtic languages were historically spoken, with the diffusion of Indo-European speakers into Britain and Ireland at the start of the Neolithic ~4000 BC. Cunliffe (2001) appears to go further, describing the coalescence of the Celtic languages along the coastline of the Atlantic façade of Europe, from southern Iberia to the Shetland Islands, *via* maritime networks

reaching back into the late Mesolithic period. The similarities in prehistoric monumental architecture and the spread of the Bronze Age 'Beaker package', to take two examples, attest to the likely sharing of beliefs and attitudes through social networks extending from one end of the Atlantic zone to the other.

This view implies an uncoupling of the link, established by Lhuyd, of a necessary connection between the various aspects of what in the past 200 years has come to be thought of as a 'Celtic package'—including, in particular, the peoples encountered and described as 'Celts' by the classical authors, the producers of Iron-Age La Tène art and their descendants, and speakers of Celtic languages. Modern Celtic speakers should, on this view, be thought of rather as 'Atlantic Celts', whose putative continental Iron-Age ancestry is open to question (James 1999; Cunliffe 2001). At the same time, many archaeologists and a seeming majority of historians retain the traditional view with some vigour (Megaw and Megaw 1996; Megaw and Megaw 1998).

2.1.6 Present Study

In this Chapter, mtDNA diversity in 200 Irish individuals is examined to investigate further the maternal origins of the Irish and the extent of any asymmetry with the paternal history recorded by the Y chromosome. These are examined in a wider genetic context, with the assembly and testing for errors of published European and Middle Eastern mtDNA Hyper-variable segment 1 (HVS-1) sequences from 8532 individual samples comprising 44 additional populations. Analyses of genetic affinities within these data are used to address Irish population origins as well as broader questions of mtDNA structure within Europe and the extent to which loci of different inheritance patterns record similar histories.

2.2 MATERIALS AND METHODS

2.2.1 Subjects and Samples

The majority of samples used in this study were assembled by Dr. Emmeline Hill (1999). An additional 122 samples were collected from Natural Science students in Trinity College, Dublin to supplement this existing sample archive. Buccal cheek cells were obtained from the participants using *CYTOSOFT*^{TM} cytology brushes (Medical Packaging Corporation). Each volunteer was also asked to complete a questionaire giving their surname and county of origin as well as a place of origin for their parents and grandparents. No other information was retained with the samples ensuring effective anonymity. Sampling was carried out in accordance with the principle of informed consent with both verbal and written explanation of the research aims and consequences of participation.

A total of 200 individuals were genotyped from the sample archive. mtDNA data for 100 previously published Irish individuals (Richards et al. 1996) were also included to give a total composite sample size of 300. There were divided into East (n=128) and West Ireland (n=127) regional sub-groupings based on maternal ancestral, usually grandmother's, origin. The eastern grouping includes the province of Leinster and the three East Ulster countries of Armagh, Antrim and Down. The west was defined as the province of Connacht and western counties of Munster (Kerry, Limerick and Clare) as well as Donegal in West Ulster. The remaining 45 samples either had no specific origin information within Ireland or a place of origin outisde the above defined eastern and western regions.

2.2.2 DNA Extraction

DNA was isolated from buccal cheek cells using a standard phenol/choroform extraction method with ethanol precipitation. Brushes were incubated in 500 µl of cell lysis buffer (0.1M NaCl, 55mM Tris (pH 8.0), 1mM EDTA (pH 8.0), 0.5% SDS, 1.5 µg/µl of Proteinase-K) overnight at 56°C. After removing the brush, the cell lysate was mixed with an equal volume (500µl) of phenol, thoroughly vortexed (5 minutes) and centrifuged (13,000 RPM for 5 minutes). The resulting aqueous layer was added to an equal volume of a chloroform/isoamyl alcohol mixture (50:1 V/V) and vortexed and centrifuged as above. The aqueous layer was again retrieved and mixed with 50µl of Sodium Acetate (3M) and 1ml of 100% ethanol. The DNA was allowed to precipitate for at least 24 hours at -20°C and then pelleted by centrifugation (13,000 RPM for 10 minutes). The supernatant was discarded and the DNA pellet air-dried for 1 hour before re-suspension in 200µl of 1xTE buffer (0.8M Tris-HCl, 0.01M EDTA). Samples were stored at -20°C.

2.2.3 mtDNA HVS-1 Sequencing

The highest mtDNA mutation rates occur in the non-coding control region, particularly in two small regions termed Hyper-Variable Segment 1 and 2 *(*HVS-1 and HVS-2), which are consequently commonly assayed. A 571 base pair (bp) portion of the mtDNA control region from positions 15927 to 16497, all position numbers after Anderson et al. (1981), was initially amplified using the forward primer (L15926) 5'- TCA AAG CTT ACA CCA GTC TTG TAA ACC – 3' and reverse primer (H16498) 5'-CCT GAA GTA GGA ACC AGA TG –3'. Primer nomenclature refers to the light (L) and heavy (H) mtDNA strands and the position number of the last primer base before the amplicon proper.

Polymerase Chain Reaction (PCR) amplifications were carried out in 25µl volumes containing 0.336µM of each primer, 50mM KCl, 10mM Tris-HCl, 1% Triton-X-100 (V/V), 2.5mM $MgCl_2$, 0.32mM of each nucleotide dATP, dCTP, dGTP, dTTP, approximately 0.05 Units/µl of *Taq* Polymerase and about 15ng of template DNA. Amplifications involved an initial denaturation step of 3 minutes at 94°C; followed by 30 cycles of 40 seconds at 55°C, 1 minute at 72°C, 40 seconds at 94°C and completed with one final extension cycle of 1 minute at 55°C and 2 minutes at 72°C.

A 440 bp portion of the PCR amplicon, including HVS-1, was then sequenced after cleaning using a *CONCERT*^{TM} Rapid PCR purification system (Life Technologies). Sequencing was undertaken from both the 5' and 3' ends of each amplicon using internal sequencing primers L15977, 5'- CAC CAT TAG CAC CCA AAG CT –3', and H16438, 5'-CGG AGC GAG GAG AGT AGC –3' respectively, on an ABI Prism^{TM} 377 DNA sequencer (MWG-Biotech, Milton Keynes, United Kingdom). It was necessary to sequence from both directions as a T to C transition at position 16089 creates a 10 base pair polycytosine tract midway through the PCR fragment which leads to band blurring beyond the tract distal to the sequencing primer (Richards et al. 1996).

The forward and reverse sequences for each sample were aligned to construct a single continuous sequence, which generally ran from positions 16000 to 16500, and mutations relative to the Cambridge Reference Sequence (CRS) (Anderson et al. 1981) recorded. However, to allow maximum comparability between the new and previously published Irish data, only sequence from positions 16030 to 16394 was considered**.**

2.2.4 mtDNA SNP Genotyping

The high mutation rate of the mtDNA control region can sometimes lead to difficulties in inferring broader haplogroups from individual haplotypes. This can often be resolved using additional informative single nucleotide polymorphisms (SNPs) from other areas of the mtDNA genome. Three of these sites, 12308, 4577, 7028, found in the coding region, define haplogroups U, V, and H respectively. The remaining site at position 73 in the non-coding HVS-2 can be informative in distinguishing haplogroup pre-HV (Torroni et al. 1996; Macaulay et al. 1999). In order to ensure maximum comparability of haplotypes generated as part of this study and those from Richards et al. (1996), samples from the latter were also retrospectively assayed for position 73 where possible.

The state of each SNP was assayed through restriction fragement length polymorphsim (RFLP) analysis. PCR reaction conditions are as described above for HVS-1 amplication, with details of primers and RFLP typing procedures summarised in **Table 2.1**. In all cases, PCR products were digested using the appropriate restriction enzyme (RE) for 3 hours at 37°C. The restriction pattern was resolved using a 2% agarose gel containing Ethidium Bromide and visualised under Ultra-Violet (UV) light.

	PCR AMPLIFICATION		RFLP GENOTYPING			
SNP Position and Character	Primer Sequences (5' to 3') [1]	PCR Product Size	Restriction Enzyme (RE)	Digestion Conditions [2]	SNP State (RFLP Pattern)	Notes
7028 C/T	Forward (L6909): AAG CAA TAT GAA ATG ATC TG Reverse (H7115): CGT AGG TTT GGT CTA GG	232	*Alu* I	0.25 U/µl RE 50mM NaCl 10mM Tris-HCl 10 mM $MgCl_2$ 1 mM DTT	C (167bp + 65 bp) T (137bp + 65bp +30bp)	C allele defines Hg H
12308 A/G*	Forward (L12124) : CTC AAC CCC GAC ATC ATT ACC Reverse (H12309): ATT ACT TTT ATT TGG AGT TGC ACC AAG	235	*Hinf* I	0.5 U/µl RE 50mM NaCl 10mM Tris-HCl 10 mM $MgCl_2$ 1 mM DTT	A (168bp + 67bp) G (138bp + 67bp + 30bp)	G allele define Hg U
4577 C/T	Forward (L4325): GGA GCT TAA ACC CCC TTA Reverse (H4720): GGA TAA GAT TGA GAG AGT G	645	*Nla* III	0.5 U/µl RE 50mM K Acetate 20mM Tris-Acetate 10mM Mg-Acetate 1mM DTT 0.1 µg/µl BSA	T (556bp + 89bp) C (287bp + 269bp +89bp)	T allele defines Hg V
73 G/A	Forward (L29): GGT CTA TCA CCC TAT TAA CCA C Reverse (H223) : TCA ATT GTT ATT ATT ATG TCC TAC AA	234	*Alw* 44 I	0.5 U/µl RE 66mM K-Acetate 33mM Tris-Acetate 10mM Mg-Acetate	A (234bp) G (169bp + 65bp)	Aids identification of Haplogroup pre-HV

*Assayed using a mutagenic primer as described in Torroni et al. 1996

[1] L and H in primer nomenclature refer to the Light (forward) and Heavy (Reverse) mtDNA strands respectively and are followed by the position number of the the last primer base before the amplicon proper

[2] Using Restriction enzyme manufacturers conditions and buffers

Table 2.1 *mtDNA SNP genotyping showing details of PCR primer sequences and RFLP SNP assay procedures.*

2.2.5 Analysis of mtDNA Phylogeography

Initially, the most basic level of phylogeography, that of the geographical provenance of matches to Irish mtDNA haplotypes, was investigated. This was implemented by comparing each haplotype found in Ireland (considered from positions 16093-16362) to a world database of mtDNA HVS-1 sequences assembled from previous studies (Röhl et al. 2001). Using information on the location and frequency of the closest matching haplotypes, the geographical information system 'mtradius' (Forster et al. 2002) calculates a centre of gravity or centre of distribution. Each centre of gravity is accompanied by a standard deviation (SD) in kilometres (km) as an indication of the dispersal range of haplotypes. Higher standard deviations tend to occur with common ancestral haplotypes, such as the CRS, which have widespread distributions and are thus phylogeographically uninformative. Locally spread haplotypes provide a crisper signature of maternal gene flow and were identified here as point estimates with a SD of less than 500 km. An intermediate category of 500-1000 km was also defined, while uninformative (diffuse) distributions were associated with a SD of greater than 1000 km. The centres of gravity for every haplotype in each category were plotted on a map of Europe with the area of each circle proportional to the frequency of that sequence in Ireland. Mtradius runs were carried out by Peter Forster at the University of Cambridge.

2.2.6 mtDNA Population Affinities

A second approach examined the affinity of the Irish and neighbouring European and Middle Eastern data grouped as populations. Control region sequences from 8533 individuals were assembled from previous studies and divided into 45 geographically defined population samples. These divisions were primarily based on current national boundaries. However, regional and/or ethnic group units were also used where first, sufficient numbers existed, and second, where their inclusion would not lead to a loss of data that might otherwise be included under a national grouping. The details of these

divisions, sample sizes and references are given in **Table 2.2**. Samples from some small and/or isolated populations were excluded from the analysis because of the possibility of unusually strong genetic drift which would confound the perception of broader phylogeographic patterns. These included the Western Isles of Scotland, Orkney and Skye (Helgason et al. 2001).

2.2.7 HVS-1 Sequence Quality Checks

The quality of the mtDNA sequence data used was tested using the methods set out in Bandelt et al. (2002). First, this involved dividing the range of polymorphisms found in each population into fast changing 'speedy' transitions and slower evolving sites consisting of 'weighty' transitions as well as transversions/Indels. Departures in the ratio of weighty transitions to transversions/indels (WTTI ratio) when compared to a dataset of high quality suggest potential anomalies, for example from sequencing errors. The significance of deviations in WTTI ratios was assessed using a one tailed Fisher's exact test compared to two benchmark data sets: Ireland, where quality can be controlled directly, and Iceland, a large sample size, whose WTTI ratio is close to the median value of the populations examined.

The second approach aims to identify systematic errors at particular sites. Phantom mutations of this type lead to incompabilities in the networks of the sequence data that are marked by the presence of excessive numbers of higher dimensional cubes. This property of the network can be summarised and expressed numerically through the calculation of a 'cube spectrum' for each population sequence set. Transversions and indels in the poly C tract (16182-16194) of HVS-1, as well as all insertions were ignored in the above sequence quality checks and in subsequent population analysis as they are potential sequencing artefacts (Bendall and Sykes 1995) and are not reported consistently across studies.

2.2.8 Inter-population Genetic Distances

Genetic distances as linearised Φ_{ST} statistics (Slatkin 1995) were calculated between all populations using ARLEQUIN Version 2.000 (Schneider et al. 2000). The Φ_{ST} values were based on pairwise sequence difference between positions 16090 and 16365 only so as to allow for maximum comparability between all populations. To account for mutation rate heterogeneity in the mtDNA control region, site rates were modelled as i.i.d (independently and identically distributed) gamma with $\alpha = 0.26$ (Meyer et al. 1999). The significance of Φ_{ST} values was gauged by permuting samples across populations over 10,000 replicates to generate a distribution of values under the null hypothesis.

2.2.9 Multidimensional Scaling

The resulting matrix of inter-population Φ_{ST} values was summarised in two significant dimensions using Multidimensional Scaling (MDS) analysis implemented by the ALSCAL program included in the SPSS package (Version 11.0, SPSS Inc.). Like principal component analysis, MDS (see Young 1987) is a general data reduction technique. In population genetics it can be used to reduce the information contained in the n-1 dimensions of a Φ_{ST} matrix, where n is the number of populations, to a more readily interpretable format. Each population is arranged in a reduced conceptual space (here two-dimensional) such that the distance between populations preserves as strongly as possible the genetic similarity between them.

Population	n	Latitude[1]	Longitude[1]	Reference(s) (Number of Samples)
Albania	42	41.3	20.3	Belledi et al. 2000
Armenia	191	40.2	44.5	Richards et al. 2000
Austria	99	47.5	14.9	Parson et al. 1998
Azerbaijan	48	40.4	49.8	Richards et al. 2000
Basque Country	156	43.4	-3.2	Bertranpetit et al. 1995 (45), Corte-Real et al. 1996 (61), Richards et al. 2000 (50)
Belgium	33	50.8	4.3	Decorte, 1996[2]
Brittany	62	47.9	-2.1	Dubut et al. 2004
Bulgaria	141	42.6	25.4	Calafell et al. 1996 (30), Richards et al. 2000 (111)
Cornwall	92	50.6	-4.3	Richards et al. 1996 (69), Richards et al. 2000 (23)
Czech Republic	83	49.8	15.1	Richards et al. 2000
Denmark	38	55.9	9.9	Richards et al. 1996 (33), Richards et al. 2000 (5)
England	242	52.3	-0.8	Helgason et al. 2001 (142), Piercy et al. 1993 (100)
Estonia	149	58.6	26.5	Sajantila et al. 1995 (28), Sajantila et al. 1996 (20), Richards et al. 2000 (111)
Finland	153	62	25.7	Sajantila et al. 1995 (50), Richards et al. 1996 (29), Kittles et al. 1999 (74)
France	379	46.9	2.3	Richards et al. 1996 (47), Rousselet and Mangin 1998 (50), Cali et al. 2001 (110), Dubut et al. 2004 (148), CEPH Database (24)
Galicia	135	43	-7.6	Salas et al. 1998 (92), Gonzalez et al. 2003 (43)
Germany	582	51	10.3	Richards et al. 1996 (156), Hofmann et al. 1997 (67), Baasner et al. 1998 (49), Lutz et al. 1998 (200), Pfeiffer et al. 1999 (109)
Greece	179	39.6	22	Richards et al. 2000 (125), Kouvatsi et al. 2001 (54)
Hungary	78	47.2	19.8	Lahermo et al. 2000
Iceland	447	64.5	-18	Sajantila et al. 1995 (39), Richards et al. 1996 (14), Helgason et al. 2000 (394)
Iraq	116	33.3	44.1	Richards et al. 2000
Ireland	300	52.9	-7.8	Richards et al. 1996 (100), present study (200)
Italy	248	43.1	12.4	Francalacci et al. 1996 (49), Richards et al. 2000 (48), Mogentale-Profizi et al. 2001 (68), Tagliabracci et al. 2001 (83)
Jordan	146	31.5	36.5	V.M. Cabrera personal communication
Karelia	83	61.8	34.3	Sajantila et al. 1995
Kurdistan	53	37.6	43.8	Richards et al. 2000
Northern Ossetia	106	43.1	44.6	Richards et al. 2000
Norway	629	60.7	8.9	Opdal et al. 1998 (216), Richards et al. 2000 (16), Helgason et al. 2001 (324), Passarino et al. 2002 (74)
Palestine	117	31.8	35.2	Di Rienzo and Wilson 1991 (9), Richards et al. 2000 (108)
Poland	473	52.1	19.9	Richards et al. 2000 (37), Malyarchuk et al. 2002 (436)
Portugal Centre	162	39.5	-8	Pereira et al. 2000 (84), Gonzalez et al. 2003 (78)
Portugal North	183	41.1	-7.8	Pereira et al. 2000 (99), Gonzalez et al. 2003 (84)
Portugal South	195	37.7	-7.9	Pereira et al. 2000 (58), Gonzalez et al. 2003 (137)
Romania	92	45.7	25.6	Richards et al. 2000
Russia	379	54.2	37.6	Orekhov et al. 1999 (103), Richards et al. 2000 (25), Malyarchuk and Derenko 2001 (50),Malyarchuk et al. 2002 (162)
Sardinia	115	39.9	9.2	Di Rienzo and Wilson 1991 (69), Richards et al. 2000 (45)
Scotland	895	56.4	-3.8	Helgason et al. 2001
Sicily	196	37.6	14.2	Richards et al. 2000 (90), Cali et al. 2001 (106)
Spain North	106	41.2	-3.5	Corte-Real et al. 1996 (45), Larruga et al. 2001 (61)
Spain Centre/South	103	38.1	-3.8	Corte-Real et al. 1996 (88), Larruga et al. 2001 (15)
Sweden	32	59.2	15.2	Sajantila et al. 1996
Switzerland	224	46.9	8.4	Pult et al. 1994 (70), Dimo-Simonin et al. 2000 (154)
Syria	69	34.5	38.2	Richards et al. 2000
Turkey	290	38.7	35.5	Richards et al. 2000 (218), Di Benedetto et al. 2001 (72)
Wales	92	52.5	-3.4	Richards et al. 1996
TOTAL	**8733**			

[1] Geographic coordinates are in decimal degree format. Positive values indicate Northern latitudes and Eastern longitudes.

[2] HVS-I sequence data for Belgium were retrieved from the Ferrara Human Genetics Database: (http://web.unife.it/progetti/genetica/pdata.htm)

Table 2.2 *European and Middle Eastern mtDNA HVS-1 population samples included in this study with details of sample sizes and references as well as the geographic coordinates used for each population in spatial analyses.*

Population	n	Latitude[1]	Longitude[1]	Reference(s) (Number of Samples)
Armenia	89	40.2	44.5	Rosser et al. (2000)
Basque Country	76	43.4	-3.2	Rosser et al. (2000) (26) and Wilson et al. (2001) (50)
Belgium	92	50.8	4.3	Rosser et al. (2000)
Belorus	41	53.5	27.6	Rosser et al. (2000)
Bulgaria	24	42.6	25.4	Rosser et al. (2000)
Cornwall	51	50.6	-4.3	Rosser et al. (2000)
Cyrpus	45	35	33.5	Rosser et al. (2000)
Czech Republic	53	49.8	15.1	Rosser et al. (2000)
Denmark	56	55.9	9.9	Rosser et al. (2000)
England (East Anglia)	172	52.3	-0.8	Rosser et al. (2000)
Estonia	207	58.6	26.5	Rosser et al. (2000)
Finland	57	62	25.7	Rosser et al. (2000)
France	40	46.9	2.3	Rosser et al. (2000)
Georgia	64	42	44.1	Rosser et al. (2000)
Germany	110	51	10.3	Rosser et al. (2000)
Greece	36	39.6	22	Rosser et al. (2000)
Hungary	36	47.2	19.8	Rosser et al. (2000)
Iceland	28	64.5	-18	Rosser et al. (2000)
Ireland	221	52.9	-7.8	Hill et al. (2000)
Italy	99	43.1	12.4	Rosser et al. (2000)
Latvia	34	56.9	25.6	Rosser et al. (2000)
Lithuania	38	55.1	24.2	Rosser et al. (2000)
Netherlands	84	52.2	5.9	Rosser et al. (2000)
Norway	130	60.7	8.9	Rosser et al. (2000) (52) Wilson et al. (2001) (78)
Orkney	71	59	-2.9	Wilson et al. (2001)
Ossetia	47	43.1	44.6	Rosser et al. (2000)
Poland	112	52.1	19.9	Rosser et al. (2000)
Portugal North	328	37.7	-7.9	Rosser et al. (2000)
Portugal South	57	41.1	-7.8	Rosser et al. (2000)
Romania	45	45.7	25.6	Rosser et al. (2000)
Russia	122	54.2	37.6	Rosser et al. (2000)
Scotland	163	56.4	-3.8	Rosser et al. (2000)
Slovakia	70	48.8	19.7	Rosser et al. (2000)
Slovenia	70	46.1	14.8	Rosser et al. (2000)
Spain	126	39.9	-4	Rosser et al. (2000)
Sweden	112	59.2	15.2	Rosser et al. (2000)
Syria	72	34.5	38.2	Wilson et al. (2001)
Turkey	235	38.7	35.5	Rosser et al. (2000) (167) Wilson et al. (2001) (68)
Ukraine	27	49.5	32.1	Rosser et al. (2000)
Wales	88	52.5	-3.4	Wilson et al. (2001)
West Friesland	94	53.3	6	Wilson et al. (2001)
Yugoslavia	100	43.9	19.8	Rosser et al. (2000)
TOTAL	**3822**			

[1] Geographic coordinates are in decimal degree format. Positive values indicate Northern latitudes and Eastern longitudes.

Table 2.3 *European and Middle Eastern Y chromosome population samples included in this study with details of sample sizes and references as well as the geographic coordinates used for each population in spatial analyses.*

2.2.10 Y Chromosome and Autosomal Diversity

In order to place European mtDNA variation in a broader context, published data from loci with different inheritance modes was re-examined. Y chromosome data comprised haplotypes defined by the 9 binary polymorphisms (SRY1532, SRY8299, 92R7, TAT, M9, 12f2, YAP, sY81, LLY22g) typed in common between Hill et al. (2000), Rosser et al. (2000) and Wilson et al. (2001), from where the data was drawn. In ensuring strict

comparability, it was sometimes necessary to collapse some haplogroups defined by a marker not typed in other data sets back into preceding groups in the Y chromosome hierarchy; for example, Haplogroup 22 into Haplogroup 1. A total of 3822 individuals across 10 haplogroups were divided into 42 populations (**Table 2.3**). Linearised Φ_{ST} distances, based on pairwise difference between binary marker haplotypes, and Multidimensional Scaling were used as for mtDNA analysis but without any correction for 'multiple hits' as the mutation rate is far lower. Autosomal variation was analysed in a similar manner using a matrix of inter-population F_{ST} values based on classical gene frequencies of 88 alleles in 25 European populations taken from Cavalli-Sforza et al. (1994, pg 270).

2.2.11 Synthetic Surface Maps

The geographic pattern of each dimension of mtDNA, Y chromosome and autosomal variation was visualised by interpolating observed values to produce a synthetic surface map of Europe using the Spatial Analyst Extension of ArcView (Version 3.2, Environmental Systems Research Institute Inc.). A geographic co-ordinate was assigned to each population (**Table 2.2 and Table 2.3**) and interpolation carried out by the Inverse Distance Weighted (IDW) method utilising the 12 nearest neighbours. The interpolated surface was divided into 12 equal (software determined) classes or contours.

2.2.12 Spatial Autocorrelation

Congruence between dimensions in different genetic systems was assessed using Pearson's correlation coefficient. However, gene flow between neighbouring populations leads to spatial autocorrelation or the prior non–independence of genetic variation across loci with respect to geography. To correct for this in assessing the true extent of correlation between variables (markers with different inheritance modes), the

Dutilleul (1993) method, as implemented in the PASSAGE package (Rosenberg 2001), was applied. An effective or reduced sample size (and thus fewer degrees of freedom) is calculated to reflect spatial structure; this can then be applied in a *t*-test to gauge the significance of the correlation coefficient. The correlation is carried out over a number of geographic distance classes. An objective rule of thumb, Yule's rule, was applied to determine the most appropriate number of classes:

$$\text{Number of Distance classes} = 2.5\sqrt[4]{n} \qquad \textbf{(Equation 2.1)}$$

where n is the number of object pairs; for a matrix this is (P x (P-1))/2 with P the number of populations. Physical distances (in kilometres) between populations were calculated from central geographic coordinates using the great circle method.

Prior spatial autocorrelation is also a consideration when determining whether geographic variation displays gradient or clinal trends. The Moran's I coefficient is commonly used to measure the similarity between samples for a given variable as a function of geographic distance. Values range between –1 and +1 indicating strong negative or positive correlation respectively; that is within any particular geographic distance range or class, values are less or more similar to each other than is expected by chance alone. The process is repeated over several distance classes to produce a 'correlogram' (Moran's I values versus geographic distances categories) allowing broader spatial patterns to be assessed. The overall significance of the correlogram can be gauged to any threshold α (for example $\alpha = 0.05$) if at least one of the Moran's I coefficients is significant at α/k, where k is the number of distances classes (Bonferroni Criterion). Clinal spatial autocorrelation analysis was also carried out using the PASSAGE package (Rosenberg 2001).

2.3 RESULTS

2.3.1 Irish mtDNA Diversity

A total of 155 haplotypes were observed amongst the 300 Irish individuals studied (including 100 from previous studies), with all but one sample falling into the main west Eurasian haplogroups U, HV, JT, I, W and X (Richards et al. 1998). Haplogroup frequencies are given in **Table 2.4** and are within the range found in most European regions (Richards et al. 2000). In common with other European populations, H is by far the most frequent haplogroup in Ireland. The Cambridge Reference Sequence (CRS) type of Hg H is the most common individual lineage observed at 18.3 %.

The significance of regional differences in mtDNA haplogroup frequencies between East (n = 127) and West (n = 128) Ireland was examined using a χ^2 test. Smaller sub-groupings were collapsed into over-arching haplogroups to allow sufficient numbers for testing. Using these haplogroups (H, I, J, K ,T,U ,V ,W) there is no significant difference between the two regions (p=0.115). In addition, the genetic distance (Φ_{ST} value) between East and West Ireland based on HVS-1 is small and not significantly greater than zero. This contrasts with the Y chromosome pattern within Ireland, where East and West complements have been shown to be substantially different, a discrepancy attributed to the preferential settlement of subsequent migrants to the accessible east coast after initial colonisation (Hill et al. 2000).

	Ireland (All)		West Ireland		East Ireland		Other Ireland*	
HAPLOGROUP	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
D	1	0.3	0	0	1	0.8	0	0
H	133	44.3	59	46.1	54	42.5	20	44.4
HV*	4	1.3	3	2.3	1	0.8	0	0
I	9	3	7	5.5	2	1.6	0	0
J*	26	8.7	14	10.9	9	7.1	3	6.7
J1a	1	0.3	0	0	1	0.8	0	0
J1b1	2	0.7	1	0.8	1	0.8	0	0
J2	3	1	0	0	2	1.6	1	2.2
K	36	12	13	10.2	18	14.2	5	11.1
T*	3	1	2	1.6	1	0.8	0	0
T1	4	1.3	2	1.6	1	0.8	1	2.2
T2	14	4.7	3	2.3	5	3.9	6	13.3
T4	1	0.3	1	0.8	0	0	0	0
U*	1	0.3	0	0	1	0.8	0	0
U2	4	1.3	0	0	4	3.1	0	0
U3	3	1	0	0	3	2.4	0	0
U4	4	1.3	3	2.3	0	0	1	2.2
U5*	5	1.7	2	1.6	3	2.4	0	0
U5a	1	0.3	0	0	1	0.8	0	0
U5a1	4	1.3	1	0.8	1	0.8	2	4.4
U5a1a	6	2	2	1.6	4	3.1	0	0
U5b	9	3	3	2.3	5	3.9	1	2.2
V	17	5.7	5	3.9	7	5.5	5	11.1
W	7	2.3	6	4.7	1	0.8	0	0
X	2	0.7	1	0.8	1	0.8	0	0
TOTAL	**300**	**100**	**128**	**100**	**127**	**100**	**45**	**100**

* Includes Irish samples from outside the East/West region and those which have no specific information on place of origin within Ireland

Table2.4 *Frequencies of mtDNA haplogroups in Ireland, as a whole and in East and West Ireland considered as separate regions.*

2.3.2 Neolithic and Paleolithic Haplogroups in Ireland

mtDNA haplogroups display little obvious geographical specificity within Europe. An alternative approach examines the distribution of lineages by their age. Founder analysis has dated the entry of different mtDNA lineages into Europe by examining the levels of nucleotide diversity accumulated around haplotypes that have matches in the Near East (Richards et al. 2000). Approximately 20% of Europeans, principally those belonging to haplogroups J, T1 and U3, are proposed to descend from Neolithic settlers, with the remainder attributed to earlier Palaeolithic inhabitants. 13% of the Irish sample belong

to these groupings, a value toward the lower end of the range found in Europe and similar to areas such as Scandinavia and the Western Mediterranean (Iberia). However, it is nearly twice as great as the value for the Basque population (Richards et al. 2000). Within Ireland, there is no east-west regional variation in the frequencies of these groups, occurring at 13% in both areas.

2.3.3 Phylogeography of Irish mtDNA Lineages

The estimated centres of geographical distribution for matches to HVS-1 sequence haplotypes found in Ireland, calculated with reference to sampled haplotypes throughout the world, are shown in **Figure 2.1**. Reassuringly, the vast majority of dispersal points are found in Europe and when the standard deviation (SD) of each type is considered it is possible to discern further details. The most frequent Irish haplotypes, for example the CRS, and the root motifs of haplogroups K and J also have high SD (shown in white in **Figure 2.1**) indicating a widespread distribution throughout Europe. However, the grey (intermediate SD category 500-1000KM) and particularly black (low SD; <500KM) centres of gravity, though less frequent in Ireland, are more informative on Irish matrilineal affinities. Haplotypes with intermediate SD are more common in Western Europe, while haplotypes with low SD are concentrated almost exclusively in Atlantic and (to a much lesser extent) Mediterranean Europe. However, the most striking result is the very strong sharing of localised haplotypes with Britain, particularly Scotland. The mtradius analysis does not indicate directionally and thus the links could be due to movements from Britain to Ireland, vice versa or indeed both. The Scottish matches are widely distributed throughout Ireland, and are not concentrated in geographically proximate areas like Northeast Ulster. A lesser degree of sharing is also apparent between Ireland and Pyrenean Spain. Finally, it is also noteworthy that particular mtDNAs that are characteristic of central Europe, such as J1a (Richards et al. 1998), are virtually absent from the Atlantic façade.

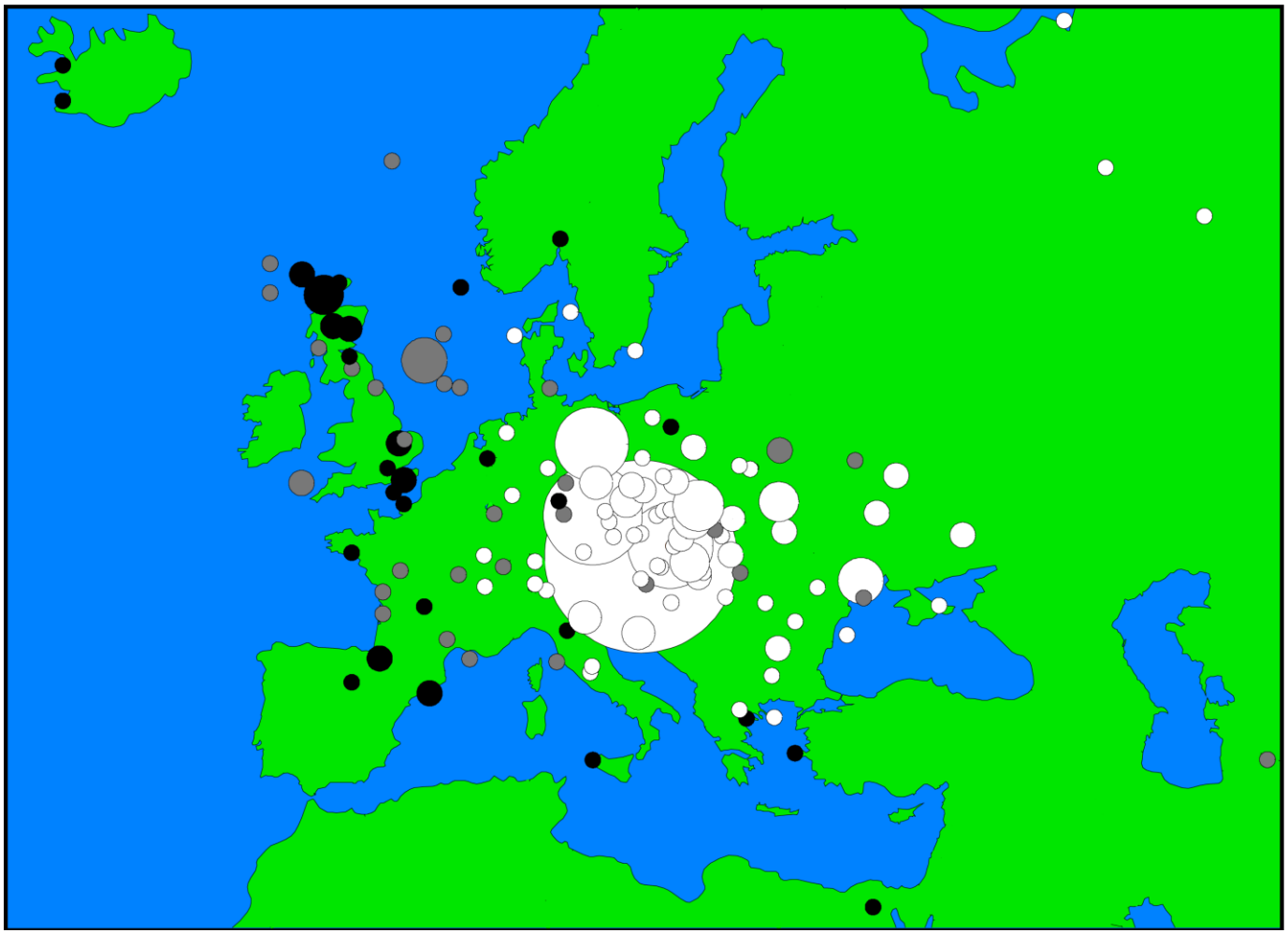


Figure 2.1. *Estimated dispersal points or centres of gravity for the 146 mtDNA HVS-1 haplotypes (positions 16093-16362) found in Ireland. Each circle represents a distinct haplotype. Circle size indicates the frequency of that type in Ireland with the largest representing the CRS (n=56) and the smallest indicating a frequency of 1; intermediate frequencies are proportional to circle area. Standard deviation is indicated as follows: black =<500 km; grey =500–1000 km; white = >1000 km. 11 centres are outside the range of this map in Asia (10) and Africa (1).*

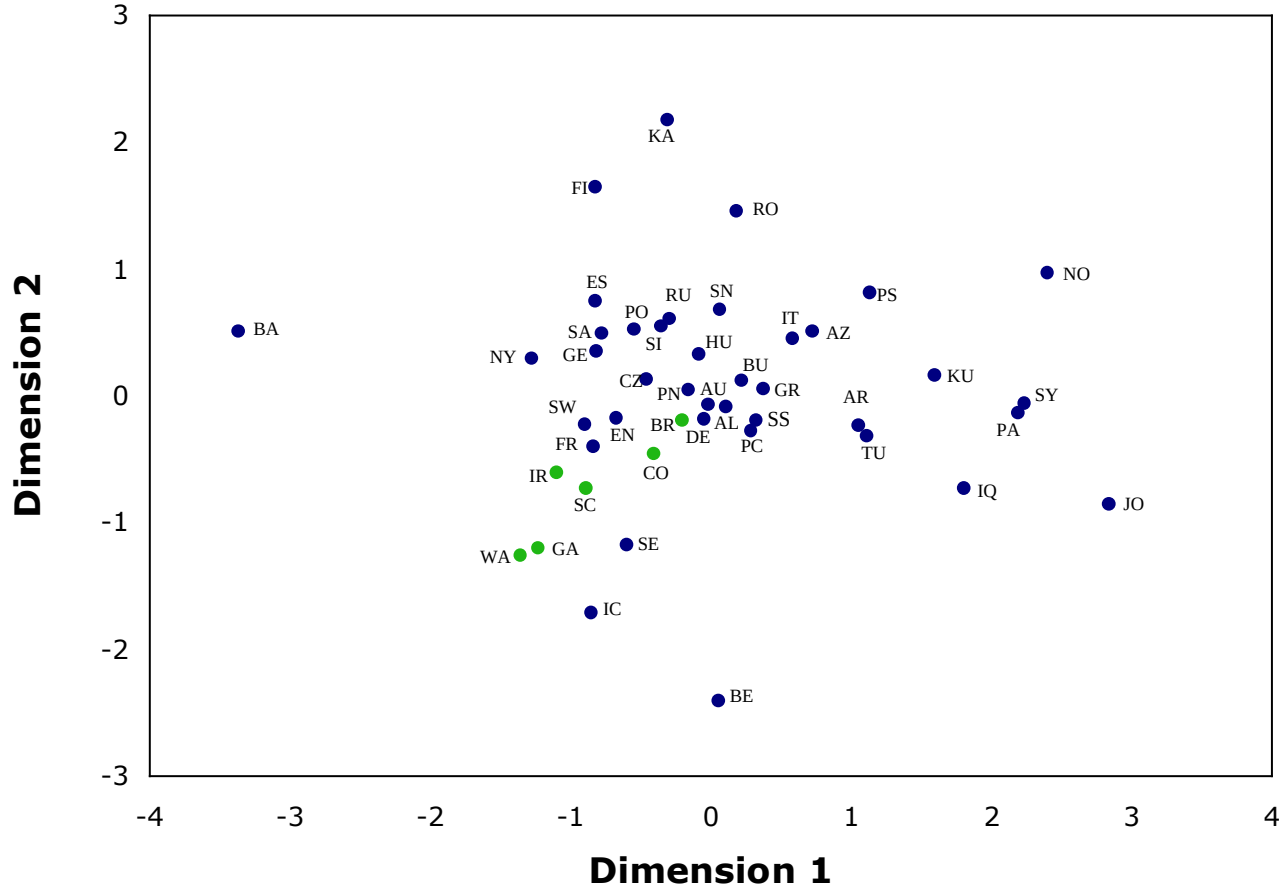


Figure 2.*2* *Multidimensional scaling (MDS) plot of inter-population Φ_{ST} values calculated from mtDNA control region sequence data. The matrix has been condensed to two dimensions, which account for 82% of the original variation. Populations with continuing or historic Celtic language use are shown in green. Population labels are as follows: AL = Albania; AR = Armenia; AU = Austria; AZ = Azerbaijan; BA=Basque Country; BE = Belgium; BR= Brittany; BU = Bulgaria; CZ = Czech Republic; CO = Cornwall; DE = Denmark; EN = England; ES = Estonia; FI = Finland; FR = France; GA = Galicia; GE = Germany; GR = Greece; HU = Hungary; IC = Iceland; IQ = Iraq; IR=Ireland; IT = Italy; JO = Jordan; KA = Karelia; KU = Kurdistan; NO = Northern Ossetia; NY= Norway; PA = Palestine; PC = Portugal Central; PN = Portugal North; PO = Poland; PS = Portugal South; RO = Romania; RU = Russia; SA = Sardinia; SC = Scotland; SE = Sweden; SI = Sicily; SN = Spain North; SS = Spain South-Central; SW = Switzerland; SY = Syria; TU = Turkey; WA = Wales.*

	Polymorphic Sites and WTTI Ratio								Cube Spectrum			
Population	**n**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**0**	**1**	**2**	**3**
Albania	42	44	14	3	11	3.67	0.343	0.459	13	13	1	0
Armenia	191	108	45	15	30	2	0.56	0.437	56	68	14	1
Austria	99	70	26	5	21	4.2	0.192	0.316	24	23	0	0
Azerbaijan	48	55	14	2	12	6	0.174	0.267	14	13	0	0
Basque Country	156	56	20	5	15	3	0.394	0.525	20	19	0	0
Belgium	33	37	12	2	10	5	0.253	0.355	12	11	0	0
Brittany	62	45	10	2	8	4	0.366	0.47	10	9	0	0
Bulgaria	141	75	27	8	19	2.38	0.516	0.649	28	30	3	0
Cornwall	92	56	19	4	15	3.75	0.285	0.41	19	18	0	0
Czech Republic	83	51	11	4	7	1.75	0.732	0.456	8	7	0	0
Denmark	38	31	7	1	6	6	0.318	0.406	7	6	0	0
England	242	97	36	10	26	2.6	0.43	0.638	43	48	6	0
Estonia	149	62	24	7	17	2.43	0.509	0.639	21	20	0	0
Finland	153	67	20	10	10	1	0.155	0.113	18	17	0	0
France	379	119	55	18	37	2.06	0.58	0.45	64	73	11	1
Galicia	135	74	25	7	18	2.57	0.467	0.603	24	23	0	0
Germany	582	131	64	21	43	2.05	0.574	0.442	70	75	6	0
Greece	179	82	36	12	24	2	0.63	0.448	43	51	10	1
Hungary	78	67	25	7	18	2.57	0.467	0.603	21	20	0	0
Iceland	447	77	28	8	20	2.5	0.477	NA	28	27	0	0
Iraq	116	91	33	9	24	2.67	0.418	0.656	32	33	2	0
Ireland	300	92	37	12	25	2.08	NA	0.477	38	39	2	0
Italy	248	102	37	7	30	4.29	0.144	0.267	41	43	3	0
Jordan	146	92	35	6	29	4.83	0.11	0.218	36	39	4	0
Karelia	83	41	14	3	11	3.67	0.343	0.459	15	15	1	0
Kurdistan	53	59	16	5	11	2.2	0.599	0.705	15	14	0	0
Northern Ossetia	106	58	15	5	10	2	0.656	0.504	15	14	0	0
Norway	629	136	71	15	56	3.73	0.146	0.295	80	88	9	0
Palestine	117	102	42	17	25	1.47	0.307	0.223	40	41	2	0
Poland	473	121	53	13	40	3.08	0.278	0.444	55	57	3	0
Portugal Centre	162	80	29	6	23	3.83	0.405	0.351	28	27	0	0
Portugal North	183	84	34	7	27	3.86	0.196	0.332	33	33	1	0
Portugal South	195	82	30	9	21	2.33	0.522	0.568	35	38	4	0
Romania	92	57	16	6	10	1.67	0.478	0.388	17	16	0	0
Russia	379	103	36	9	27	3	0.329	0.484	36	36	1	0
Sardinia	115	65	20	7	13	1.86	0.252	0.435	19	19	1	0
Scotland	895	128	63	23	40	1.74	0.425	0.312	69	77	9	0
Sicily	196	73	27	12	15	1.25	0.236	0.173	26	25	0	0
Spain North	103	66	23	4	19	4.75	0.164	0.275	22	21	0	0
Spain South/Centre	106	62	16	5	11	2.2	0.599	0.705	16	15	0	0
Sweden	32	37	8	2	6	3	0.518	0.611	8	7	0	0
Switzerland	224	92	33	9	24	2.67	0.418	0.656	32	33	2	0
Syria	69	73	24	7	17	2.43	0.509	0.639	20	19	0	0
Turkey	290	128	63	30	33	1.1	0.101	0.07	69	79	11	0
Wales	92	49	13	3	10	3.33	0.398	0.513	13	13	1	0

n = sample size
A = number of polymorphic sites between HVS-1 positions 16090-16365
B = Total number of "Weighty" polymorphic sites
C = Number of "Weighty" Transitions
D = Transversions and INDELS
E = WTTI Ratio
F = Fisher's Exact test *p* -value for significant deviation of the WTTT ratio compared to the Irish population
G = Fisher's Exact test *p* -value for significant deviation of the WTTT ratio compared to the Icelandic population

Table 2.5 *Summary of HVS-1 quality checks in 45 European and Middle Eastern sample populations.*

2.3.4 European mtDNA Population Affinities

The relationship of 45 European and Middle Eastern populations was summarised by calculating genetic distances between them based on HVS-1 sequences. Each population dataset had previously been checked for errors as outlined in Materials and Methods (**Section 2.2.7**) and these findings are shown in **Table 2.4.** None of the Weighty transitions to transversions/Indels (WTTI) ratios were significantly different, at the 0.05% level using Fisher's exact test, from the benchmark Irish or Icelandic datasets. In addition, no population shows an unambiguously inflated cube spectrum or the presence of large numbers of higher dimensional cubes in their networks. These methods cannot exclude the possibility of all error but do strongly argue that the dataset as whole is not unduly influenced by this potential confounding factor.

The 44 dimensions of the inter-population Φ_{ST} matrix are reduced to two-dimensional space in **Figure 2.2**. A broadly east–west or southeast–northwest trend is evident in the first dimension with the Jordanian and Basque population samples occupying the respective poles. This echoes the trend observed in a principal component analysis of haplogroup frequencies in European regions (Richards et al. 2002). While the extreme value of the Basques in Dimension 1 is unique, many Northwest Atlantic area populations, including Ireland and other Celtic regions of the British Isles (but excluding) Cornwall, occupy positions on the edge of the general European range toward the Basque pole.

This can be seen more clearly in **Figure 2.3A** where the 1st Dimension values are displayed as a synthetic surface map of Europe. Atlantic European samples, including those from Ireland, Wales, Scotland and Galicia, as well as Iceland and Norway, show a common affinity centred on the Basque country. It is notable that most of the centres of gravity estimates with low and intermediate diffusion (SD <500 km and 500-1000 km

respectively) are encompassed within or adjacent to this Atlantic zone. There is also evidence for a degree of mtDNA differentiation between western/northern areas of the British Isles and Southern England. The Iberian Peninsula is notable as an area of steep north–south gradient, with the north more similar to Central and Western Europe and the south more similar to Mediterranean Europe and the Near East. The 2nd Dimension does not appear to display any obvious geographical pattern, but does distinguish Ireland, Scotland, Wales, Iceland and Galicia from areas in Fenno-Scandinavia which also have low values in Dimension 1. Spatial autocorrelation analysis confirms that Dimension 1 values are consistent with a clinal pattern across Europe while those in Dimension 2 are not (**Figure 2.4**).

2.3.5 European Y chromosome and Autosomal Variation

In addition to mtDNA, variation in other loci with different modes of inheritance was also examined. Dimension 1 of Y chromosome binary marker haplotype variation, displayed as a synthetic map (**Figure 2.3B**), shows a broadly similar pattern to mtDNA Dimension 1 with a prominent gradient from the Near East to Western Europe. A similarity between Atlantic coastal areas is again evident and indeed shows a stronger affinity between Ireland/Western Britain and the Basque region than is found with mtDNA. Cornwall (where the last Cornish speaker died in 1891) shares this Atlantic affinity in Y chromosome data but not mtDNA. Brittany, an area with a Celtic language and putative close links to Cornwall, also appears to be an exception in the general mtDNA landscape. Autosomal variation was analysed using a matrix of inter-population F_{ST} values calculated from classical gene frequencies in 25 European populations. This did not include any Near Eastern populations, but nonetheless a similar trend is observed (**Figure 2.3C**). However, unlike the Y chromosome or mtDNA this is clearer in the 2nd Dimension.

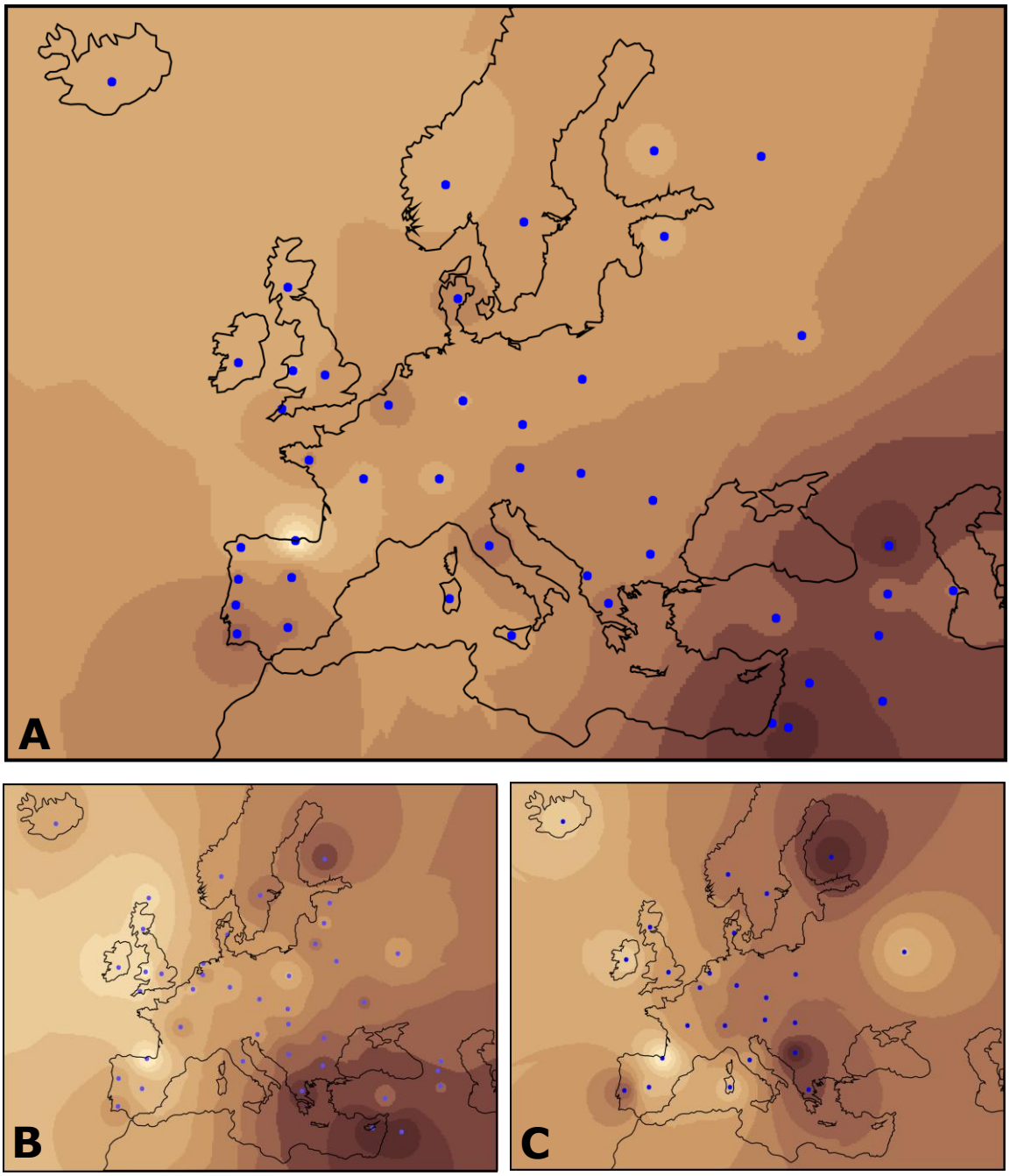


Figure 2.3 *Synthetic surface maps of Europe displaying the three significantly correlated dimensions of genetic variation. These are (A) the 1st Dimension of mtDNA variation; (B) 1st Dimension of Y chromosome diversity and (C) 2nd Dimension derived from classical gene frequencies. Blue points indicate sample locations.*

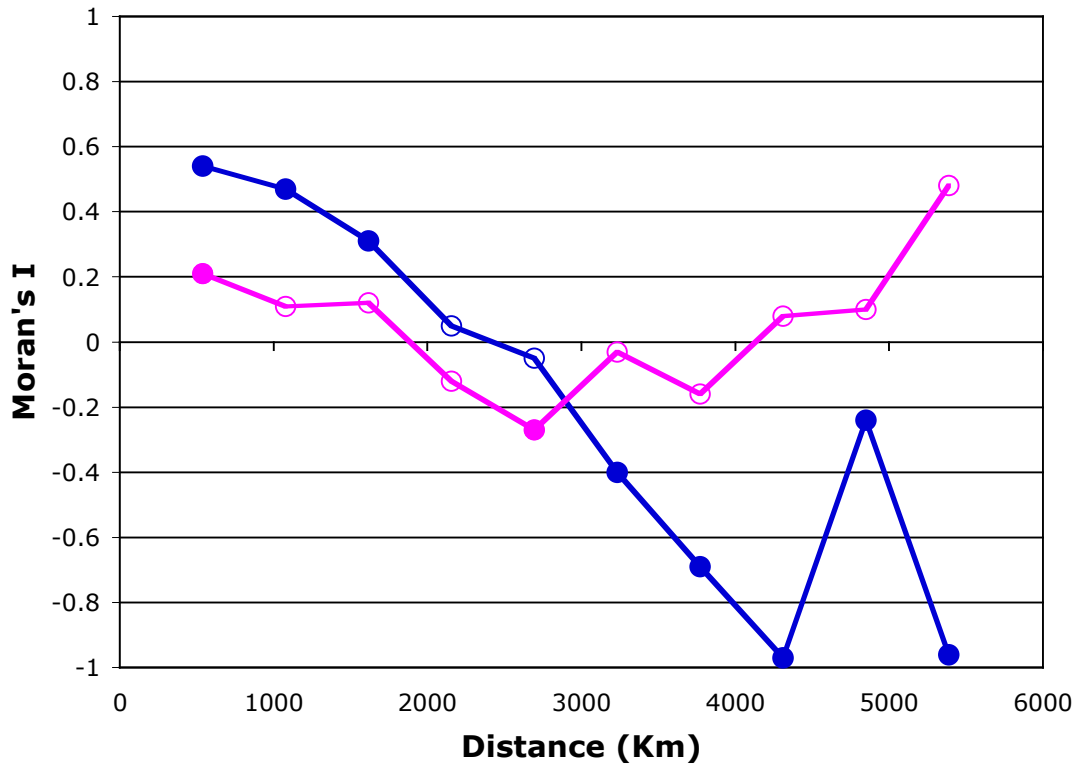


Figure 2.4 *Spatial autocorrelation analysis of mtDNA Dimension 1 (blue) and 2 (pink) values from 45 European and Middle Eastern populations. Correlograms were generated using the PASSAGE package (Rosenberg 2001) with 10 equal interval geographic distances classes. Significant individual distance classes in each dimension are represented by filled symbols. The gradual change from significant positive correlation over short distances to significant negative correlation over longer distances is consistent with a broad clinal trend of Dimension 1 across Europe. The overall significance of the correlogram for Dimension 1 is p <0.05. In contrast, mtDNA Dimension 2 shows no significant continental wide pattern.*

2.3.6 Congruence of mtDNA and other Genetic Systems

Pearson's correlation coefficient was used to assess the congruence of the major trend in mtDNA variation (Dimension 1, **Figure 2.3A**) with, firstly, the two dimensions of Y chromosome variation and, secondly, both dimensions of autosomal diversity. Using 28 populations in common, mtDNA Dimension 1 correlated strongly and significantly with Y chromosome Dimension 1 (r=0.714 and p=0.033, corrected for two comparisons) but not with the second dimension (r=0.317, p=0.309). Conversely, Dimension 2 in autosomal variation is correlated with the predominant mtDNA trend over 19 shared populations (r=0.528, p=0.043), while Dimension 1 is not (r=0.141, p=0.86).

2.4 DISCUSSION

2.4.1 European mtDNA Structure

Studies of Y chromosome variation have previously demonstrated strong levels of differentiation within Europe (Rosser et al. 2000), with variation in autosomal loci often exhibiting similar structure (Cavalli-Sforza et al. 1994). However, a detailed portrait of mtDNA structure in Europe has hitherto remained elusive. Yet concordance between different marker systems is an important means of demonstrating that geographical patterns are the result of demographic history and not, for example, of selection. These results strongly suggest for the first time that the demographic histories of Europe, in general, and Ireland, in particular, are very similarly recorded in loci with different modes of inheritance.

The use of an extensive European and Middle Eastern data-set, checked for quality and divided into fine population units is likely to have aided the resolution of hitherto undetected mtDNA patterns. Direct use of sequence, with appropriate correction for the high mutation rate of the mtDNA control region, takes full advantage of intra-haplogroup variability. It is a corollary of this approach that analysis will be less confounded by the potential error from placing disparate lineages into broad and sometimes still ill-defined haplogroups. For example, haplogroup H can be subdivided into finer lineages, each with a potentially distinct history (Achilli et al. 2004; Loogväli et al. 2004).

2.4.2 Atlantic Façade Genetic Legacy

The affinity of Atlantic regions of Europe is a recurring and enduring feature of the European genetic landscape seen here. Particular affinities within the Atlantic zone have been suggested by the distribution of Y chromosome haplogroup R1b (which reaches

frequencies approaching 100% in some parts of Western Europe) and mtDNA haplogroup V along with the H1 and H3 sub-groupings of haplogroup H (Torroni et al. 1998; Hill et al. 2000; Semino et al. 2000; Torroni et al. 2001; Wilson et al. 2001; Achilli et al. 2004). During the last glacial maximum, human habitation is thought to have been largely restricted to refugial areas in Southern Europe. One of the most important of these is likely to have been in Southwest France/Iberian Peninsula (Dolukhanov 1993; Housley et al. 1997; Gamble et al. 2004). The re-colonisation of Western Europe from an Iberian refugium, following the retreat of the ice-sheets from ~15,000 years ago, could explain the common genetic legacy in this area.

An alternative, but not mutually exclusive, model would place Atlantic fringe populations at the 'Mesolithic' extreme of a Neolithic demic expansion into Europe from the Near East (Ammerman and Cavalli-Sforza 1984). In any event, the preservation of this signal within the Atlantic Arc certainly suggests that this region was relatively undisturbed by any subsequent migrations either from Central Europe or beyond. The identification of likely dispersal points for some Irish haplotypes in Northern Spain and Western France is further evidence for links between Atlantic populations. Cunliffe (2001) has used the French historian Fernand Braudel's term, '*longue durée*', or the concept of a long-term historical perspective, to describe the sedimentation of traditions on the Atlantic façade. He suggests these links may stem from the late Mesolithic, perhaps even predating the coming of agriculture to the region. These results support the view that the genetic legacy, at least, of the region may trace back this far and perhaps even to the earliest settlements following re-colonisation after the Last Glacial Maximum (LGM).

Alternatively, the genetic affinity of these regions might be explained by restricted patterns of long-term gene flow facilitated by the Atlantic seaways. It is difficult to

distinguish genetically between a common Palaeolithic origin and more recent contacts nor are they mutually exclusive. However, haplogroup R1b3f Y chromosomes, which have a recent origin in Iberia (Hurles et al. 1999), have not been found in Ireland (Hill et al. 2000), arguing against the migration of very large numbers of men by this route, at least in the last 2000–3000 years. This would be consistent with the suggestion that most contacts over this period would have been small-scale, perhaps in the manner of the *kula* ring in the western Pacific (Cunliffe 2001). On the female side, the presence of putatively Neolithic mtDNA haplogroups in Ireland does indicate some gene flow with the continent after the initial peopling of the island, about 9000 YBP, following the post-glacial population re-expansion (cf. Wilson et al. 2001). However, this could have been at any time in the last 6000 years or so; the beginning of the Neolithic in the area simply represents their earliest possible arrival.

2.4.3 Irish Maternal and Paternal Past

While the patterns seen in different loci are significantly correlated, there appears to be some variation in scale, with sharper differentiation seen in the Y chromosome relative to mtDNA. This general European feature finds local expression in Ireland with the significant differences between East and West Irish Y chromosomes, attributed to the preferential settlement of migrants on the accessible East coast after initial colonisation (Hill et al. 2000), not matched in mtDNA. Such an observation is not unexpected given the greater subtlety of mtDNA differentiation between the island and potential source populations. Higher mtDNA diversity levels in East Ireland relative to Western areas may represent the faint indication of parallel movements. However, the even distribution of putative Neolithic haplogroups around the island suggests that females who arrived post initial settlement were not restricted to East facing regions. By contrast, Y chromosome lineages of putative Near Eastern Neolithic origin (Semino et

al. 2000) appear to be virtually absent from the West of Ireland (Hill et al. 2000). There are two obvious potential explanations from the maternal perspective for this discrepancy. Women may have been more mobile than men (see Seielstad et al. 1998; Oota et al. 2001) after arrival in the East or secondly, other regions of the island could have been in direct contact with the continent and the migrant source populations perhaps *via* the proposed Atlantic façade maritime network. Alternatively and/or additionally, aspects of the paternal past such as high reproductive variance, either before or after settlement in Ireland, may have exaggerated the Y chromosome pattern. In any event, some asymmetry is expected to result from inexact matching of sample populations and molecular marker types (Wilder et al. 2004; Stoneking 1998).

2.4.4 Genetic Heterogeneity in the British Isles

These results also demonstrate the existence of a degree of genetic heterogeneity in the British Isles, at least in the Y chromosome and much more tentatively with mtDNA. Southeast England tends to show a greater affinity to neighbouring areas of continental Europe compared to Ireland and, more surprisingly, other physically adjoining areas of Britain. Anglo-Saxon mass-migration has been proposed, and indeed sometimes presumed, to be the explanation for this pattern in Y chromosome variation (Weale et al. 2002; Capelli et al. 2003). Such explanations may seem feasible for the Y chromosome, given the high levels of drift that might be associated with disproportionately high numbers of offspring amongst conquering elite males. However, the weight of archaeological evidence is against population replacement associated with the Anglo-Saxons conquest (Esmonde-Cleary 1989), suggesting that alternative explanations should be considered. It may be that the genetic landscape of Southeast Britain has been shaped by older links with the continent, perhaps during the

Neolithic, or even before the filling of the North Sea when Britain was connected to the continent *via* the Doggerland plain (Coles 1998).

2.4.5 Close Affinity of Scotland and Ireland

While grouped population relationships have proved extremely informative, a parallel and sophisticated phylogeographic approach can augment these findings and also detect specific and subtle contacts that might otherwise be missed at population level comparisons. The tentative indications of gene flow between Ireland, Iceland and possibly Scandinavia are in agreement with historical sources and other genetic evidence (Helgason et al. 2000a; Helgason et al. 2000b; Helgason et al. 2001) that the Viking founders of Iceland sourced some of the initial female settlers from their activities in the British Isles (see also Chapter 6).

However, the multiple mtDNA links between Ireland and Britain, particularly Scotland, are the most striking finding of the phylogeographic analysis. The strong maternal relationship it suggests is reflected in the close mtDNA population level affinity of the two countries. In addition, there is a wealth of independent evidence of interface between the two, especially through Northern Ireland (Ulster), facilitated by close physical proximity. The geographical distribution of mutations in the phenylhydroxylase gene has provided separate genetic confirmation of Ulster as a zone of considerable admixture (O'Donnell et al. 2002). Archaeological evidence concurs, demonstrating contacts during prehistory, while early historical accounts describe the establishment of Irish colonies in Scotland from at least ~500 AD. Indeed, the name 'Scotland' derives from the Latin for Ireland (*Scotia*) at this time (Duffy 2000). Linguistically, modern Scottish Gaelic is a clear derivative of the Irish language (see Forster and Toth (2003) for a novel genetics inspired philological demonstration of this). During the 16th and 17th centuries the substantial direction of movement was

reversed with the plantation of Ulster and arrival of Scottish settlers ('planters'). However, Irish haplotypes with distribution centres in Scotland are not simply concentrated in Ulster, but rather are widely distributed throughout the island, strongly suggesting the links are primarily the result of earlier contacts (see also Chapter 4). Evidence, perhaps, of a local '*longue durée*' relationship dating back to the earliest peopling of Ireland.

2.4.6 Conclusions

These results demonstrate for the first time that the same prominent trends in European genetic variation are recorded in loci with different inheritance modes. It seems clear that neither mtDNA, Y chromosome nor autosomal markers support a substantially central-European Iron-Age origin for most Celtic- or former Celtic-speakers of the Atlantic façade. The affinities of the areas where Celtic languages are spoken, or were formerly spoken, are generally with other regions in the Atlantic zone, from Northern Spain to Northern Britain. Whilst some level of Iron-Age immigration into Ireland and Britain could probably never be ruled out using modern genetic data, these results emphasise genetic continuity in the British Isles and point to a distinctive Atlantic genetic heritage whose roots probably lie in processes at the end of the last Ice Age.

CHAPTER 3

LONGITUDINAL AUTOSOMAL GENETIC VARIATION IN IRELAND

3.1 INTRODUCTION

3.1.1 Longitudinal Gradients of Variation in Ireland

Significant differences between eastern and western regions of Ireland have been a recurring finding of studies into Irish population variation. The earliest attempts to glean historical inference from human diversity inevitably focused on phenotypic differences. In the 19th century, the English 'anthropologist' John Beddoe (1885) noted some distinctive physical characteristics of people in the West of Ireland, compared to the East, such as a high frequency of dark hair. However, his dubious methods and opinions make his work more notable today as a historical curiosity rather than reliable scientific study. For example, on some people of Western Ireland, he remarked that "though head is large, intelligence is low and there is a great deal of cunning and suspicion"!

A somewhat more enlightened 'Harvard Anthropological Survey of Ireland' was conducted between 1934 and 1936 and collected anthropometric information (such as head circumference, nose length and other physical characteristics) on 11,000 Irish people. This study also provided support for longitudinal regional differences with, for example, a gradient in skin type through density of freckles from east to west (Croke 2000 and reference therein). Later and more sophisticated statistical treatments of the same data claimed similar findings with a prominent longitudinal gradient from East to West Ireland seen in the variation of 17 such characteristics (Relethford and Crawford 1995).

3.1.2 Irish Blood Group Frequencies

Advancing technology allowed the investigation of human variation more closely connected to its underlying genetic source. Large-scale studies of blood group frequencies provided some of the clearest evidence for east-west substructure, detecting converse reciprocal clines across the island in the frequency of ABO blood group A and O alleles. Parallel trends were also seen in the Rhesus positive/negative blood group alleles (Hackett and Dawson 1958; Dawson 1964). The O allele increases in frequency from East to West Ireland while the A and rhesus negative (Rh-) alleles show the reverse pattern (**Figure 3.1**). The trend was attributed to the increased influence of England, where blood group A is more common, on the East Coast, while blood group O was proposed to represent an earlier indigenous type. The latter interpretation is consistent with the observation that the intra-Ireland pattern is the culmination of a general gradient of increasing blood group O frequency across Europe in a southeast to northwest direction. However, further (univariate) studies of different blood proteins provided little additional evidence for any substructure within Ireland (Tills 1977; Tills et al. 1977).

3.1.3 Intra-Ireland Molecular Genetic Variation

More recent molecular genetic studies have continued to provide new examples of Irish longitudinal variation. Hyperphenylalaninaemia are a group of inherited metabolic disorders that include Phenylketonuria (PKU), in which levels of plasma phenylalanine are elevated. They can result from various mutations in the phenylalanine hydroxylase gene on chromosome 12. One of the most common mutations (R408W) displays a local cline of increasing frequency across Northwest Europe climaxing in Ireland.

Furthermore, echoing the O blood group, the trend continues within Ireland, with the R408W mutation more common in Connacht than Leinster (O' Donnell et al. 2002).

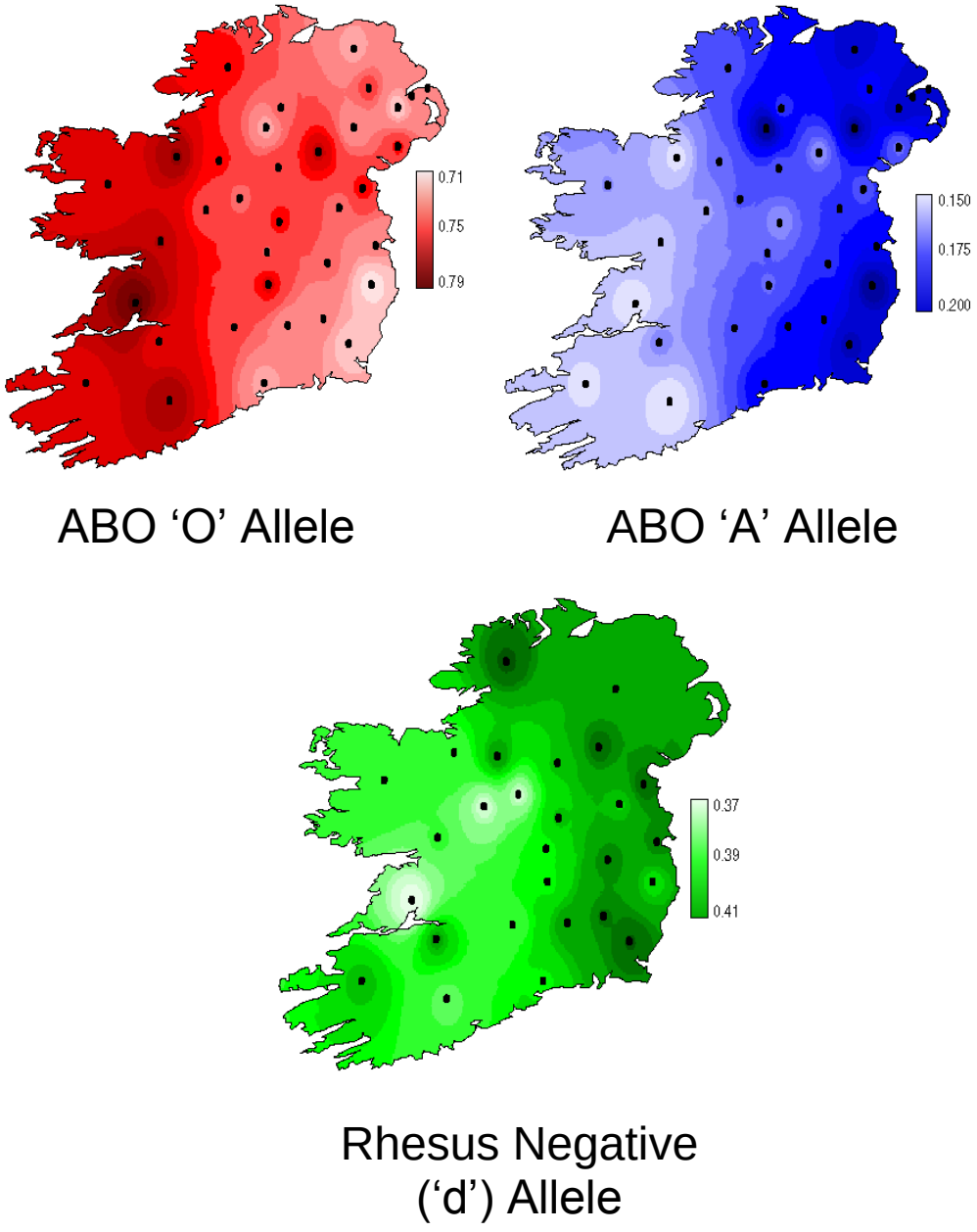


Figure 3.1 *Synthetic surface maps of Ireland showing geographic variation in blood group frequencies. The O and A alleles of the ABO blood group demonstrate converse frequency gradients across the island in a roughly east-west direction. A similar trend is observed in the rhesus negative blood group allele. Sample locations are indicated by black points. Data for the Republic of Ireland was taken from Dawson (1964) while that for Northern Ireland was extracted from the database of CGF assembled by Cavalli-Sforza et al. (1994) and provided by Christopher A. Edmonds, Stanford University, California. Maps were constructed in Arcview (Version 3.2, Environmental Systems Research Institute Inc.) by interpolating observed values to generate 20 equal interval contours.*

The paternally inherited Y chromosome shows a similar trend, with increasing frequency of haplogroup R1b from east to west leading to significant regional differences that once again slot into a broader European cline of increasing R1b frequency from Southeast to Northwest Europe. In contrast, neither maternally inherited mtDNA (Chapter 2) nor X chromosomes (Moore 2004) show any obvious longitudinal differences. Nonetheless, the repeated culmination in Ireland of continental wide trends in different loci is frequently interpreted as reflecting the relatively undisturbed Palaeolithic genetic legacy of the Irish. In this scenario east–west regional differentiation is the product of greater accessibility and thus contact of Eastern Ireland with Britain and the continent.

3.1.4 Detecting Regional Substructure

Investigation of group affinities traditionally involves prior population specification, using a criterion such as place of origin or linguistic affiliation, followed by the reconstruction of genetic relationships. Although convenient, all 'population' definitions are inevitably subjective to a greater or a lesser extent. They may not correspond to real genetic sub-divisions, which will consequently go undetected. An alternative approach begins without prior assumptions on populations but rather infers these divisions from the genetic relationships. For example, genetic distances between individuals can be used to construct phylogenetic trees and groupings or clusters thus identified. However, this can also be subjective and influenced by the choice of distance measure and output visualisation.

To circumvent these problems, Pritchard et al. (2000) developed a model-based clustering approach implemented in the program *structure*. The method uses multi-locus genotypes to infer population groupings and assign each individual to these

populations. Its operation stems from the 18th century theorem by the Reverend Thomas Bayes (1702-1761) (Shoemaker et al. 1999; Beaumont and Rannala 2004). Under the Bayesian framework there is no logical distinction between an outcome (set of results) and the processes (parameters) that control that outcome. Assuming Hardy-Weinberg equilibrium (HWE), the genotype (*X*) for any individual is dependent on the population of origin (*P*) and the allele frequencies that characterise that population (*Z*). Using a Bayesian approach it possible to work 'backwards' and find these using only genotypes and certain prior assumptions on the nature of *P* and *Z*. These 'priors' are a convenient way of incorporating any additional (or a lack of) background information about the parameters of interest. In effect the method seeks to introduce substructure or creates populations to ensure total HWE. Solutions are explored iteratively using a Markov chain Monte Carlo (MCMC) scheme. The result of one step in the MCMC is dependent on the preceding step, such that over a sufficiently large number of iterations, the chain follows a trail to the result with the highest probability.

3.1.5 Present Study

Genetic differences between East and West Ireland have often been attributed to the preferential migration of post-foundation settlers to eastern regions of the island, presumably due to the geographic proximity of Britain and, *via* there, the rest of Europe. Under this scenario Britain and the West of Ireland represent migration source and destination populations respectively, with eastern regions of the island a zone of contact and admixture between the two. The impact and signature of any such process on the autosomal genome is examined using two classes of markers. Initially, intra-Ireland classical gene frequency (CGF) variation is explored in the context of a European dataset. Secondly, regional British Isles variation is examined in 380 STR loci, spread across the entire autosomal genome, using both traditional genetic distance and model-based clustering approaches.

3.2 MATERIALS AND METHODS

3.2.1 Classical Gene Frequency Analysis

Cavalli-Sforza et al. (1994) assembled Classical Gene frequency (CGF) data from numerous world populations, including information on 29 loci for Ireland. An examination of source literature identified regional East-West Irish allele frequencies, defined as the provinces of Leinster and Connacht respectively, for 19 of these loci (**Table 3.1**). Their frequencies in 27 additional European populations were taken from Cavalli-Sforza et al. (1994).

CLASSICAL GENE	CHROMOSOME	REFERENCE
Rhesus Blood Group	1	Tills et al. 1977
MNS Blood Group	4	Tills et al. 1977
ABO Blood Group	9	Tills et al. 1977
Lutheran Blood Group	19	Tills et al. 1977
Kell Blood Group	7	Tills et al. 1977
Duffy Blood Group	1	Tills et al. 1977
Haptoglobin	16	Tills 1977
P Blood Group	22	Tills et al. 1977
Acid Phosphatase	2	Tills 1977
Phosphoglucomutase 1	1	Tills 1977
Adenylate Kinase 1	9	Tills 1977
Phosphogluconate Dehydrogenase	1	Tills 1977
Adenosine deaminase	20	Tills 1977
Beta lipoprotein, Ag System	2	Tills 1977
Transferrin	3	Tills 1977
Beta lipoprotein, LPA System	6	Tills 1977
Group Specifc Component	4	Tills 1977
Phenylthiocarbamide tasting	7	Sunderland et al. 1973

Table 3.1 *19 Classical Gene loci used to examined intra-Ireland autosomal substructure*

Genetic distances between populations as conventional F_{ST} statistics, based on allele frequencies, were calculated for each locus using a purpose written Microsoft ® Excel Macro (Stephen Park) as follows:

$$F_{ST} = \frac{\sum_{i=0}^{i} (p_1 - p_2)^2}{\sum_{i=0}^{i} 2p(1-p)} \qquad \textbf{(Equation 3.1)}$$

where *i* is the number of alleles at each locus, P_1 and P_2 are the frequencies of the allele in the first and second population respectively and $\bar{P}$ is the mean frequency of both populations. An inter-population F_{ST} matrix was calculated for each locus and these were averaged across all loci by summing denominators and numerators separately. The combined matrix was then subjected to MDS analysis (see **Section 2.2.9**) using the ALSCAL program included in SPSS package (Version 11.0, SPSS Inc.)

3.2.2 Autosomal STR Markers and Samples

Several hundred genome-wide STR loci were typed in British and Irish families in the course of a linkage scan for chromosomal regions contributing to bipolar affective disorder (Bennett et al. 2002). These loci (Applied Biosystems' Medium Density 10cM linkage mapping set) included 380 loci from the autosomal Chromosomes 1 to 22. The permission of Professor Michael Gill and the assistance of David Lambert, Ricardo Segurado and Aidan Corvin (Department of Genetics, Trinity College, Dublin) in providing the raw genotypic data from this study are gratefully acknowledged.

Linkage studies are family based and therefore, by definition, include related individuals. To adapt the data for population purposes, it was initially filtered to remove related individuals yielding a total sample population of 194. Compete 380 loci genotypes were then constructed for each individual using specially written PERL language software to manipulate and assemble the data. 119 individuals originated in the United Kingdom while the remaining 75 came from Ireland. More detailed information on a place of origin for 61 of the latter was possible and an East (*n*=21) and West Irish cohort (*n*=40) was defined from this information. The Eastern cohort was composed of samples (*number*) from Dublin (*14*), Wicklow (*1*), Wexford (*3*), Antrim (*1*) and Kildare (*1*). The Western grouping included samples from Sligo (*3*), Kerry (*3*), Mayo (*1*), Tipperary (*1*), Roscommon (*2*), Cork (*27*), Clare (*1*), Longford (*1*) and Galway (*1*).

3.2.3 Analysis of Autosomal STR Allele Frequencies

The first analytical approach uses the information on sample origin in Ireland to examine the relationship of predefined populations. Although the number of loci is large there was no extensive European comparative data available. A variety of approaches were therefore undertaken to maximise the amount of information that could be drawn from the dataset.

Initially, to maximise the number of loci used, analysis was internally reliant on the British Isles sample population alone. Genetic distances as linearised F_{ST} values based on allele frequencies were calculated between the UK population and the East and West Irish cohorts for each locus in turn using ARLEQUIN (Version 2.000) (Schneider et al. 2000). The numbers of loci showing a closer relationship (smaller F_{ST} value) of the UK to either East or West Ireland were then summed.

In the second approach, most loci were sacrificed in a compromise to include an additional comparative European population. Frequency data for 69 STR loci in a Danish population, typed in common with the present data set, were retrieved from the **al**lele **fre**quency **d**atabase (**ALFRED**) (www.alfred.med.yale.edu/alfred/index.asp). Genetic distances were calculated as above between Denmark, as a representative of continental Europe, and the UK and Ireland (as whole and East and West regions separately).

Finally, maximisation of the number of comparative populations was prioritised. 12 loci (D1S468, D3S1262, D3S1311, D4S403, D5S408, D10S189, D10S212, D13S285, D15S128, D15S165, D16S516, D17S784) had been typed in common between the British Isles samples and 1056 individuals from 52 world populations described in Rosenberg et al. (2002). Nine European and Middle Eastern populations were extracted from this data (*sample size*): France (*29*), Basques (*24*), Russia (*25*), Sardinia (*28*), Orkney Islands (*16*), Druze (Israel) (*48*), Palestinians (*51*) and Italy (*22*) (composed of Bergamo and Tuscan samples). Principal component analysis (PCA) of the 134 alleles observed across the 12 loci was carried out using the program POPSTR (Henry Harpending, University of Utah). Secondly, genetic distances as linearised R_{ST} values (an F_{ST} analog for STR data under the stepwise mutation model) were calculated using ARLEQUIN, based on individual 12 STR loci genotypes within each population. The resulting inter-population matrix of genetic distances was then subjected to MDS analysis.

3.2.4 Autosomal STR Model-based Clustering

The extent of any substructure in the British Isles was also explored using a model-based clustering method as implemented in the program *structure* (Version 2.0) (Pritchard et al. 2000; Falush et al. 2003), (http://pritch.bsd.uchicago.edu/). Although this approach creates populations based essentially on genotypic data for each individual, it does require the prior specification of the number of populations (*K*) into which the data are to be divided. In the absence of any prior knowledge of this value and in line with the Bayesian approach, *ad hoc* estimates of the likelihood of the data under various *K* values can be used to determine the number of populations that are most compatible with the observed genotypes.

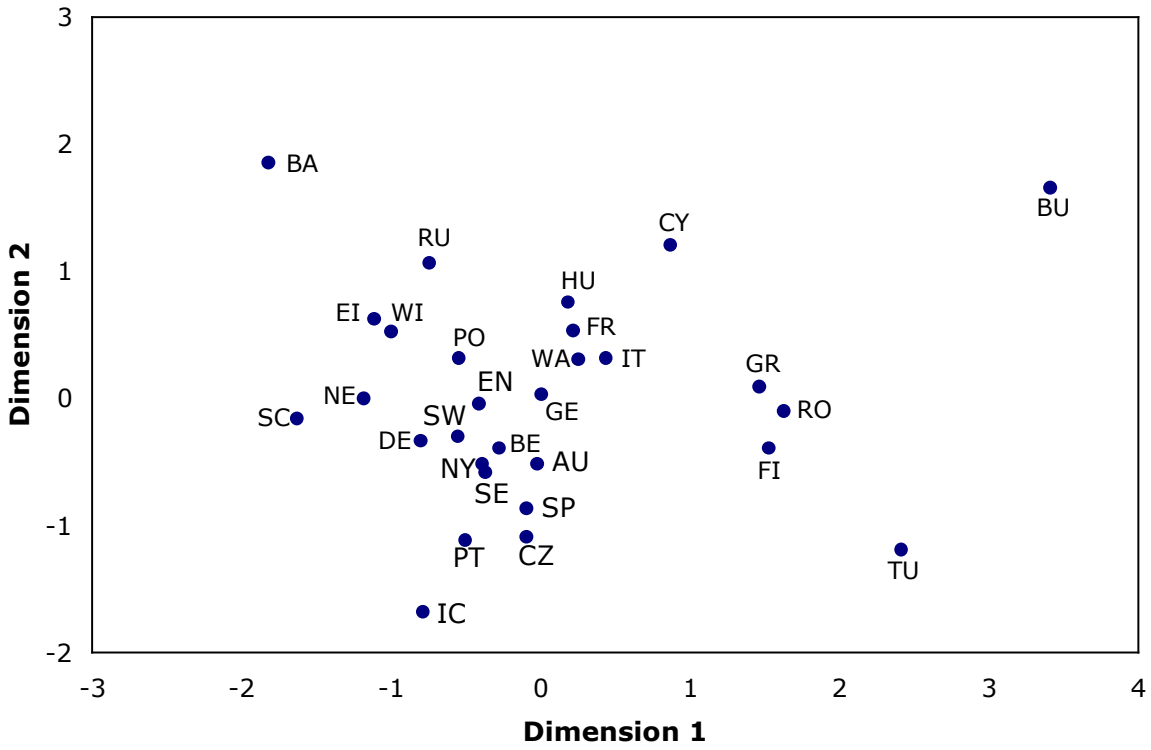
Each *structure* run used 250,000 MCMC steps following an initial burn-in of 50,000 iterations, which were subsequently discarded. It was found by exploration that significantly longer chains (up to 10^6 iterations with a 100,000 burn-in steps) did not alter the findings. Five independent runs were carried out for values of *K* 1 to 5 (a total 25 runs.) All analysis was carried out under the correlated allele frequencies model. This assumes that allele frequencies between neighbouring populations are likely to be similar rather than independent, a reasonable assumption given the small region under investigation. The possibility of admixture was also permitted, where individuals may derive proportions of their genome from different populations. The likelihood of the data under each *K* was averaged over the 5 runs and the probability of each *K* value calculated as described in Pritchard et al. (2000).

3.3 RESULTS

3.3.1 Classical Gene Frequencies

A total of 19 classical gene loci were identified where regional East and West Irish frequencies were available. MDS analysis of inter-population genetic distances between the two regions and 27 other European populations is shown in **Figure 3.2**. Despite the relatively small number of loci the most prominent trend (Dimension 1) identifies the major southeast to northwest trend of European variation, with Basque and Bulgarian (closely followed by Turkey) population samples anchoring the respective poles. Ireland and other places in Northwest Europe like Scotland, occupy a position on the edge of the main concentration of European populations toward the Basque pole. At the regional level, East and West Ireland have similar values in both dimensions. There is no evidence of a stronger relationship of East Ireland to England or other continental populations.

Figure 3.2 *MDS plot showing the relationship of 29 European populations calculated from 19*



classical gene frequencies. **(Full legend overleaf)**

Figure 3.2 (preceding page) *MDS plot showing the relationship of 29 European populations calculated from 19 classical gene frequencies. The two dimensions preserve 83% of the original variation from the inter-population F_{ST} matrix. East and West Ireland (EI and WI respectively) occupy a similar position on the fringe of most other European populations and toward the Basque pole of Dimension 1. Other population labels are as follows: AU = Austria; BA=Basque Country; BE = Belgium; BU = Bulgaria; CY=Cyprus; CZ = Czech Republic and Slovakia; DE = Denmark; EN = England; FI = Finland; FR = France; GE = Germany; GR = Greece; HU = Hungary; IC = Iceland; IT = Italy; NE= Netherlands; NY= Norway; PO = Poland; PT = Portugal; RO = Romania; RU = Russia; SC = Scotland; SE = Sweden; SP = Spain; SW = Switzerland; TU = Turkey; WA = Wales.*

3.3.2 Autosomal STR Genetic Distances

380 STR loci spread across all 22 nuclear autosomal chromosomes were analysed in 194 individuals from Ireland and Britain. Substructure was initially investigated using traditional distance based methods with predefined geographic populations. F_{ST} values were calculated between East/West Ireland and the UK for each locus. 124 (32.6%) of these returned zero values for both East and West Ireland and were thus completely uninformative. Of the remaining 256, the majority (141 or ~55%) display a closer relationship of the UK to West rather than East Ireland; the converse was true in 115 instances (**Figure 3.3A**). However, a χ^2 test shows this difference to be insignificant (p= 0.104). Accordingly, there is no evidence for a general greater affinity of East Ireland to the UK or the regional substructure this would imply. A similar analysis for a subset of 69 loci was carried out using Denmark as an external continental reference population. There was no evidence for a significantly greater affinity of the UK over Ireland, nor East over West Ireland, to the continental Danes (p=0.475 and p=0.456 respectively) based on the observed distribution of F_{ST} relationships in these loci (**Figure 3.3B** and **3.3C** respectively).

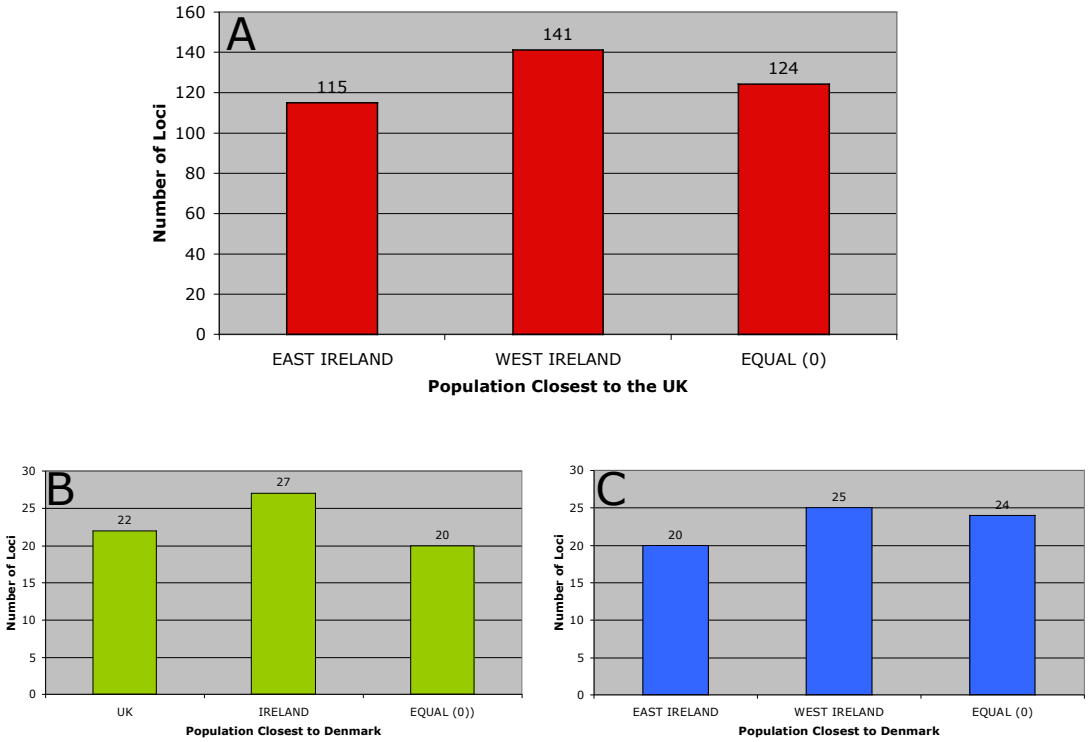


Figure 3.3 *Summary of autosomal STR locus-by-locus genetic distance analysis. (**A**) The relationship of the UK to East and West Ireland over 380 loci. (**B**) The relationship of Denmark to the UK and Ireland, or (**C**) Denmark to East and West Ireland sepearately, over 69 loci. Equal or zero F_{ST} values indicate that the locus could not provide any information on relative population relationships.*

Next, the broader relationship of East and West Ireland to nine other European and Middle Eastern populations was explored using a subset of 12 loci. PCA of the frequency of 134 alleles is shown in **Figure 3.4A**. Despite the relatively small number of loci, a general Basque to Middle East trend forms the first principal component. Both East and West Ireland occupy an intermediate position in this trend close to other British Isles populations (United Kingdom and Orkney Island). Parallel MDS treatment of genetic distances (R_{ST} values) (**Figure 3.4B**) using the same data once again identifies Basque-Middle East polarity. However, the positions of East and West Ireland are somewhat divergent. Although, both are far from the Basque pole, Western Ireland is relatively closer compared to East Ireland. However, the variability in the same data using different approaches indicates that little robust inference can be drawn.

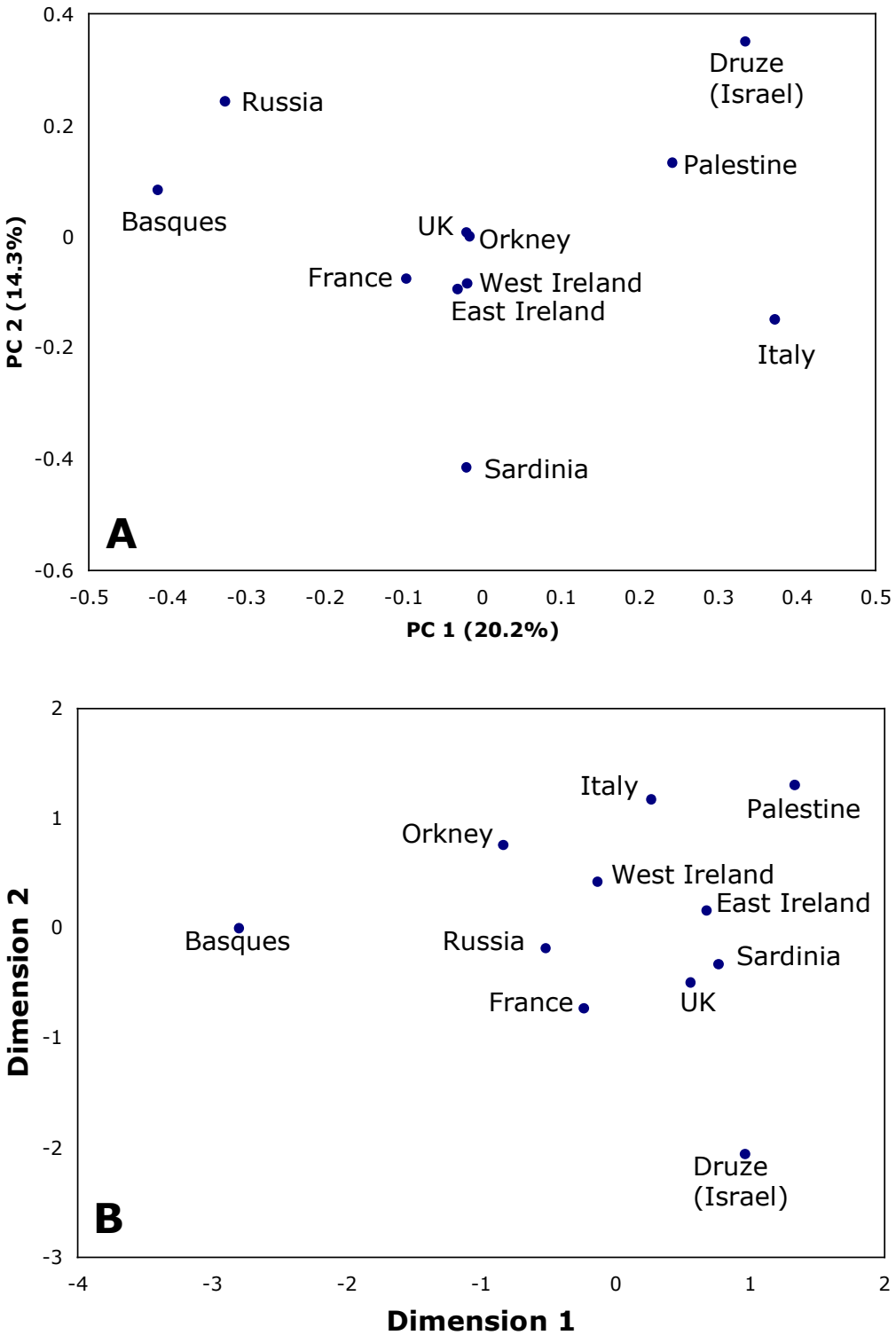


Figure 3.4 *Genetic relationship of 11 European and Middle-Eastern populations based on 12 autosomal STR loci. (**A**) PCA of 134 allele frequencies. Principal Component (PC) 1 accounts for 20.2% of the variation with 14.3% contained in PC 2. (**B**) MDS analysis of a matrix of inter-population R_{ST} values. The two dimensions retain 84% of the original variation. While both show a Basque-Middle East polarity as their major trend, the relative positions of East and West Ireland are changeable.*

3.3.3 *Structure* Analysis of Autosomal STR Genotypes

A model-based clustering method, as implemented in the program *structure*, was also employed in the search for substructure. The method requires the prior specification of the number of populations (*K*) into which data are to be divided. Calculations from the likelihood of the data under different values of *K* placed a ~100% probability on *K*=2. However, it should be noted that the average difference in likelihood between *K*=1 and *K*=2 is small compared to the variance within runs of the same *K* indicating low reliability in what is, in any event, an *ad hoc* approximation. **Figure 3.5** illustrates the typical results for a series of *K* values from 2 to 5. Regardless of the number of population divisions, most individuals are not strongly (unambiguously) assigned to any particular grouping indicating little robust or meaningful structure even for *K*=2, the highest likelihood *K*.

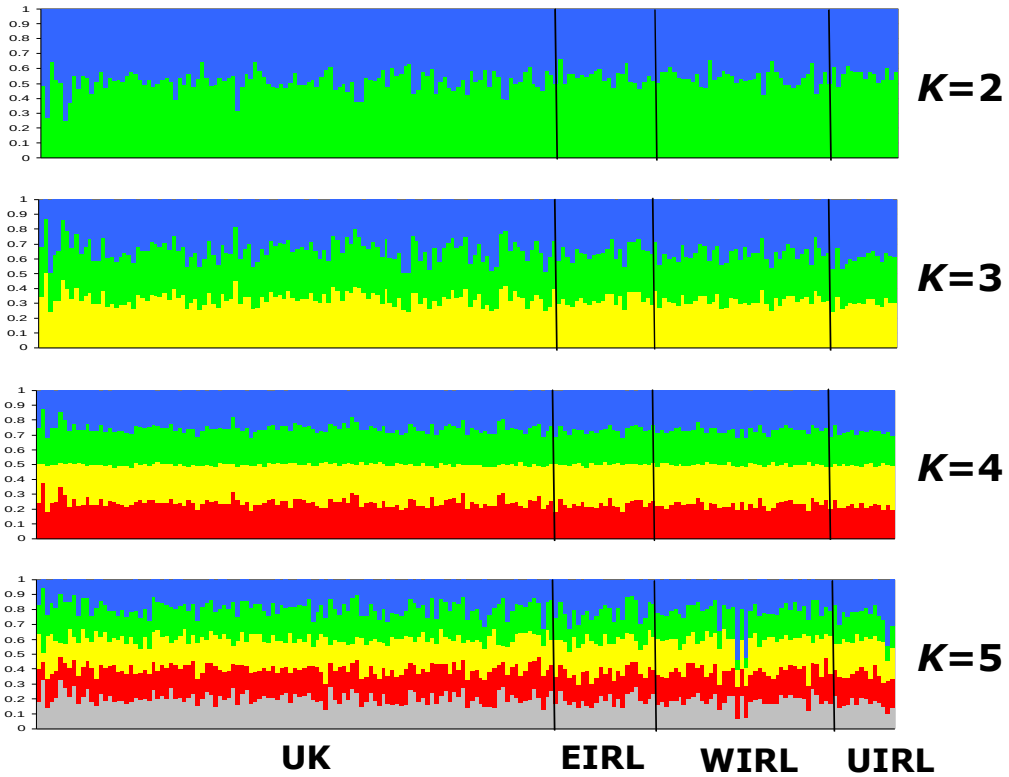


Figure 3.5 *Summary of the **structure** model-based clustering approach. Analysis was based on complete 380 STR loci autosomal genotypes in 194 individuals. Each vertical line represents one individual and is partitioned into K (number of populations) coloured segments representing the fraction of the genome derived from each of K populations. Black lines separate populations: UK = United Kingdom, EIRL= East Ireland, WIRL = West Ireland, UIRL= Unknown Ireland (no information on regional origin).*

3.4 DISCUSSION

3.4.1 Classical Gene Substructure within Ireland

A range of autosomal loci, examined using a variety of methodologies, provided no compelling or robust evidence for large-scale differences between East and West Ireland. Some individual classical genes, notably the ABO blood group system, display (relatively shallow) gradients across the island. Novel simultaneous analysis of multiple classical gene loci in a broad European context suggests that differences within Ireland, while individually real, are collectively slight especially compared to dramatic Y chromosome differentiation (Hill et al. 2000). The detection of the major southeast to northwest pattern, that typically extents into Ireland to separate east and west regions, suggests a lack of regional Irish distinction is not simply a function of a small number of classical genes and/or uninformative loci.

3.4.2 Genome-wide Autosomal STR Variation

As with classical gene loci, little evidence of general substructure emerges from 380 STR autosomal loci despite several different analytical approaches. However, there are a number of potential confounding factors that make inference from this difficult. Firstly, the regional East-West Irish sample populations are far from ideal being relatively small and with a heavy influence of Dublin and Cork respectively. As major urban centres they are likely to encompass a variety of now unknown ancestral locations from across Ireland. Secondly 'United Kingdom' is a poor population label given the heterogeneity of Britain, including regions like Scotland and Wales that have a close affinity with Ireland (see Chapter 2). It is also conceivable that the UK cohort contains individuals from Northern Ireland or British people with recent Irish ancestry. These observations may explain why the UK does not show any closer affinity to the continental Danes than does Ireland.

In additional, a single genome wide assessment, adopted due to the absence of comparative data, may lack the finesse for examining substructure. Although autosomal loci can be closely linked, those used here are widely dispersed across the genome and each may represent a separate realisation of population history. Some will have different primarily influences, for example, through different population movements. However, even under a common gene flow process, many loci and their alleles are expected to be uninformative either because source and destination populations are not sufficiently differentiated in the first instance or as a result of the subsequent distorting effect of genetic drift (Sokal et al. 1989; Ammerman and Cavalli-Sforza 1984).

As neutral markers, STRs offer the advantage of being invisible to selection unlike some classical gene loci, for example, the Duffy blood group (Hamblin and Di Rienzo 2000). However, they have a generally high mutation rate, particularly the dinucleotide repeat markers used here with some displaying hyper-mutability, a feature that may confound the reconstruction of even relatively short-term relationships (Zhivotovsky 2003). The large number of loci that could not distinguish any relative difference between groupings (zero F_{ST} values) questions the overall information content of these markers. Furthermore, combined analysis of a dataset whose variation was moulded by a palimpsest of processes may confound the detection of real trends in individual or small subsets of loci.

3.4.3 Model Based Clustering versus Genetic Distances

The results of the model based clustering method implemented in *structure* should be immune to the criticisms of poorly defined or heterogeneous populations as these are not defined in advance but rather are essentially formed from the genotypic data. Notwithstanding this, *structure* failed to detect any form of substructure in the British Isles, let alone longitudinal Irish variation. However, this finding is also unsurprising, as the method seems to perform relatively poorly when applied to closely related human populations. Within continent divisions are often poorly resolved, particularly in Europe with its relatively low levels of divergence and diversity (Rosenberg et al. 2002). A smaller study using 34 X chromosome STRs was also unable to detect any meaningful structure between four widely dispersed European populations including the frequent polar opposite Basques and Turkish populations (Wilson et al. 2001).

In the context of this and the results of the present study, with its large number of loci, it seems reasonable to conclude that the *structure* approach is not sufficiently powerful to detect subtle differences in small geographic regions like the British Isles. Roughly speaking, the method uses departures from Hardy-Weinberg equilibrium as the basis for defining different populations. Britain and Ireland may not be sufficiently diverged or isolated from each other for this to have occurred. Incidentally, this points out the overall similarity of populations despite a strong emphasis on divergence that often characterises population genetic studies. Despite the sophistication and attraction of the *structure* method, traditional distance based approaches seem superior for investigating the relationship of local human groupings.

3.4.4 Conclusions

While this study did not detect general longitudinal autosomal variation in Ireland, it is less clear what population inference may be draw independent of sample and analytical limitations. It tentatively suggests, unsurprisingly, that genome wide intra-Ireland east-west substructure is certainly not a feature of all or even a large number of loci. Secondly, when any such differences are viewed in a broader European context, they are not as severe as found in the Y chromosome. The contrast with the latter, and agreement with mtDNA (see Chapter 2) adds weight to the idea that Y chromosome differentiation may have been driven (and relatively exaggerated) by male specific factors either before or during Irish population history. A fuller and more meaningful assessment of autosomal regional variation in Ireland, that large numbers of loci could potentially provide, will only be realised in the context of greater European comparative data. A broader background would allow the identification of informative loci that display (intelligible) variation along with a greater understanding of the processes that shaped these patterns; knowledge that could then be applied in the investigation of intra-Ireland diversity.

CHAPTER 4

PATRILINEAL ANCESTRY IN IRISH SURNAMES

4.1 INTRODUCTION

4.1.1 Surnames and Genetics

The hereditary nature of surnames, indicating common and specific ancestry, has a long tradition of exploitation in the investigation of population history. Charles Darwin (1875) used the frequency of marriages between people of the same surname (isonymy) to deduce inbreeding levels. Since then, similar surveys have been carried out throughout the world, while the geographic and ethnic specificity of surnames has also been used to track and measure levels of gene flow between different regions (reviewed by Colantonio et al. 2003).

However**,** rather than using surnames to learn about population genetic diversity, the reverse is made possible by the patrilineal coinheritance of Y chromosomes and surnames in many societies, including Ireland. In theory, if just one man founds a surname all his patrilineal descendents will carry both the name and Y chromosome of their 'founding father'. Alternatively, if established by multiple men the modern Y chromosome pool of that surname will display numerous disparate lineages. In practice, factors such as mutations, adoptions, non-paternity events, and in short, anything that breaks the paternal coinheritance of Y chromosomes and surnames, make historical inference from extant diversity more complicated than these simple scenarios (Jobling 2001).

4.1.2 Y Chromosomes and Surnames

While the patrilineal coinheritance of surname and Y chromosome is not a recent observation, an application in learning about surname history was necessarily limited by an inability to distinguish between Y chromosomes. However, some carry translocations from autosomal chromosomes that are selectively neutral (do not impact

on fitness) and are easily identified cytogenetically. In 1970s, these 'satellited Y chromosomes' were observed in seventeen Canadian men bearing the same surname, providing the first genetic evidence of an underlying paternal ancestry to surnames (Genest 1973). An additional cytogenetically distinct Y chromosome was later found in three out of four bearers of the same surname in Columbia (Giraldo et al. 1981).

The discovery of further Y chromosome polymorphism, particularly fast changing STR markers in the 1990s, finally provided the means of fully exploiting Y chromosome diversity to give finer detail and greater finesse to the reconstruction of paternal relationships. Low resolution investigation of Y chromosomes in the English surname "Sykes", using just four STR loci, revealed that about 44% of 48 surname bearers carried the same Y chromosome haplotype, which was not observed in local or general English population samples (Sykes et al. 2000). These observations were interpreted as a single origin of the name. Other Sykes males, who mostly showed no close relationship to the founding haplotype or to each other, were attributed to cumulative non-paternity events. Notwithstanding this study, Y chromosome investigation of surname adoption and history is under-utilised and largely the preserve of haphazard personal 'genealogical' testing provided by commercial concerns (Brown 2002; Elliot and Brodwin 2002).

The use of surnames as indicators of ethnic and/or geographic origin has, however, found wider use in the exploration of population-level paternal affinities. In Ireland, the Y chromosome complements of men with English, Scottish and 'Gaelic' (indigenous Irish) surnames were found to differ significantly, demonstrating the genetic legacy of recent migrants (Hill 1999; Hill et al. 2000). In addition, the general high frequency of R1b chromosomes in men with Gaelic Irish surnames along with its uneven distribution

in those originating in East and West Ireland supported the conclusion that this type represented the primary Y chromosome legacy of the earliest Irish settlers (see also Chapters 2 and 3). However, in the absence of large numbers of men with the same surname, this study could say little about individual or collective surname origin and history in Ireland.

4.1.3 Early Irish Names

Although paternally inherited surnames are the norm across Europe, Ireland was one of the earliest cultures to adopt this system. On the continent and in Britain their development tends to date from the 12th to 15th centuries and frequently much later (Adams 1979). One of the earliest recorded examples of a hereditary surname in Ireland (O'Clery) comes from 916 AD (Woulfe 1923, Introduction, pg xv-xvi). However, their adoption was only one step in the 1500-year historical record of Irish personal nomenclature.

The earliest contemporary records of Irish names are found on 'Ogham stones', memorial markers dating from the 4th to 7th centuries AD inscribed with a primitive Irish alphabet (Ogham script). These mention a first name followed by the phrase *maqi muccoi*, meaning "son of the race of", and the name of the local population grouping with which the person identified (Ó Crónín 1995, pg 33-36). In the 2nd half of the 1st millennium AD, Ireland was a patchwork of such population groups and local kingdoms, each of varying and shifting power. The later names of these reflect a pre-surname sense of identity defined by ancestor and were typically formed by adding a prefix or suffix to the putative founding ancestor's name. The former include *Cinéal* (kindred of), *Dál* (tribe of) and *Uí* (grandson/descendent of); while the suffix *–acht* meaning 'people of' was also commonly used, for example *Eóganachta* (Eogan's

people) (MacNiocaill 1972, Chapter 2; Woulfe 1923, pg 685-696). However, whether this should be understood as a real biological link between inhabitants of the same local kingdom or simply an expression of the power by the ruling dynasty is uncertain. By at least the 8th century AD, transient surnames had appeared by prefixing *Mac* (Son of) or *Ó* (grandson [or looser male descendent] of) to the relevant ancestor (McLysaght 1985a, pg 15). However, these early surnames were not hereditary, the bearers were literally the son or grandson of the person referred to in their names, a system similar to that currently used in Iceland.

4.1.4 Hereditary Surnames in Ireland

True hereditary indigenous Irish surnames began to form from the 10th century AD. In common with the prior transient system, they emphasised relationship to a (frequently prominent) male ancestor through the prefixes *Mac* and *Ó*. However, once the name formed it passed from generation to generation intact. The reasons behind early surname adoption are uncertain and poorly explored. Increasing population and administrative concerns may have made accurate identification of people more important. Other changes in Irish and European society during the eleventh and twelfth centuries whereby nobility came to be associated with birth into a particular family or class, rather than deriving directly from landownership or profession may also have been responsible (Dr. Katherine Simms, Department of Medieval History, TCD, personal communication). Whatever the reasons, many of major names were formed by the 12th century though the system continued to evolve in later centuries. As surnames grew in frequency they would often split and give rise to new offshoots names either by the addition of a suffix, for example O'Sullivan divided into O'Sullivan Mór and O'Sullivan Beare branches, or more dramatically, by the formation of an otherwise entirely distinct name. However, perhaps the biggest post-foundation influence on Irish

surnames was the English conquest beginning in 1169. Not only did this lead ultimately to the introduction of large numbers of Scottish and English settlers and their names, it also profoundly impacted pre-existing Gaelic Irish surnames. There was both overt and indirect pressure to emulate the conquering culture or 'anglicise' leading to some native Irish dropping their indigenous names for English replacements. The pressure was somewhat relieved by the tendency of the early Anglo-Norman settlers to slip into native Gaelic culture becoming in the words of the historical cliché "more Irish than the Irish themselves". With a degree of irony, this led to a reverse process of 'gaelicisation' of foreign names through the addition of *Mac* and *Ó*. A situation so serious that the 1366 'Statutes of Kilkenny' were passed to remind the Anglo-Norman community of its identity and nomenclature (Adams 1979).

However, from the 16th century forward, the reassertion and extension of direct English control and administration across the island caused a reciprocal decline in native Irish culture and language. Anglicisation gathered pace and impacted virtually every name though the process was far from uniform and manifest in various different ways. The most obvious and general expression was the discarding of the *Mac* and *Ó* prefixes. In addition, some names were 'transliterated' by English speaking administrators into a phonetically similar English spelling, while others were substituted for, often vaguely, similar English names. In yet other instances, direct translation and frequently mistranslation from Irish to English was undertaken. The anglicisation process continued well into the 19th century in the West, which was largely Irish-speaking until the famine. However, by the last century, the situation had come full circle with resurgence nationalism and subsequent independence leading to a 'Gaelic revival'. The prefixes Mc and O were re-adopted enthusiastically and some reverted to putative Irish language versions of their names. Nonetheless, the net legacy of this fluid and fitful

evolution is, at least, a somewhat confused corpus of surnames. Some names that had no initial relationship to each other are now synonymous while other originally identical surnames currently exist in multiple different forms (Woulfe 1923; Adams 1979; McLysaght 1985a).

4.1.5 Irish Genealogical Tradition

With the heavy emphasis on ancestral origins in Irish surname nomenclature, it unsurprising that the recording of genealogical information was a passion of the Irish from at least the early historical period. Ireland has the largest surviving body of early genealogical records tracing the descent of leading Irish families over long periods of time, in some cases based on 6th or 7th century AD sources. Many genealogies continue further back in time to include pre-historic and quasi-mythical ancestors. As their main purpose was validation of claims to power and property, they were often altered and forged to accord with the shifting political circumstances. In addition, the surviving records only represent a fraction of an original collection much of which was lost in the wars of the 17th century and the neglect and poverty of the 18th century (Ó Muraíle 2003, Volume I pg 9-14).

Systematic and genealogically informative records for most of the general population are relatively recent. State registration of birth, deaths and marriages began for most of the population in the 1860s, while parallel church records are hardly of any greater time depth. Although a census had been carried as early as 1821, these and other administrative records (some dating to the 13th century) were destroyed during the Civil War (1922-1923). As a result, the earliest extensive record of surname distribution comes from the 'Primary Valuation of Ireland' conducted by Richard Griffith between 1848-1864 ('Griffith's Valuation') to assess land taxation and which incidentally noted the surname of almost every householder in the country (Grenham 1999).

4.1.6 The McGuinness Surname

McGuinness is a typical example of a Gaelic or indigenous Irish surname. The current version formed from the phonetic anglicisation of the Irish *Mac Aonghusa* meaning "Son of Aongus" (McLysaght 1985a). The eponymous Aongus is reputed to have lived about 950 AD in the area of modern County Down. Around this time and for several preceding centuries, the region was part of the over-kingdom of *Ulaid* (whence Ulster derives its name) which was a patchwork of local dynasties, principally the *Dál Riata*, *Dál nAraide*, *Uí Echach Coba* and *Dál Fiatach*. There are also suggestions that this area was ethnically diverse with a distinction drawn, in some historical writings, between two groups: the *Érainn* (including the *Dál Riata* and *Dál Fiatach*) and the *Cruithni* (*Dál nAraide* and *Uí Echach Coba*). The term *Cruithni* (Cruthin) is the Irish language or Q-Celtic (see also Chapter 2) equivalent of the P-Celtic *Prydyn* meaning 'Britons', which is more familiar as the Latin nickname *Pictti* or Picts (Mac Niocaill 1972, Chapter 2; Ó Cróinín 1995, pg 48-49; Connolly 2001, pg 136-137 and 589). Consequently, an origin outside Ireland in geographically proximate Scotland is sometimes suggested. However, the absence of a separate language, social structure or any archaeological distinctiveness argues this theory (Bardon 1992, pg 14). The progenitor of the McGuinness surname and others from the same area such as McCartan came from the *Uí Echach Coba* dynasty. Indeed the surviving genealogical corpus specifically names a common ancestor for these two names around the 6th century AD (Byrne 1997, pg 287).

The surname is notable today for the variety of spelling versions, the presumed result of haphazard anglicisation, with the two most common varieties being McGuinness and Magenis. The earliest complete geographic distribution of the McGuinness group of phonetic names, from the mid-19th century 'Griffith's Valuation', (**Figure 4.1**) clearly shows a single prominent frequency focus on County Down in agreement with the

historical narrative. The incidence of the name declines steeply from here but shows a moderate up-turn in occurrence in the far West, particularly Donegal and Leitrim. In line with the disturbed history of Irish surnames, several other names including McCreesh, McNeice and Neeson are proposed to represent derivatives or corruptions of the original McGuinness surname (McLysaght 1985b). In addition, the surname 'Guinness' has an obvious similarity and family tradition makes a link to County Down and McGuinness. However, the earliest historical records find the father of the Guinness brewing founder Arthur Guinness (**Figure 4.2**) living in Co. Kildare in the early 18th century (Wilson 1999).

4.1.7 The Donohoe Surname

The Donohoe surname is an example of the other great prefix of Irish surnames, deriving from the addition of *Ó* to the personal name *Donnchadh* to give *Ó Donnchadha*, meaning grandson (or other male descendent) of Donnchadh. As this was a reasonably common personal name, the surname is thought to have arisen multiple times across the country, principally in present day Cavan, South Munster and Galway (McLysaght 1985a). The eponymous ancestor of the name in Cavan lived *ca*. 1100 AD with the first bearer of the surname recorded in 1159 AD in the Annals of Ulster. In Munster, the name appears to be somewhat older originating with an early 11th century King of Munster. The geographic distribution of the name in the mid-19th century is in agreement with these historical accounts, showing multiple discrete frequency foci (**Figure 4.3**). Like McGuinness, there are several different phonetically equivalent spelling versions of the name most notably 'Donohoe' and 'O'Donoghue', which are generally associated with the Cavan and South Munster areas respectively.

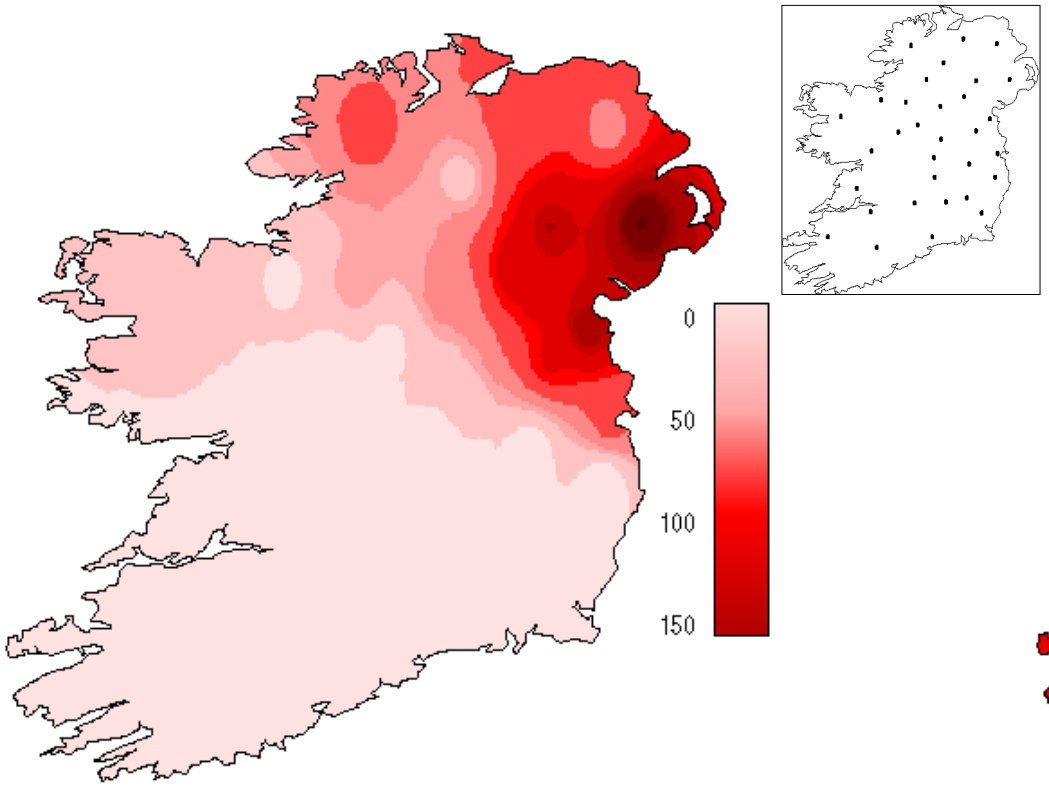


Figure 4.1 *Synthetic surface map of Ireland showing the geographic distribution of the McGuinness surname in the mid-19th century (1465 households including close spelling variants). Even some 900 years after foundation there is still a clear focus of frequency on the historical Co. Down place of origin. Data taken from the Primary Valuation of Ireland conducted by Richard Griffith 1848-1864, accessed online at http://www.ireland.com/ancestor. The contour map was generated from the county level distribution of the surname (32 county coordinates are shown inset) using the Spatial Analyst extension of Arcview (Version 3.2, Environmental Systems Research Institute Inc.).*

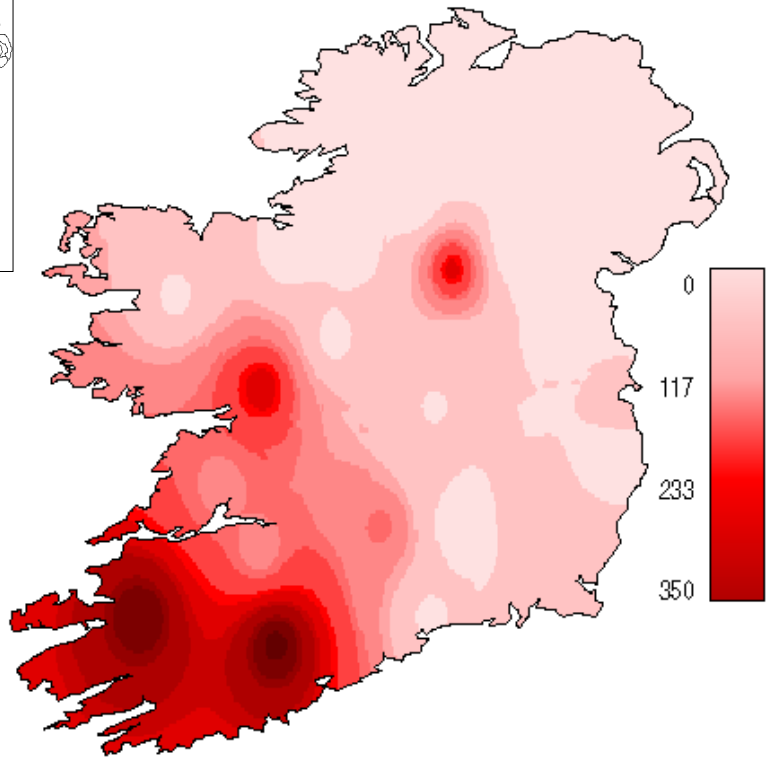


Figure 4.3 *Synthetic surface map of Ireland showing the geographic distribution of the Donohoe surname in the mid-19th century (2769 households including close spelling variants). The multiple geographically distinct frequency foci accord with historical accounts suggesting many independent origins of the surname from the historically common personal name Donnchadh. (Data and Map construction as described for* ***Figure 4.1****)*

Figure 4.2 *Arthur Guinness (1725-1803), who founded the Guinness brewery in 1759. Family history suggests a link to the McGuinness surname of Co. Down. Due to the perturbed nature of Irish surname history many once originally identical names have ended up as different versions. Neeson, McCreesh and McNeice are also thought to represent corruptions of the McGuinness surname. Painting by unknown artist ca. 1770 – 1780 © Patrick Guinness.*

4.1.8 Present Study

Historical sources and indeed the very nature of Irish surname nomenclature suggests that modern surname populations have common patrilineal descent from a single or limited number of founders. However, the considerable time-depth of Irish surname origins coupled with their turbulent history creates uncertainty in the current biological basis, if any, which underlies a surname. Beginning in this Chapter and continuing in Chapter 5, the theoretical coinheritance of Y chromosomes and surnames is exploited to investigate the early origins and subsequent histories of several prominent indigenous Gaelic Irish surnames. Furthermore, the inter-relationship of different names linked by geography and putative history is examined, ending ultimately in a broad assessment of the paternal genetic legacy of Irish surnames. To achieve this, 1158 DNA samples from groups of men bearing the surnames of interest were collected and their Y chromosomes investigated using STR and UEP markers. In this Chapter, the Y chromosome diversity of two names: McGuinness of County Down and Donohoe of County Cavan are separately examined both in detail and in concert with that found in other culturally and geographically matched surnames. Through this, an initial assessment is made of the common patrilineal ancestry underlying Irish surnames and

secondly, the role of the Y chromosome in the reconstruction of surname origin and history.

4.2 MATERIAL AND METHODS

4.2.1 Overview of Sample Collection Strategy

The sample collection strategy is an important consideration in all population genetic studies and this is especially true when examining recent events like Irish surname history. Surname populations are generally small and dispersed presenting a challenge in assembling sufficient sample sizes. Apart from size considerations, the sample group should accurately reflect the surname it purports to represent by excluding known close relatives and respecting the geographic distribution of the name. Discrete geographic peaks in surname occurrence may reflect genetically distinct groups and sampling from one region alone could lead to false conclusions on overall surname origin.

The samples used in this study, described here and in Chapter 5, were collected over a two and a half year period between Spring 2001 and Autumn 2003. Most samples were sourced through postal requests for participation to surname bearers identified from telephone directories. Attempts were made, in so far as possible, to reflect the geographic distribution of the name, using the earliest available information contained in the mid-19th century 'Griffith's Valuation' (**Section 4.1.5**), accessed online at http://www.ireland.com/ancestor. Using this approach it was not possible to formally exclude the participation of close relatives but only one request was sent to smaller villages and townlands to minimise this possibility. Each postal request contained a letter explaining the research, as well as a Cytosoft ^{TM} nylon cytology brush (Medical Packaging Corporation) and instructions for the self-collection of buccal check cells. Volunteers returned brushes by prepaid post, along with a completed questionnaire giving their surname and county of origin, as well as a place of origin for their father and paternal grandfather. Depending on the age of the participant, the latter gives a location for the Y chromosome up to 100 years ago, which is more likely to hold continuity with

ancestral origins than a current place of residence. A minority of samples were collected *via* other channels including personal contacts (especially for common names) and the participation of staff and students in Trinity College, Dublin. In a small number of cases the distribution of participation requests by surname associations was a valuable source of samples for specific names. All sampling was carried out in full accordance with the principle of informed consent.

4.2.2 McGuinness and East Ulster Sample Collection

The sample collection for the McGuinness name is outline in detail in **Table 4.1**. 99 usable DNA samples were obtained from men belonging to the McGuinness phonetic group of surnames representing twelve spelling variants. The most common types in the general population, McGuinness and Magenis, were also the most frequent in the sample population at 64 and 16 respectively. The surname sample also encompassed the geographic distribution of the surname in mid-19th century (**Figure 4.1**) In line with this, most came from County Down or surrounding counties but representatives from West Ulster and other parts of the island were also included. In addition, samples from four putative related names were collected (*sample size*): McNeice (*4*), Neeson (*8*), Guinness (*3*) and McCreesh (*7*). To place these results in a suitable local Y chromosome context, a further 194 individuals from 12 surnames with generally similar geographic and cultural (Gaelic) origins were also assembled as follows (*sample size*): Coulter (*11*), Dunleavy (*12*), Haughey (*19*), Heaney (*17*), McAvoy/McEvoy (*26*), McCartan (*13*), McGinn (*19*), McVeigh (*10*), Murphy (*13*), O'Hanlon (*13*), O'Hare (*19*) and Rooney (*22*). In addition, samples from two predominantly West Ulster names, McGinley (*17*) and McGinty (*12*) were collected to investigate the origin of McGuinness in this area.

NAME	SAMPLE COLLECTION					SAMPLE QUALITY			
	Samples Collected	Samples from Postal Requests	Postal Request Male Response Rate (%)	Samples from Other Sources[1]	NoDNA	Female DNA	Usable DNA	Usable DNA (%)	
MCGUINNESS SURNAMES:									
M(A)CINNES	2	1	4	25	1	0	0	**2**	100
MACAONG(H)USA	3	3	5	60	0	1	0	**2**	66.7
MAGENIS	1	1	3	33.3	0	0	0	**1**	100
MAGENNIS	16	15	36	41.7	1	0	0	**16**	100
MAGINESS	5	5	12	41.7	0	0	0	**5**	100
MAGINNES	1	1	16	6.3	0	0	0	**1**	100
MAGINNESS	1	1	11	9.1	0	0	0	**1**	100
MAGINNIS	2	2	10	20	0	0	0	**2**	100
MAGUINESS	0	0	1	0	0	0	0	**0**	NA
MAGUINNESS	1	1	11	9.1	0	0	0	**1**	100
MCGENNIS	2	2	4	50	0	0	0	**2**	100
MCGINNESS	0	0	1	0	0	0	0	**0**	NA
MCGINNIS	2	2	12	16.7	0	0	0	**2**	100
MCGUINESS	0	0	9	0	0	0	0	**0**	NA
MCGUINNESS	68	65	199	32.7	3	4	0	**64**	94.1
MCGUINNIS	0	0	1	0	0	0	0	**0**	NA
TOTAL	**104**	**99**	**335**	**29.6**	**5**	**5**	**0**	**99**	**95.2**
PUTATIVE MCGUINNESS DERIVED SURNAMES:									
GUINNESS[2]	4	NA	NA	NA	4	1	0	**3**	75
M(A)CNEICE	4	4	21	19	0	0	0	**4**	100
NEESON	9	9	39	23.1	0	1	0	**8**	88.9
MCCREESH	7	7	30	23.3	0	0	0	**7**	100
TOTAL	**24**	**20**	**90**	**22.2**	**4**	**2**	**0**	**22**	**91.7**
EAST ULSTER SURNAMES:									
COULTER	13	13	84	15.5	0	1	1	**11**	84.6
HEANEY	19	18	79	22.8	1	2	0	**17**	89.5
HAUGHEY	19	19	66	28.8	0	0	0	**19**	100
O'HARE	20	20	82	24.4	0	1	0	**19**	95
MCVEIGH	12	12	56	21.4	0	2	0	**10**	83.3
MCEVOY/MCAVOY[3]	26	23	73	31.5	3	3	0	**26**	88.5
ROONEY	22	21	82	25.6	1	0	0	**22**	100
O'HANLON	15	15	41	36.6	0	2	0	**13**	86.7
DUNLEAVY	12	12	32	37.5	0	0	0	**12**	100
MCCARTAN	13	13	31	41.9	0	0	0	**13**	100
MCGINN	22	22	48	45.8	0	3	0	**19**	86.4
MURPHY[4]	NA	NA	NA	NA	13	NA	NA	**13**	NA
TOTAL	**193**	**188**	**674**	**27.9**	**18**	**14**	**1**	**194**	**92.8**
WEST ULSTER SURNAMES:									
MCGINLEY	17	17	43	39.5	0	1	1	**15**	88.2
MCGINTY	12	12	22	54.5	0	0	0	**12**	100
TOTAL	**29**	**29**	**65**	**44.6**	**0**	**1**	**1**	**27**	**93.1**
PROJECT TOTAL[5]	**350**	**336**	**1164**	**28.9**	**27**	**22**	**2**	**342**	**93.4**

[1] Mainly through personal contacts.

[2] The assitance of Patrick Guinness in recruiting volunteers of this name is acknowledged.

[3] Final total includes 3 McEvoy/McAvoys collected during an earlier sampling of the McEvoy name in the Midlands (see Chapter 5).

[4] Ulster Murphy samples drawn from a general Murphy sample (see Chapter 5).

[5] Project Total response rate excludes the McEvoy and Murphy samples described in notes 3 and 4

Table 4.1 *Summary of sample collection for McGuinness and other geographically (general East Ulster area) and culturally (Gaelic) matched surnames. In addition to numerous spelling variants, samples from putative McGuinness derivatives were also collected along with samples from two West Ulster surnames to investigate the origin of the McGuinness surname in this region.*

4.2.3 Donohoe and Cavan Area Sample Collection

As with McGuinness, the Donohoe surname has a number of very similar phonetic spelling variants. A total of 73 usable DNA samples were collected from men with four versions (*sample size*): Donohoe (*36*), O'Donoghue (*24*), Donohue (*7*), and Donoghue (*6*) (**Table 4.2**). The sample population also aimed to reflect variation in the geographic distribution of the name, so as to allow inter-region comparisons, and thus included samples from Munster as well as from Cavan and surrounding counties. The local genetic context for the Cavan area was provided by 62 samples spread across 10 surnames as follows (*sample size*): Maguire (*8*), Farrelly (*5*), O'Connor (*9*), McCabe (*6*), McMahon (*5*), McGovern (*7*), O'Rourke (*5*), Brady (*3*), O'Reilly (*8*) and McKiernan (*6*).

	SAMPLE COLLECTION					SAMPLE QUALITY			
NAME	Samples Collected	Samples from Postal Requests	Postal Request Made	Response Rate (%)	Samples from Other Sources[1]	NoDNA	Female DNA	Usable DNA	Usable DNA (%)
DONOHOE SURNAMES:									
DONOGHUE	6	6	22	27.3	0	0	0	6	100
DONOHOE	37	17	49	34.7	20	1	0	36	97.3
DONOHUE	8	6	15	40	2	1	0	7	87.5
O'DONOGHUE	25	22	60	36.7	3	1	0	24	96
O'DONOHUE	0	0	2	0	0	0	0	0	NA
TOTAL	**76**	**51**	**148**	**34.5**	**25**	**3**	**0**	**73**	**96.1**
CAVAN AREA SURNAMES:									
MAGUIRE	*8*	8	15	53.3	0	0	0	**8**	100
FARRELLY	*5*	5	28	17.9	0	0	0	**5**	100
O'CONNOR	*9*	9	30	30	0	0	0	**9**	100
MCCABE	*6*	6	15	40	0	0	0	**6**	100
MCMAHON	*5*	5	15	33.3	0	0	0	**5**	100
MCGOVERN	*8*	8	15	53.3	0	1	0	**7**	87.5
O'ROURKE	*5*	5	15	33.3	0	0	0	**5**	100
BRADY	*3*	2	16	12.5	1	0	0	**3**	100
O'REILLY	*9*	7	15	46.7	2	1	0	**8**	88.9
MCKIERNAN	*6*	6	15	40	0	0	0	**6**	100
TOTAL	***64***	**61**	**179**	**34.1**	**3**	**2**	**0**	**62**	**96.9**
PROJECT TOTAL	**140**	**112**	**327**	**34.3**	**28**	**5**	**0**	**135**	**96.4**

[1] The assitance of Joseph A. Donohoe in collecting additional samples from the Donohoe group of names is acknowledged.

Table 4.2 *Summary of sample collection for Donohoe (including spelling variants) and other geographically and culturally matched Cavan area surnames.*

4.2.4 Estimation of Surname Frequency

In the absence of any alternative figures, the frequency of each surname was calculated using an *ad hoc* approach based on information from telephone directories. The number of households with each surname was estimated by direct and approximate count from phone directories encompassing all areas of Ireland. 83% of residences in the Republic of Ireland have a fixed phone-line and 70% of these are included in a telephone directory (Source: Eircom P.L.C). The average household in Ireland contains slightly fewer than 3 people (Source: Census of Ireland 2002). Conflating these statistics (multiplying 0.83 by 0.70 and diving the result by 3), and assuming similar figures for Northern Ireland, indicates that the number of individuals mentioned in the phone directory (each representing one residence) is 19.4% of the total population. Surname frequency estimates based on this observation are for Ireland only; the total number of surname bearers worldwide is much greater given high historic (and indeed recent) Irish emigration, particularly to North America, Britain and Australia.

4.2.5 Y Chromosome Short Tandem Repeats

DNA was extracted from buccal cheek cells essentially as described in **Section 2.2.2**. The relatively recent timeframe of Irish surnames means that short tandem repeats or STR (microsatellite) markers with their high mutation rates are the most appropriate and, therefore, principal marker type used in this study. The absence of recombination ensures that multiple Y chromosome STR loci together produce highly informative haplotypes that can be used to infer fine relationships between individual Y chromosomes.

19 Y chromosome specific STR loci were genotyped for each sample. These were amplified in three polymerase chain reaction (PCR) multiplexes and the number of

repeats at each marker (size difference) detected through gel electrophoresis on ABI Prism^{TM} 377 DNA sequencer (Applied Biosystems/Perkin-Elmer) using fluorescently labelled primers. Two multiplexes 'MS1' (Thomas et al. 1999) and 'CTS' (Ayub et al. 2000) included six markers while the third, 'EBF' (Bosch et al. 2002), contained 7 STR loci. The latter included DYS385, which gives two homologous products (A and B) simultaneously that cannot be differentiated using these primers. EBF also contains the complex DYS389 locus, which encompasses DYS389I and DYS389II in a single amplicon. The DYS389II (AB fragment) repeat number was determined by subtracting the DYS389I (CD fragment) score from the former (Rolf et al. 1998). The Amelogenin gene sex test was also included in this multiplex. This allowed accidental female participation, during unsupervised sample collection, to be distinguished from failed Y chromosome amplification due to poor quality DNA.

Details of the STR markers, primers sequences, their fluorescent modifications and PCR cycling conditions are given in **Table 4.3**. Primers sequences are as described by Bosch et al. (2002) but primer concentrations were substantially modified, for many loci, to generate a more robust PCR. Each multiplex amplification was carried out in a 10µl volume containing 0.4mM each of dATP, dCTP, dGTP, dTTP, 10mM Tris-HCl, 50mM KCl, 0.1% (V/V) Triton-X-100, 0.01% (W/V) Gelatin, 3.0 mM $MgCl_2$, 1 unit of *Taq* polymerase and an average of approximately 50 ng of template DNA. 1µl of PCR product was mixed with 2.0 µl of loading dye (50mg/ml Blue dextran, 25mM EDTA, 20% Formamide (V/V)) and 0.3µl of Genescan-ROX 500^{TM} (Applied Biosystems) fluorescent internal size standard. 1.2µl of this mixture was then electrophoresed through a 6% polyacrylamide gel for 3-4 hours on an ABI Prism^{TM} 377 DNA sequencer. DNA fragment size was determined using GeneScan^{TM} Analysis Software (Version 3.1.2, Applied Biosystems).

MULTIPLEX		PCR CYCLING	LOCUS	REPEAT SIZE	REPEAT UNIT	FINAL PCR [PRIMER] (μM)	5' FLUORESCENT DYE	PRIMERS (5' TO 3')
MS1	*1*	95°C for 2 mins	DYS19	4	TAGA	0.35	TAMRA	ACT GAG TTT ATG TTA TAG TGT TTT T ATG GCA TGT AGT GAG GAC A
	2	94°C for 1 min						
	3	57.5°C for 1 min -0.5°C each cycle	DYS388	3	ATT	0.35	TAMRA	GTG AGT TAG CCG TTT AGC GA CAG ATC GCA ACC ACT GCG
	4	72°C for 1 min						
	5	Go to Step 2 x 5 times						
	6	94°C for 1 min	DYS390	4	TCTG/A	0.35	6'FAM	TAT ATT TTA CAC ATT TTT GGG CC TGA CAG TAA AAT GAA CAC ATT GC
	7	55°C for 1 min						
	8	60°C for 1 min						
	9	Go to step#6 x 32 times	DYS391	4	TCTA	0.12	6'FAM	CTA TTC ATT CAA TCA TAC ACC CAT AT ACA TAG CCA AAT ATC TCC TGG G
	10	72°C for 10 mins						
			DYS392	3	TAT	0.425	HEX	AAA AGC CAA GAA GGA AAA CAA A CAG TCA AAG TGG AAA GTA GTC TGG
			DYS393	4	AGAT	0.1	HEX	GTG GTC TTC TAC TTG TGT CAA TAC AAC TCA AGT CCA AAA AAT GAG G
CTS	*1*	94°C for 2 mins	DYS434	4	CTAT/ TAAT	0.2	TAMRA	CAC TCC CTG AGT GCT GGA TT GGA GAT GAA TGA ATG GAT GGA
	2	94°C for 1 min						
	3	60°C for 1 min -0.5°C each cycle	DYS435	4	TGGA	0.05	TAMRA	AGC ATC TCC ACA CAG CAC AC TTC TCT CTC CCC CTC CTC TC
	4	72°C for 1 min						
	5	Go to Step 2 x 8 times						
	6	94°C for 1 min	DYS436	3	GTT	0.065	6'FAM	CCA GGA GAG CAC ACA CAA AA GCA ATC CAA CTT CAG CCA AT
	7	56°C for 1 min						
	8	72°C for 1 min						
	9	Go to Step#6 x 30 times	DYS437	4	TCTA/G	0.065	HEX	GAC TAT GGG CGT GAG TGC AT AGA CCC TGT CAT TCA CAG ATG A
	10	72°C for 5 mins						
			DYS438	5	TTTTC/A	0.175	HEX	TGG GGA ATA GTT GAA CGG TAA GTG GCA GAC GCC TAT AAT CC
			DYS439	4	GATA	0.075	TAMRA	TCC TGA ATG GTA CTT CCT AGG TTT GCC TGG CTT GGA ATT CTT TT
EBF	*1*	95°C for 5 mins	DYS385AB	4	GAAA	0.3	HEX	AGC ATG GGT GAC AGA GCT A CCA ATT ACA TAG TCC TCC TTT C
	2	94°C for 1 min						
	3	60°C for 1 min -0.5°C each cycle	DYS389I/II	4	TCTG/A	0.2	6'FAM	CCA ACT CTC ATC TGT ATT ATC TAT GT CCT GAG TAG CAG AAG AAT GTC ATA
	4	72°C for 1 min						
	5	Go to 2 x 10 times						
	6	94°C for 1 min	DYS460	4	ATAG	0.2	TAMRA	GCC AAA CTC TTT CCA AGA AG TCA TCT ATC CTC TGC CTA TCA TT
	7	55°C for 1 min						
	8	72°C for 1 min						
	9	Go to Step#6 x 28 times	DYS461	4	T/CAGA	0.15	6'FAM	AGG CAG AGG ATA GAT GAT ATG GAT TTC AGG TAA ATC TGT CCA GTA GTG A
	10	72°C for 5 mins						
			DYS462	4	TATG	0.3	HEX	TGT GCT GTA CCA GTT GCC TA CCA GCC TGA GCA AGA GAG TA
			AMEL	NA	NA	0.075	HEX	CCC TGG GCT CTG TAA AGA ATA GTG ATC AGA GCT TAA ACT GGG AAG CTG

Table 4.3 *PCR amplification of 19 Y chromosome short tandem repeat (STR) markers. These were amplified in 3 separate PCR multiplexes (MS1, CTS, and EBF) using primers sequences (Bosch et al. 2002) and modified concentrations described above. Differences in repeat number/size were resolved by polyacrylamide gel electrophoresis using fluorescently labelled primers on an ABI Prism™ 377 DNA sequencer (Applied Biosystems/Perkin-Elmer).*

4.2.6 Y Chromosome Unique Event Polymorphisms

Y chromosome variation was also investigated through unique event polymorphisms (UEP), typically single nucleotide polymorphisms (SNP). These have substantially slower mutation rates than STR markers and consequently allow Y chromosomes to be placed into broader haplogroups (Y Chromosome Consortium 2002; Jobling and Tyler-Smith 2003). These often display good geographical structure in Europe, allowing wider inference on Y chromosomal origins. Each Y chromosome was typed in a sequential and hierarchical manner, illustrated in **Figure 4.4**, guided by prior knowledge on the frequency of haplogroups in Ireland (Hill et al. 2000). The M269 SNP that defines haplogroup R1b3 (closely corresponding to Haplogroup 1 in the nomenclature of Hill et al. 2000) was typed for all samples initially. Any samples not in R1b3 were typed for the remaining markers in series leading to haplogroup designations of IxI1b2, R1a and E3b (Y Chromosome Consortium 2002; Jobling and Tyler-Smith 2003). Four samples out of 1105 successfully genotyped here and in Chapter 5 fell outside these groups and remained unclassified (U/C)

Each marker was amplified using the same basic PCR conditions described for STR markers (previous section), with specific details on primers and $MgCl_2$ concentrations summarised in **Table 4.4**. The state of each marker except YAP was subsequently determined using restriction fragment length polymorphism (RFLP) analysis. Appropriate restriction enzymes (RE) and digestion conditions are also given in **Table 4.4**. For the M269, SRY1532 and M35 SNPs, RFLP patterns were resolved using a 2% agarose gel, containing Ethidium Bromide and visualised under UV light. The YAP polymorphism did not require RFLP analysis as the presence or absence of the *Alu* insertion element is easily determined using agarose gel resolution. However, the smaller size differences for markers M170 and M26 necessitated the higher discriminatory power of an 8% polyacrylamide gel.

UEP	PRIMERS (5' to 3')	PCR AMPLFICATION				PCR PRODUCT SIZE (bp)	RFLP GENOTYPING					
		PCR[PRIMER] (µM)	PCR[$MgCl_2$] (µM)	PCR CYCLING			RESTRICTION ENZYME(RE)	RE DIGEST CONDITIONS	DIGEST TEMPERATURE (°C)	DIGESTION TIME (HOURS)	UEP STATE (RFLP PATTERN bp)	METHOD REFERENCE
M269	Forward: CAT GCC TAG CCT CAT TCC TC Reverse: GGC TGG ACG TAG TCT TGC TC	0.3	2.5	1 94°C for 5 mins 2 94°C for 30 secs 3 60°C for 30sec 4 72°C for 30 secs 5 Go to Step#2 x 35 times 6 72 °Cfor 5 mins		416	BstN I	0.2 U/µl RE 50mM NaCl 10mM Tris-HCl 10 mM $MgCl_2$ 1 mM DTT 0.1 µg/µl BSA	60	2	Ancestral: T (247 + 169) Derived: C (94 + 153 + 169)	Moore 2004
M170	Forward: CTA TTT TAT TTA CTT AAA AAT CAT TGA TC Reverse: AGA CCA CAC AAA AAC AGG TC	0.4	3	1 94°C for 2mins 2 94°C for 30 3 50°C for 30 secs 4 72°C for 30 secs 5 Go to Step#2 x 40 times 6 72°C for 5 mins		88	Bcl I	0.2 U/µl RE 100mM NaCl 50mM Tris-HCl 10 mM $MgCl_2$ 1 mM DTT	50	3	Ancestral: A (63 + 25) Derived: C (88)	Flores et al. 2003
M26	Forward: TTT TTC TGA ATT AGA ATG ATC Reverse: GTA CAC CTT TCT TAG GTT GC	0.4	3	1 94°C for 2mins 2 94°C for 30 3 48°C for 30 secs 4 72°C for 30 secs 5 Go to Step#2 x 35 times 6 72°C for 5 mins		239	Bcl I	0.2 U/µl RE 100mM NaCl 50mM Tris-HCl 10 mM $MgCl_2$ 1 mM DTT	50	3	Ancestral: G (239) Derived: A (222/17)	Flores et al. 2003
SRY1532	Forward: TCC TTA GCA ACC ATT AAT CTG G Reverse: AAA TAG CAA ATG ACA CAA GGC	0.48	2.5	1 94°C for 2mins 2 94°C for 30 3 58°C for 30 secs 4 72°C for 30 secs 5 Go to Step#2 x 34 times 6 72°C for 5 mins		167	Dra III	0.15 U/µl RE 100mM NaCl 50mM Tris-HCl 10 mM $MgCl_2$ 1 mM DTT 0.1 µg/µl BSA	37	4	Ancestral*: G (112 + 55) Derived*: A (167)	Hill 1999
YAP	Forward: CAG GGG AAG ATA AAG AAA TA Reverse: ACT GCT AAA AGG GGA TGG AT	0.4	2.5	1 94°C for 2mins 2 94°C for 30 3 54°C for 30 secs 4 72°C for 30 secs 5 Go to Step#2 x 37 times 6 72°C for 5 mins		455 or 150	NA	NA	NA	NA	Ancestral: YAP- (455) Derived: YAP +(150)	Hill 1999
M35	Forward: TAA GCC TAA AGA GCA GTC AGA G Reverse: AGA GGG AGC AAT GAG GAC A	0.4	2.5	1 94°C for 2mins 2 94°C for 30 3 65°C for 30 secs 4 72°C for 30 secs 5 Go to Step#2 x 33 times 6 72°C for 5 mins		513	Bsr I	0.25 U/µl RE 100mM NaCl 50mM Tris-HCl 10 mM $MgCl_2$ 1 mM DTT	65	3	Ancestral: G (345 + 168) Derived: C (513)	Moore 2004

*SRY-1532 G to A recurrence defines haplogroup R1a. Therefore, the ultimate ancestral state of this marker is also A.

Table 4.4 *PCR amplification and RFLP genotyping of Y chromosome UEPs. The six markers divide Y chromosomes into broader haplogroups as shown in* ***Figure 4.4***

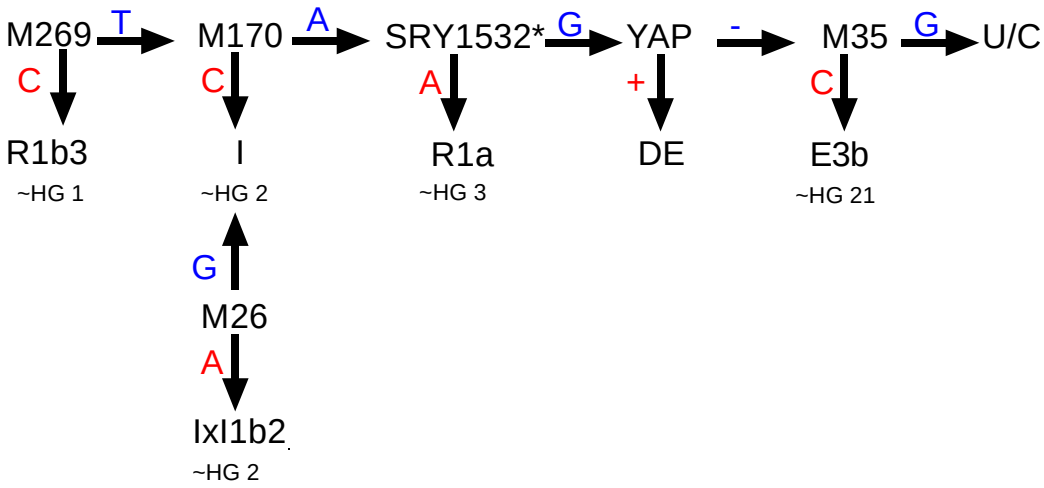


Figure 4.4 *Y chromosome unique event polymorphisms (UEPs) and haplogroups. Each sample was typed in a sequential and hierarchical manner for up to 6 markers (M269, M170, M26, SRY1532, YAP, M35). The ancestral and derived character of each marker is shown in blue and red respectively. The haplogroup defined by each derived state (Y Chromosome Consortium 2002) is shown with the approximate equivalent in the nomenclature used by Hill et al. (2000). Samples not falling into any of these haplogroups remained unclassified (U/C). *Note: SRY-1532 G to A recurrence defines haplogroup R1a. Therefore, the ultimate ancestral state of this marker is also A.*

4.2.7 Median-Joining Networks

The reconstruction and visualisation of relationships between Y chromosomes was carried out through median-joining (MJ) networks (Bandelt et al. 1995; Bandelt et al. 1999). Unlike traditional phylogenetic trees, which show one possible relationship, the network can summarise uncertainty within a single structure through reticulations or multiple pathways between haplotypes (effectively showing multiple trees). However, large numbers of individuals and/or markers can lead to increased uncertainty in true relationships manifesting in the formation of multiple hyper-dimensional cubes that can make networks unintelligible. The MJ method combats this problem by introducing a limited number of 'median vectors' in line with maximum parsimony principles. These are hypothetical intermediate types (nodes) not observed in the real data but whose inclusion can facilitate and ease the construction of phylogenetic pathways between all individuals. In biological terms these represent extant but unsampled or extinct haplotypes. MJ networks were constructed using the program *Network* (Version 4.1) (http://www.fluxus-engineering.com/sharenet.htm).

4.2.8 Y Chromosome STR Weighting

To further reduce reticulations in the MJ networks, a weighting system was adopted to reflect the mutability and thus phylogenetic information content of different loci. Fast changing loci are prone to back mutations and parallelisms that can confound relationship reconstruction. In addition, the method of construction implies a simple stepwise mutation model (SMM) of STR evolution (Kimura and Ohta 1978; Goldstein et al. 1995a, Shriver 1993; Shriver 1995). Loci that exhibit departures from this (multi-step movements) risk confusing the phylogeny. Direct mutation rates are not available for most of the loci used here but variance in the STR repeat length (V) is related to mutation rate (μ) and coalescence time (t) as $V=\mu t$ (Di Rienzo 1994). The V of each locus was calculated over 984 Y chromosomes (described here and in Chapter 5) in the discrete R1b3 haplogroup as $V= n\Sigma x^2 -(\Sigma x)^2/n(n-1)$ where x is repeat number and n is the sample size. Since t is constant between all loci, V provides a proxy and relative indication of mutability.

On this basis seven categories were arbitrarily defined and are shown in **Figure 4.5A**. Differences were interpreted conservatively with each class weighted by successive single integral increments. The most stable markers (least mutable) were given a weighting of 9 (DYS462, DYS436, DYS434), followed by 8, (DYS388, DYS393 DYS438, DYS435, DYS437, DYS19), 7 (DYS461), 6 (DYS389I, DYS389AB, DYS392) and 5 (DYS460, DYS391). The high mutability of DYS390 and DYS439 were reflected in greater increments and a weighting of 3 and 1 respectively. In a validation of this approach, variance was found to be strongly and significantly correlated (r=0.878 p=0.0042) with mutation rates in a subset of the loci, calculated from the data presented in Dupuy et al. (2004) **(Figure 4.5B)**. The marker DYS385 produces two homologous amplicons that cannot be unambiguously differentiated and were therefore not used in network construction or further analysis.

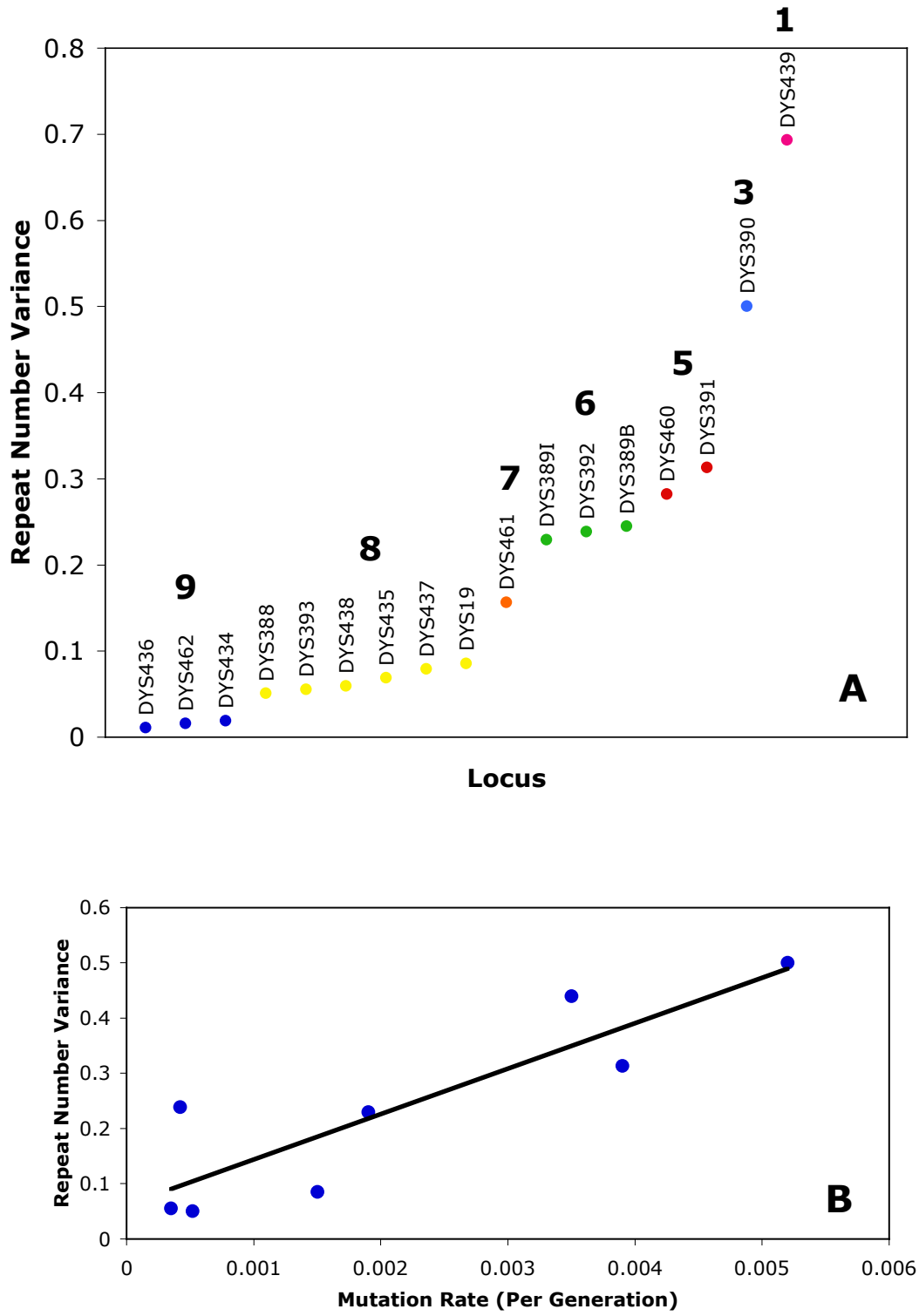


Figure 4.5 *Y chromosome STR loci variance. (**A**) The variance in repeat number distribution amongst 984 haplogroup R1b3 Y chromosomes for each locus was used to define a 7 category weighting system, that reflects locus mutability, for use in MJ network construction. Colours indicate different categories with the relevant weighting indicated above these. (**B**) The significant positive correlation (r=0.878; p=0.0042) between mutation rate and variance in a subset of 8 loci (DYS19, 388, 390, 391, 392, 392 389I, 389II) validates this approach (mutation rates from data in Dupuy et al. 2004).*

4.2.9 Y Chromosome Cluster Definition

A single or limited number of surname founders should result in the elevated frequency of their haplotype(s) in the modern surname population. However, in practice the high mutation rate of STR loci and long history of Irish surnames means even a single lineage will have diversified into a cluster of related types. These are apparent in networks as a frequent and phylogenetically central (ancestral) haplotype orbited by less frequent closely related types. The extent of a single cluster is difficult to define; however, an *ad hoc* approach, based on expectations of mutation as a Poisson distributed process, is used here to estimate these boundaries. The Poisson distribution describes the expected number of discrete events (x) each with a small probability (p) of occurring over (n) trials, which together give the expected number of observed events λ (where $\lambda = np$) as follows:

$$P(x)=\frac{e^{-\lambda}\lambda^{x}}{x!} \qquad \textbf{(Equation 4.1)}$$

As such it approximates the distribution of mutations over compound STR Y chromosome haplotypes under the SMM. From **Equation 4.1**, the probability of observing 0 or 1 mutational steps is given as $P(0)=e^{-\lambda}$ and $P(1)= e^{-\lambda}\lambda$ respectively; dividing P(1) by P(0) gives the parameter λ. Clusters were initially defined as frequent repeated (ancestral) haplotypes with all one step neighbouring haplotypes assumed to represent subsidiary diversity. The frequencies of these gives the first two elements (0 and 1 step events) in the distribution and thus λ, which can then be applied in **Equation 4.1** to calculate the maximum number of individuals expected to be exactly two mutational steps from the ancestral type. However, in an additional level of stringency, only two-step neighbours that could be directly connected to the ancestral haplotype by a filled pathway in which no node was of greater frequency than that preceding it, were

included, up to the maximum number derived from λ. While an approximation, it nonetheless allows clusters to be defined in objective and consistent manner.

4.2.10 TMRCA of Y Chromosome Lineages

The time to the most recent common ancestor (TMRCA) of each significant lineage, as identified above, was calculated based on expectations under the simple stepwise mutation model (SMM) of STR evolution. Under this scenario, the average square distance (ASD_0) in repeat size between an ancestral STR haplotype and all current Y chromosomes is linearly related to the TMRCA (t) and mutation rate (μ) as $ASD_0 = \mu t$ (Goldstein et al. 1995a; Goldstein et al. 1995b; Slatkin et al. 1995; Thomas et al. 1998). TMRCA estimates are clearly dependent on accurate specification of mutation rate and there is a large disparity between rates based on direct observations or pedigree studies (for example, Dupuy et al. 2004) and those calculated from accumulated diversity over an 'evolutionary' (hundreds of years) timeframe (Zhivotovsky et al. 2004). The latter and most current estimate is primarily relied on in this study since it is calculated over time-scales similar to Irish surname history. This rate is 2.76×10^{-5} per locus per year with heuristic 95% confidence intervals (CI), based on inter-locus rate variation, defined by the point estimates of 1.72×10^{-5} and 3.8×10^{-5} per locus per year. The alternative 'direct observation rate' of 6.57×10^{-5} per locus per year (heuristic 95% CI between 3.14×10^{-5} and 1.0×10^{-4} per locus per year) was also explored. These values were calculated from the novel data and compiled published results presented in Dupuy et al. (2004) and assume a generation time of 35 years (Tremblay and Vezina 2000). Although the two estimates differ 2.5 fold they are comfortably contained within each other's 95% CIs. TMRCA and confidence intervals calculations, using the above point estimates, were carried out using purpose-written PERL language software.

A second source of error in TMRCA estimates comes from sampling variance and while any point estimate by the ASD_0 method is independent or unbiased by the demographic history of the cluster, confidence intervals for sampling variance are not (Stumpf and Goldstein 2001). These were estimated, separately to mutation rate uncertainty, using a Monte Carlo simulation approach under the simplifying assumption of a rapidly expanding population ('Star phylogeny') as implemented in the program *Ytime* developed and described by Behar et al. (2003) (http://www.ucl.ac.uk/tcga/software/).

4.2.11 Congruence of Surnames and Y Chromosomes

The extent to which a surname label corresponds to real Y chromosome genetic division and thus common patrilineal ancestry was investigated using three different methods. The elevated frequency of distinct Y chromosome haplotypes in different surnames is the simplest legacy of surname foundation by a single or limited number of men. The expectation of greater haplotype sharing within surnames was assessed seperately for the County Down and Cavan surname populations (described in **Sections 4.2.2** and **4.2.3**). For each individual, the frequency of their haplotypes is calculated within their own surname and simultaneously in a control population of the same size constructed by drawing Y chromosomes at random from different surnames originating in the same area. The process is repeated for each individual and the values within and outside surnames averaged over all samples. Any differences give a rough indication of the 'extra relatedness' both within individual surnames and collectively over all surnames. The entire process was iterated 1000 times to account for the sampling effect involved in the control population construction and the median values reported. Calculations were executed using PERL language software.

The Analysis of Molecular Variance (AMOVA) framework (Excoffier et al. 1992) is a more sophisticated approach as it not only uses haplotype frequencies (as above) but also incorporates information on their molecular divergence, through the sum of the squared (repeat unit) size difference between each haplotype (R_{ST} values). The method assesses surname differentiation by apportioning the total variance found in the sample universe into that found within and between surnames. Variance components were calculated using ARLEQUIN Version 2.000 (Schneider et al. 2000). Significance was gauged by randomly permuting the samples across surname populations 10,000 times to generate a distribution of values under the null hypothesis.

The previous two methods implicitly account for the effect of geographic Y chromosome substructure since all surnames are generally matched for place of origin. However, the partial Mantel Test (Mantel 1967; Smouse 1986) allows for the objective assessment of the relationship between two sets of variables (surname and Y chromosome), while explicitly and simultaneously controlling for the effect of a third factor (geography). The relationships between individual samples are set out in separate matrices for each of the three factors (Y chromosome, surname and geography).

The relationships of the McGuinness and East Ulster surnames were defined in four categories. Exact spelling matches were scored as 0, while 1 was applied to phonetically equivalent names with different spellings (for example Magenis and McGuinness). A value of 2 describes the relationship between names that are distinct but which known historical sources indicate a possible common origin in the last 1000 years (for example Neeson and McGuinness). Completely different surnames were classed as 3. A geographic matrix between individuals was obtained as the distance in kilometres between central points in their county or origin calculated using the great circle method. Samples with no specific information on a place of origin or those with an origin

outside Ireland were assigned a generic central coordinate for Ireland or Northern Ireland. Correlations were performed using several different categorical descriptions of Y chromosome relationships based on mutational divergence (repeat number difference) between 17 STR marker haplotypes. A boundary figure was set and differences in repeat number below this level were each treated as a separate category and scored according. Those at or above the boundary were not distinguished but collectively classed at the boundary figure. Correlations were carried out with repeat difference boundaries from 1, which describes a simple binary haplotype relationship of exact match versus non-match, through single integral steps up to 10 repeat units. As well as examining the overall relationship between Y chromosome and surnames, this process allows for the parallel investigation of the time-depth to any congruence. Grouping larger differences considers them as a single cumulative product of an older time horizon, allowing a clearer assessment of relationships due to smaller differences and thus more recent events.

The construction of surname and Y chromosome matrices was carried out using PERL software, while the partial Mantel correlation calculations were implemented using the PASSAGE package (Rosenberg et al. 2001). Significance was gauged through the random permutation of matrix elements and recalculation of the correlation coefficients 10,000 times to obtain the null hypothesis distribution of values. A partial Mantel test was also undertaken for Donohoe and Cavan surnames but this was based on a simple binary description of surname and genetic relationship as matches (scored as 0) or mismatches (scored as 1), considering all Donohoe spelling variants as identical names.

4.3 RESULTS

4.3.1 McGuinness Surname Y Chromosome Diversity

A MJ network displaying the relationship of 99 Y chromosomes from men with the McGuinness or a phonetically equivalent surname is shown in **Figure 4.6**. From herein "McGuinness" collectively describes all these names unless otherwise stated. The sample population represents about 2.7% of the current male Irish surname population, the latter estimated at 3,650 from *ad hoc* phone book calculations. There is high diversity with the observation of 48 haplotypes distributed over two SNP defined haplogroups (R1b3 and IxI1b2), immediately refuting the possibility of a common origin within the last thousand years. Notwithstanding this, there is also extensive haplotype sharing and evidence of diversified single lineages (clusters) centred on frequent haplotypes marked **A** to **F** in **Figure 4.6.** A summary of the relevant features of each can be found in **Table 4.5.**

Despite the high frequency of R1b3 Y chromosomes in Gaelic Irish surnames (Hill et al. 2000), the most prominent lineage in McGuinness, centred on Haplotype A, falls within the IxI1b2 haplogroup. The diversified cluster is the largest group of closely related Y chromosomes in the McGuinness surname, accounting for 28% of the sample population. Haplogroup I, defined by the M170 SNP (I-M170) can be further dissected into sub-groupings I1a, I1b, 11c by the SNP markers M253, P37 and M223 respectively. These were not directly typed here but their state was inferred through characteristic compound STR profiles at DYS19, DYS388, DYS390, DYS391, DYS392 and DYS393 (Rootsi et al. 2004). From this it appears that the major McGuinness-A lineage belongs to the I1c sub-grouping.

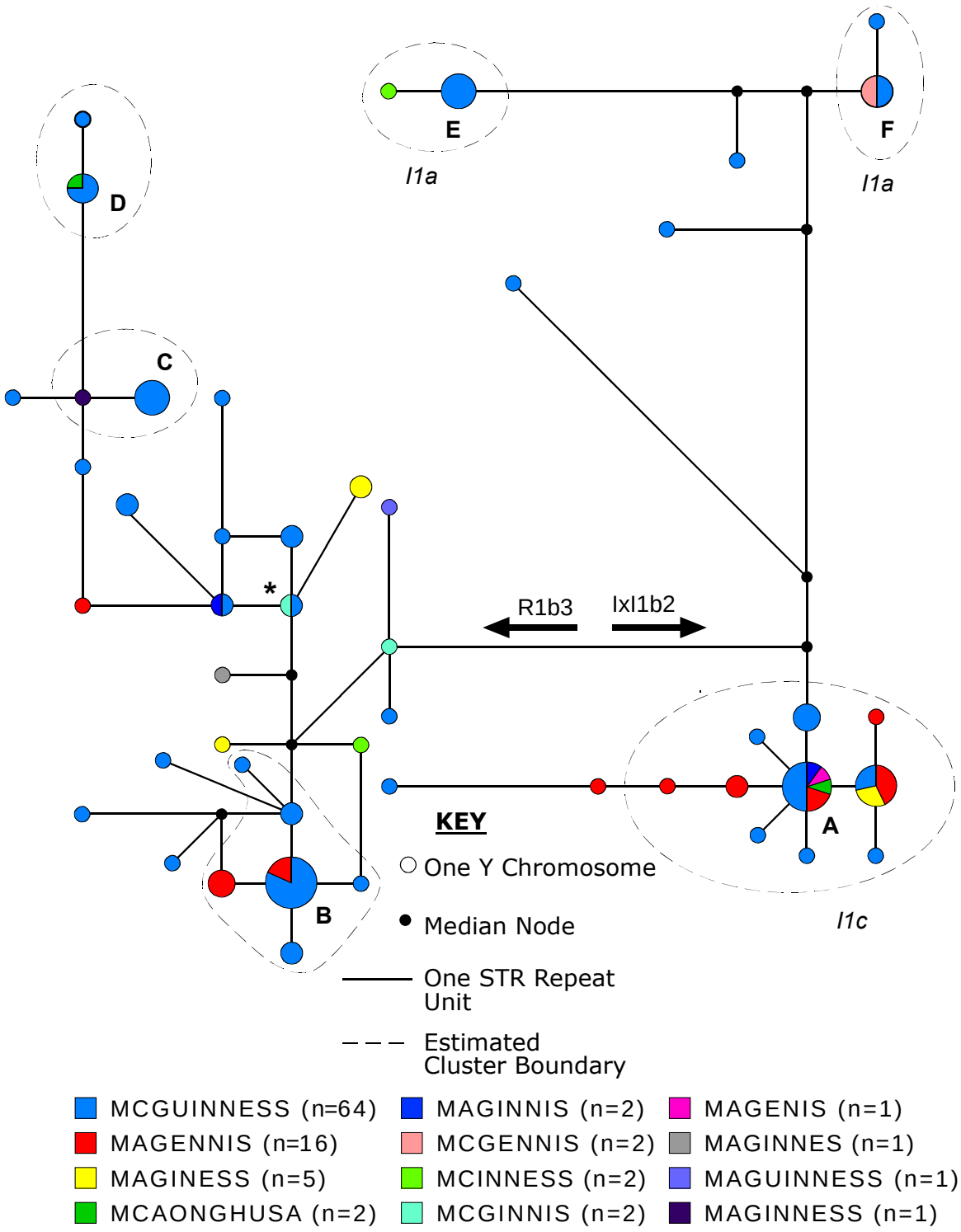


Figure 4.6 *MJ network of 99 McGuinness surname Y chromosomes. As with all networks, each circle represents a distinct haplotype (composed of 17 STR loci) with circle area proportional to frequency. The length of line between haplotypes shows their mutational divergence. Haplogroup boundaries are indicative, as the exact position of UEP events in the STR phylogeny is unknown. Despite obvious high diversity (48 haplotypes), several groups of closely related Y chromosomes are apparent and centred on ancestral haplotypes marked* ***A-F****. Inferred haplogroup I sub-groupings are indicated next to the relevant lineages (***Legend continued overleaf****)*

Figure 4.6 (preceding page) *MJ network of 99 McGuinness surname Y chromosomes. (***continued)**. *Modern spelling (different colours with "pie-slice" proportional to frequency) does not necessarily indicate closest genetic affinity, with the same version found between clusters and different versions within single lineages The most frequent haplotype in the general Irish population (3.8%) using these marker (Irish Modal Haplotype or IMH, Moore 2004) is marked *.*

The modal McGuinness haplotype (**B** in **Figure 4.6**), observed 11 times compared to 10 examples of haplotype **A**, forms the core of a second diverse cluster and accounts for 19% of the sample population. These two main lineages encompass a variety of surname spellings; conversely, many genetically unrelated individuals share the same version. Together these observations demonstrate that current spelling does not necessarily indicate closest paternal affinity within the McGuinness group of surnames. The other four clusters (**C**-**F** in **Figure 4.6**) are less frequent, each holding ~5% of the sample population. The remaining 30 individuals (31%) do not fall into any obvious broader clusters.

4.3.2 Geography and Age of McGuinness Y Chromosome Lineages

The network in **Figure 4.7** explores the phylogeography of the McGuinness samples. The two main lineages (**A** and **B**) display a strong East Ulster/North Leinster bias, centred on County Down, the historical place of origin for the surname. However, the minor clusters display markedly different geographic affinities. Three of these (**C**, **D** and **E**) are predominantly of West Ulster /North Connacht origin with no apparent link back to County Down. The final cluster (**F**) displays a third distinct and intermediate geographic affinity to South (Central) Ulster.

Differences in the size and diversity of the clusters are reflected in TMRCA estimates. The McGuinness-A lineage is the oldest, dated to 1600 years before present (YBP) using an evolutionary mutation rate and 670 YBP using a faster directly

observed/genealogical derived rate. However, these dates are not inconsistent with the putative original 10th century foundation, considering confidence intervals (CIs) associated with both mutation rate and sampling variance (see **Table 4.5** for all CIs) as well as uncertainty in exact cluster boundaries. The evolutionary rate is relied on henceforth as it is calculated over timescales similar to Irish surname history. The diverse spelling range of the McGuinness-A lineage as well as the occurrence of the ancestral haplotype in an individual with known genealogical links to the 17th century McGuinness Clan Chief provides further anecdotal support for this lineage representing the legacy of the initial surname founder. However, it also clear from a TMRCA estimate of 1010 YBP, that the McGuinness-B lineage represents an old (if relatively younger) input of a second male into the surname in the same area. The remaining geographically distinct clusters are younger than either major East Ulster area lineage but are of similar age to each other (~400 YBP) (**Table 4.5**).

4.3.3 McGuinness and Putative Subsidiary Surnames

Next, the relationship of McGuinness to the putatively derived surnames of Guinness, McCreesh, Neeson and McNeice was investigated by incorporating all names into a single network (**Figure 4.8**). None of these show a close affinity to the presumed founding McGuinness-A lineage. However, McCreesh samples are phylogenetically coherent with the McGuinness-B cluster, both sharing the modal ancestral haplotype. Furthermore, all eight Neeson Y chromosomes are identical to the mode of the distinct South Ulster McGuinness-F lineage. Y chromosomes from a small number of McNeice men show no relationship to each other nor to any McGuinness grouping. Similarly, although the three Guinness samples are closely related, they are not obviously linked to any of the main McGuinness lineages.

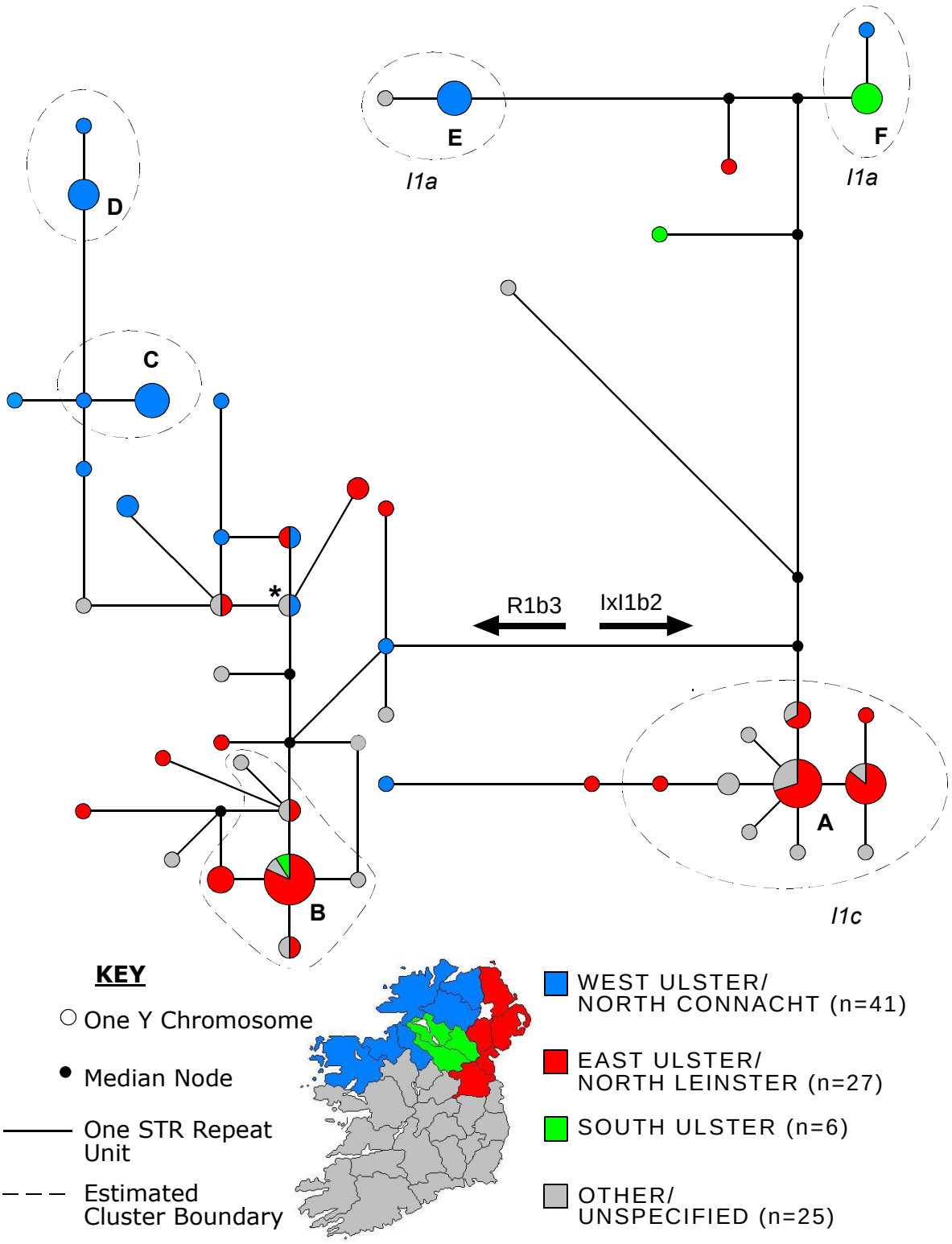


Figure 4.7 *Phylogeography of McGuinness surname Y chromosomes. The MJ network is identical to that in* ***Figure 4.6*** *with the geographic origin of each sample indicated by different colours. The major McGuinness lineages* ***A*** *and* ***B*** *are predominantly composed of samples from East Ulster and North Leinster. Lineages* ***C-E*** *are mainly West Ulster/North Connacht in geographic provenance while the final cluster (****F****) shows a third and physically intermediate South Ulster geography. * = IMH.*

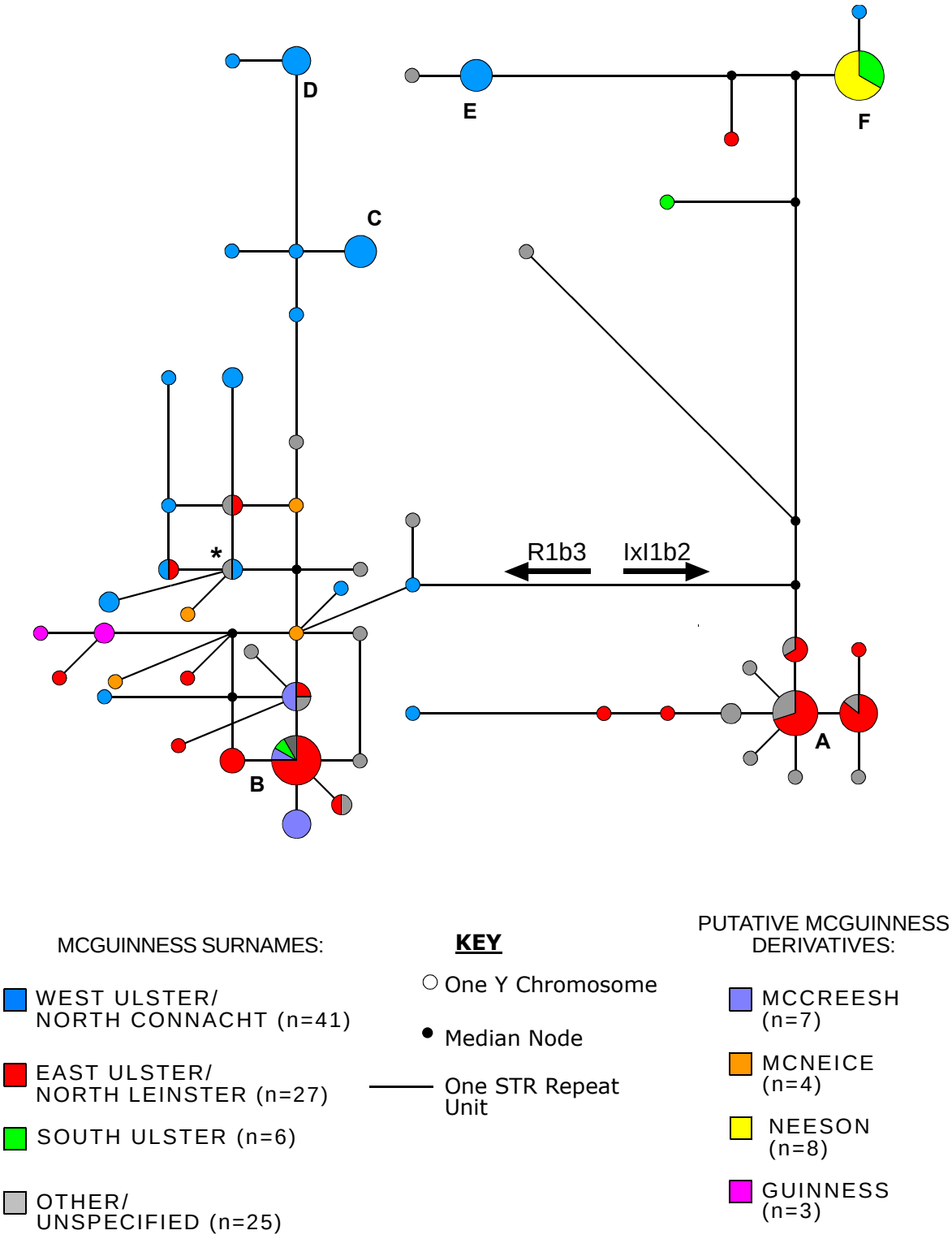


Figure 4.8 *MJ network of Y chromosomes from the McGuinness, McCreesh, Neeson, McNeice and Guinness surnames. Although the latter four are putative derivatives of the McGuinness surname, none show an affinity with the presumed founding McGuinness-**A** lineage. However, McCreesh Y chromosomes are coherently related to the other major East Ulster lineage McGuinness–**B**, while all Neeson Y chromosomes are cognate with the mode of the South Ulster McGuinness-**F** lineage. Guinness and McNeice show no obvious relationship to any major cluster. *=IMH.*

Lineage	Haplogroup	Geography	Cluster Size		Time to Most Recent Common Ancestor (TMRCA) Years Before Present (YBP)						
					Evolutionary Mutation Rate [A]				Directly Observed Mutation Rate [B]		
			n	% Surname	95 + CI[C]		95 - CI[C]	95 - 95 + CI[D]	95 + CI[C]		95 - CI[C]
McGuinness A	I1c	East Ulster	28	28.3	2560	1360 **1600** 1890	1160		1400	570 **670** 790	440
McGuinness B	R1b3	East Ulster	19	19.2	1620	810 **1010** 1260	730		890	340 **420** 530	280
MCGuinness C	R1b3	West Ulster	6	6.1	570	190 **360** 580	260		310	80 **150** 240	100
McGuinness D	R1b3	West Ulster	5	5.1	680	230 **430** 710	310		370	100 **180** 310	120
McGuinness E	I1a	West Ulster	6	6.1	570	210 **360** 590	260		310	90 **150** 250	100
McGuinness F	I1a	South Ulster	5	5.1	680	240 **430** 730	310		370	100 **180** 310	120
McCartan/ McGuinness A	I1c	East Ulster	34	NA	2820	1530 **1760** 2030	1270		1540	640 **740** 850	480
Down I1c1	I1c	East Ulster	61	NA	3700	2090 **2310** 2560	1670		2030	880 **970** 1070	640
Donohoe A	R1b3	Cavan Area	14	19.2	1710	830 **1070** 1360	770		940	350 **450** 580	290
Donohoe B	R1b3	Cavan Area	12	16.4	1710	820 **1070** 1400	770		940	340 **450** 590	290
Donohoe C	R1b3	Munster	16	21.9	2140	1050 **1330** 1660	970		1170	440 **560** 700	370

[A] Mutation rate of 2.76 x 10^{-5} per locus per year (Zhivotovsky et al. 2004)
[B] Mutation rate of 6.57 x 10^{-5} per locus per year; calculated from the data presented in Dupuy et al. (2004)
[C] 95% Confidence Intervals based on mutation rate uncertainty
[D] 95% Confidence Intervals associated with sampling variance

Table 4.5 *Summary of the major lineages in the McGuinness and Donohoe surnames giving TMRCA estimates and their 95% confidence intervals (CI) associated with mutation rate uncertainty (blue) and sampling variance (red).*

4.3.4 Y Chromosome Landscape of East Ulster

The significance of the substantial haplotype sharing within the McGuinness surname depends on the local Y chromosome genetic context, which was provided here by 194 Y chromosomes from 12 surnames with an origin in the same general East Ulster region. A single network combining these and 121 samples from McGuinness and putatively related names (315 Y chromosomes and 154 haplotypes in all) was difficult to construct and interpret. Consequently, only 48 repeated Y chromosomes, which are informative on haplotype sharing, were used in a single regional surname network (**Figure 4.9**). Although some repeated haplotypes are shared across surnames most (~71%) are not. Furthermore, there is good phylogenetic congruence between different Y chromosomes of the same name, even if they are sometimes shared between surnames.

Most Y chromosomes from this area fall into haplogroup R1b3. However, the I1c type of the founding McGuinness-A lineage also occurs in three other surnames (McEvoy, McCartan and McVeigh), suggesting that it is a general feature of the East Ulster genetic landscape. Together they form a cluster (**Figure 4.9**, inset) that displays a coherent and 'star like' topology consistent with an *in situ* expansion. The antiquity indicated by distribution over several surnames is reflected in a TMRCA estimate of 2310 YBP. In contrast, the general Irish population repertoire of haplogroup IxI1b2 displays little coherency or repetition of haplotypes (**Figure 4.10**) (Hill 1999; Hill et al. 2000; Moore 2004), features proposed to reflect its later piecemeal introduction into Ireland after initial settlement with R1b3 Y chromosomes. The appearance of the principal, and presumed founding McGuinness and McCartan lineages within this cluster is remarkable given that genealogical records name a common 6th century AD ancestor (*Mongan*) for the founders of these names (Byrne 1997, pg 287). The TMRCA

of the two lineages estimated at 1755 YBP is not inconsistent with the suggested genealogical timeframe.

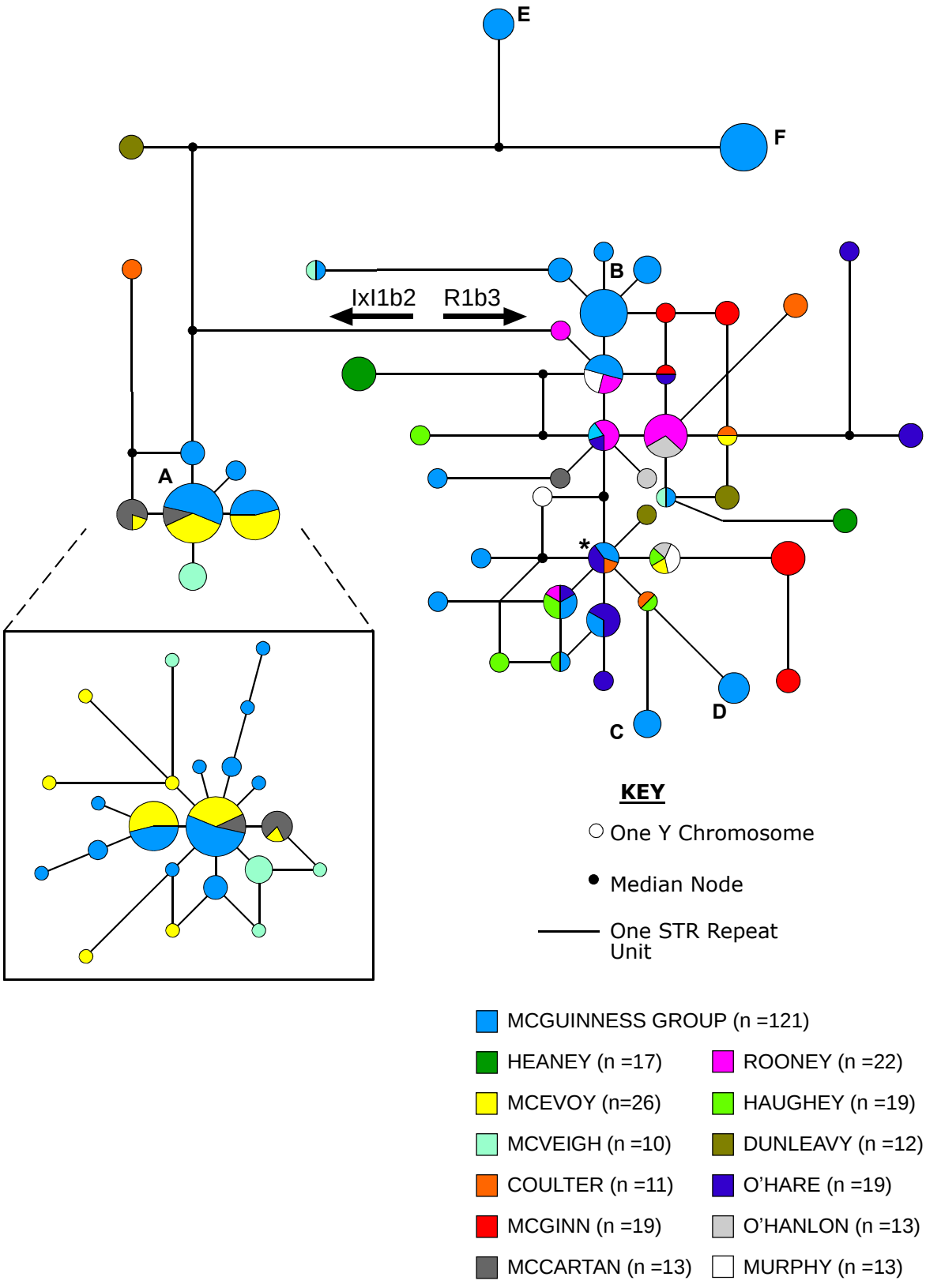


Figure 4.9 *MJ network of Y chromosome haplotype sharing between East Ulster surnames. Only the 48 repeated haplotypes from amongst the 154 observed in 315 individuals are shown. McGuinness and related surnames were treated as a single group with the major lineages of this surname marked **A-F** as in preceding Figures. Most haplotypes (~71%) are not shared between surnames indicating largely separate origins. A notable exception concerns the main McGuinness-**A** lineage, which fits into a coherent and trans-surname haplogroup I1c cluster (full extent, including singletons, is shown inset) consistent with an early in situ origin in East Ulster. *= IMH.*

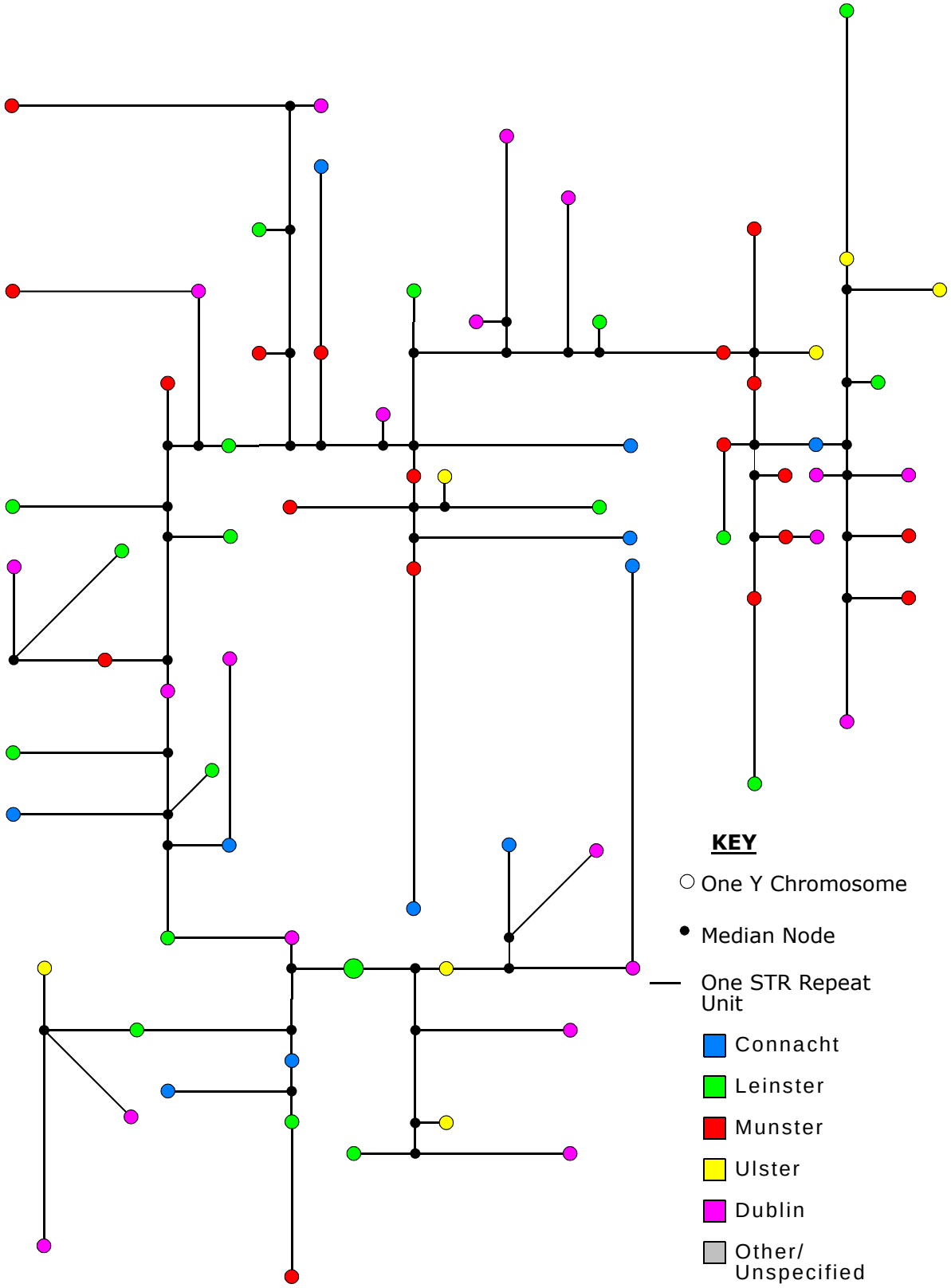


Figure 4.10 *MJ network of haplogroup IxI1b2 Y chromosomes in a general Irish population sample. The 77 individuals show virtually no haplotype repetition or coherent clustering in contrast to the local East Ulster Y chromosome landscape (see* **Figure 4.9***). Network drawn with taken data from Moore (2004).*

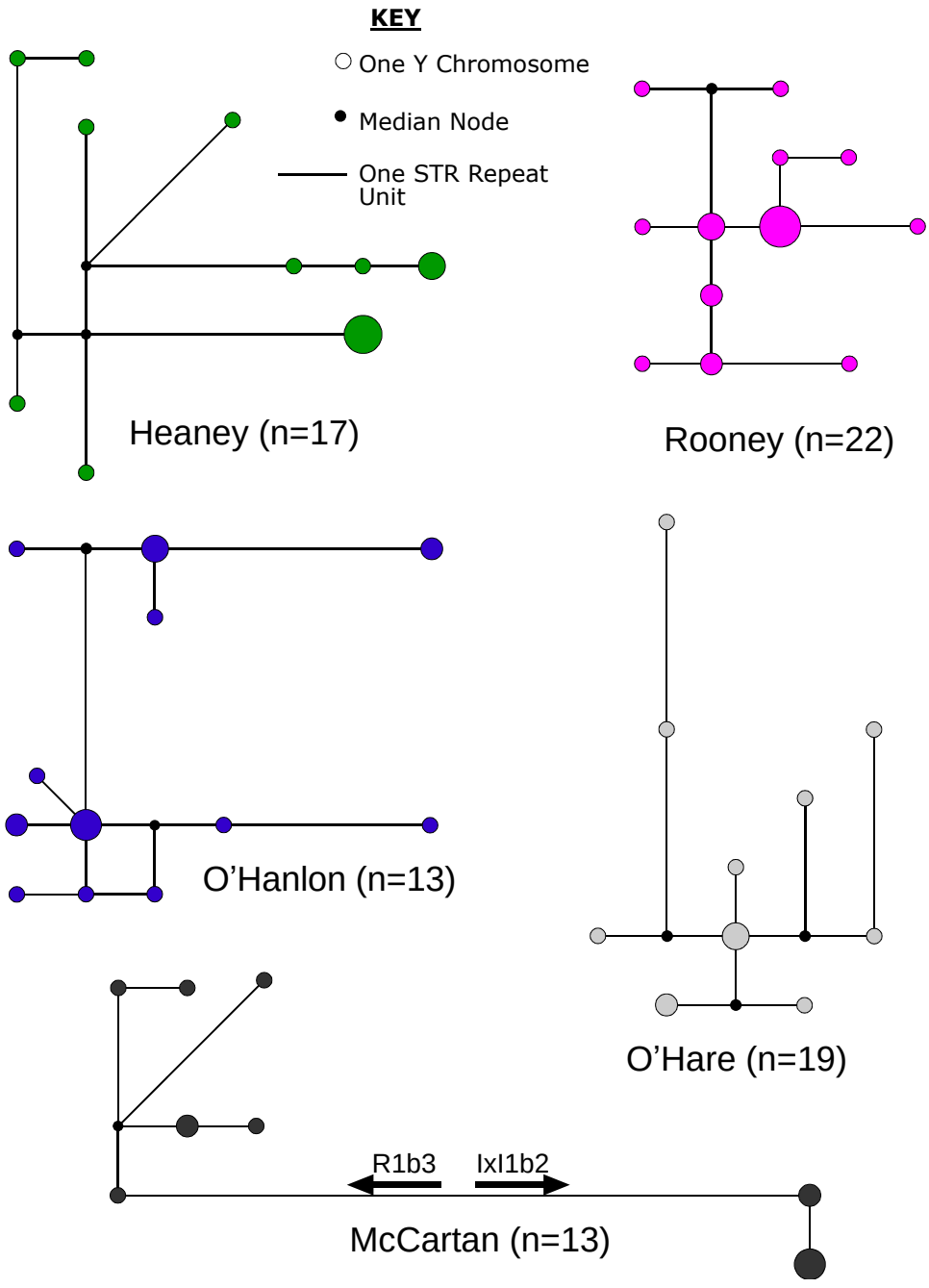


Figure 4.11 *MJ networks of individual East Ulster surname samples (***Full legend overleaf**)

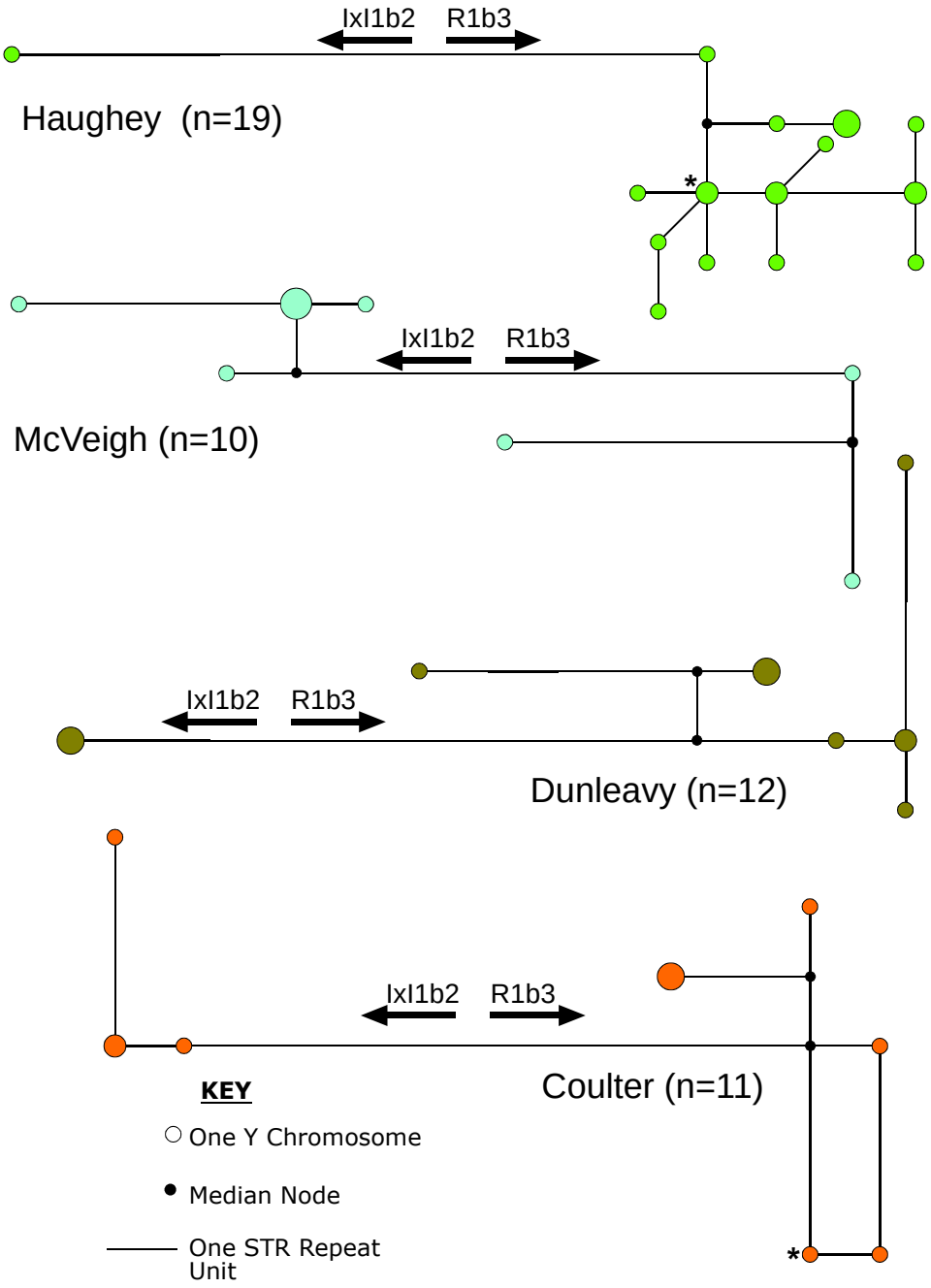


Figure 4.11 (preceding page and above) *MJ networks of individual East Ulster surname samples, used in* ***Figure 4.9*** *and not shown in full elsewhere (for McEvoy and Murphy see Chapter 5 and* ***Figure 4.13*** *for McGinn). Although sample sizes are small they illustrate the diversity of the East Ulster area. Like McGuinness, they show evidence of recent multiple founder effects through repeated and clustered lineages but overall diversity within individual surnames is generally inconsistent with a single paternal ancestor for all bearers. * = IMH.*

The presence of additional, if less prominent, haplogroup I-M170 subtypes both in McGuinness and other surnames from the area confirms the general diversity of East Ulster prior to the adoption of surnames *ca.*1000 years ago. Full networks for other East Ulster area surnames (not examined in detailed elsewhere) are shown in **Figure 4.11**. Overall these show a similar picture to McGuinness with evidence of recent paternal ancestry among subsets of surname bearers witnessed by repeated haplotypes and tentative evidence of diversified lineages. However, due to small sample sizes, detailed and definitive assessments of their individual histories are not attempted.

4.3.5 Congruence of Y Chromosomes and Surnames in East Ulster

The extent to which surnames are real indicators of biological paternal ancestry was quantitatively assessed using three different methods. Over all East Ulster surnames (conservatively treating the putative McGuinness derivatives separately) there is a 14.5-fold greater chance of finding an identical Y chromosome (17 STR marker haplotype) within men of the same surname compared to geographically and culturally controlled groups of identical sample size. The figure for individual surnames varies widely. In McGuinness it is a smaller but still robust 3.5-fold increase, possibly reflecting the substantial modern legacies of several males within the name. A second approach using analysis of molecular variance (AMOVA) confirms surname-defined populations as real Y chromosome genetic divisions. A highly significant ($p<0.00001$) 30.6 % of the variance is found between the 17 surnames. The remaining 69.4% occurs between individuals of the same name, simultaneously indicating the considerable heterogeneity within surnames.

The results of partial Mantel correlations, holding the effect of geography constant while examining the congruence of surname and Y chromosome distance, are shown in **Figure 4.12**. The correlation was carried out several times using a successively increasing number of repeat differences from 1 to 10 as a boundary at or above which all differences were classed together. While all correlations were highly significant (p<0.002), the strongest relationship (r=0.1919, p < 0.0001) occurs when Y chromosome differences greater than 1 repeat unit are treated as a single category. Further and finer distinction of relationships by higher boundaries leads to a steady decline in the positive correlation values. These observations indicate that overall congruence of surname and Y chromosome is very significant and primarily due to a relatively recent event in line with the 1000 year timeframe of Irish surname history.

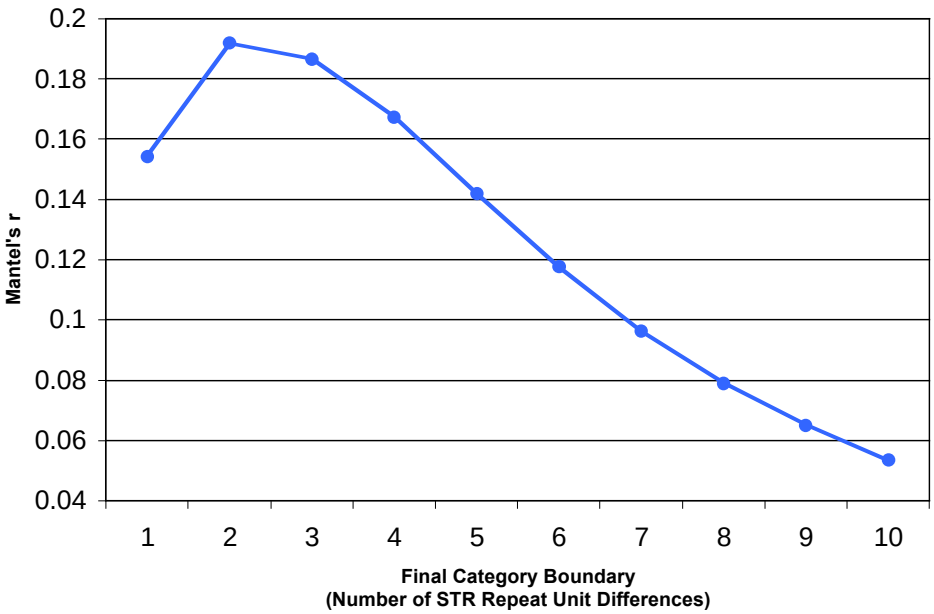


Figure 4.12 *Partial Mantel correlations between surname and Y chromosome distances, holding the effect of geography constant. The correlation was carried out 10 times using different categorical descriptions of Y chromosome relationships. These were determined by successively increasing, from 1, a boundary number of repeat differences and considering all those at or above this level as a single class. While surnames and Y chromosomes are significantly correlated (p <0.002) over the full range, the strongest values occurs when the boundary is low (1, 2 or 3 repeat units). The latter finding indicates that the relationship between surname and Y chromosome is primarily due to relatively recent events.*

4.3.6 Origin of McGuinness in West Ulster/North Connacht

The McGuinness surname appears to have multiple and relatively recent origins in the West Ulster/North Connacht region (**Table 4.5**). One explanation consistent with these observations is the transmutation of other distinct but phonetically similar names to McGuinness through relatively recent anglicisation events. The general West Ulster area coincides with the distribution of several possible source names including McGinty (*Mac Fhinneachta*) and McGinley (*Mac Fhionnghaile*). The relationship of these to 27 McGuinness Y chromosomes originating in the West Ulster area (or with membership of a predominantly West Ulster cluster) is examined in **Figure 4.13**. Although partial (repeated haplotypes) McGinn (*Mac Fhinn*) results were included in **Figure 4.9**, they are also shown in full in **Figure 4.13** because of the phonetic similarity to McGuinness and a geographic distribution extending to County Tyrone in West Ulster. However, there is no evidence of any major inter-relationship with only two haplotypes shared between the four surnames. This illustrates once again that surnames are a powerful indicator of recent paternal ancestry even in geographically restricted areas.

4.3.7 Donohoe Surname Y Chromosome Diversity

A network showing the relationship of 73 Y chromosomes from men with the Donohoe and spelling variant surnames appears in **Figure 4.14**. These samples represent about 1% of the estimated 7,500 male bearers of these surnames in Ireland. Y chromosome diversity is high and no single lineage predominates. Instead, three approximately equal-sized clusters are observed, each centred on a frequent ancestral haplotype (**A**, **B**, and **C** in **Figure 4.14).** TMRCA estimates for these three lineages are also similar at 1070 YBP for Donohoe-A and B and 1330 YBP for cluster C (see **Table 4.5** for Confidence Intervals), suggesting at least three major early foundations of the name.

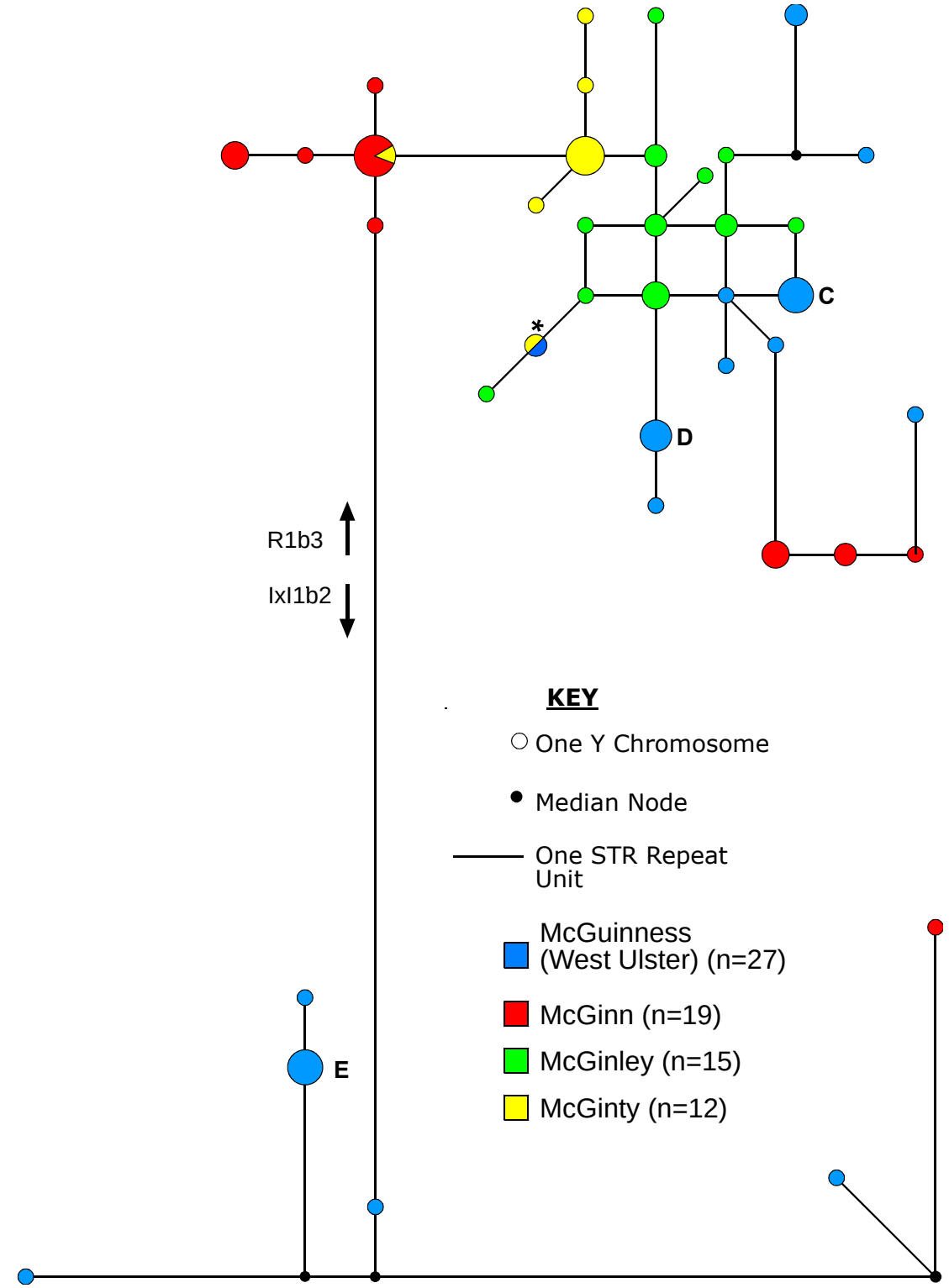


Figure 4.13 *MJ network of McGinn, McGinty, McGinley and West Ulster/North Connacht McGuinness Y chromosomes. The recent origin of several McGuinness lineages* ***(C****,* ***D*** *and* ***E*** *from previous figures) in this area may be explained by an anglicisation event from another phonetically similar surname. However, the absence of haplotype sharing argues against such a provenance, at least, from these three surnames (McGinn, McGinty and McGinley). *= IMH.*

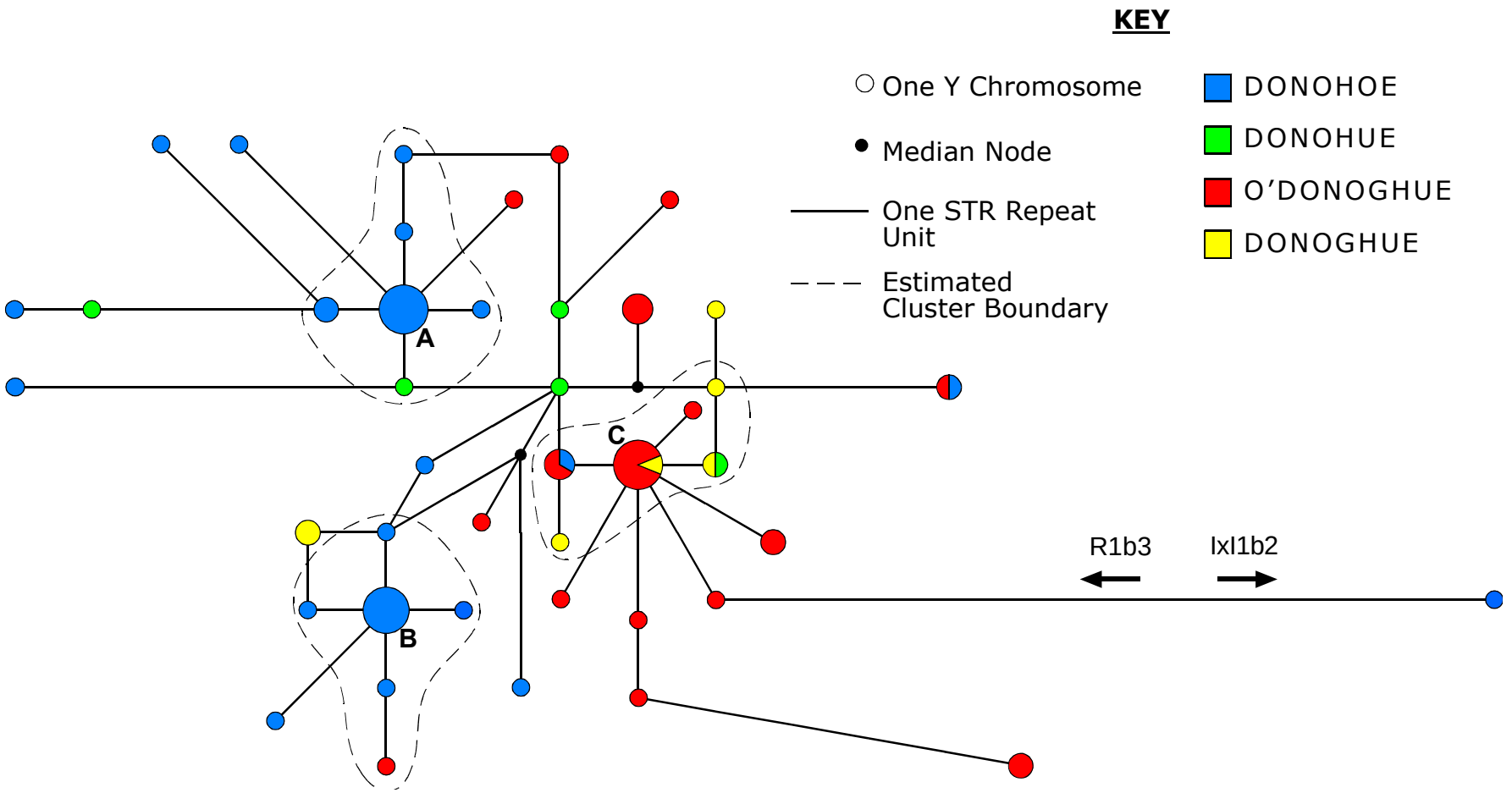


Figure 4.14 *MJ network of 73 Y chromosomes from the Donohoe surname. 43 haplotypes were observed and three diversified clusters, centred on approximately equally frequent ancestral haplotypes (marked **A**, **B** and **C**) are apparent. There is good congruence between these paternal genetic divisions and variation in surname spelling with the 'Donohoe' version predominating in clusters A and B and the '(O)'Donoghue' type characteristic of cluster C.*

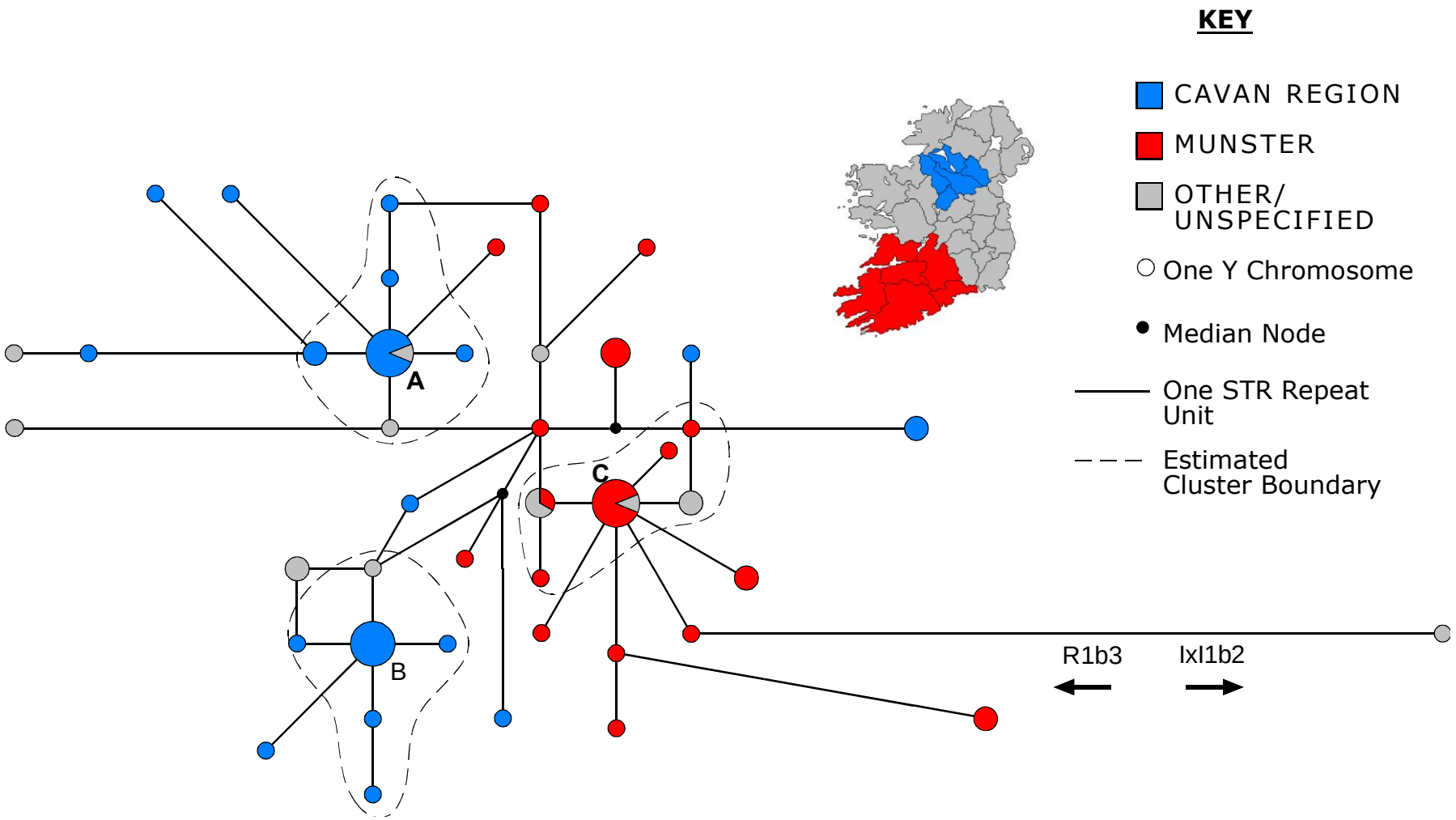


Figure 4.15 *Phylogeography of Donohoe surname Y chromosomes. The MJ network is identical to that in* ***Figure 4.14*** *with the geographic origin of each sample indicated by different colours. Lineages* ***A*** *and* ***B*** *show an obvious bias to the same area of Cavan and surrounding counties indicating a dual foundation of the surname in this region. In contrast a single major origin (cluster* ***C****) is observed in Munster. The absence of Cavan samples in the Munster cluster and vice versa suggests separate origins with little or no subsequent migration between areas over about 1000 years of surname history.*

Unlike McGuinness, different spelling versions show a broad congruence with Y chromosome relationships. Clusters A and B are composed mainly of the 'Donohoe' version while cluster C is identified with the '(O') Donoghue' spelling. Furthermore, this agreement is reflected in the geographic origin of the samples (**Figure 4.15**). 'O'Donoghue' is typical of South Munster, where most members of lineage C trace their origin. However, and surprisingly, the other two major foundations are associated with the same area, that of Cavan and surrounding counties. There is no evidence of mixing between Munster and Cavan lineages confirming independent origins and indicating little or no migration between the two regions over the 1,000 years of surname history. A substantial number (42%) of individuals are not part of any broader cluster.

4.3.8 Y Chromosome Landscape of the Cavan Region

The relationship of 62 Y chromosomes in 10 other surnames from Cavan and surrounding counties is shown in **Figure 4.16**. As noted in previous local sample populations, there is general congruence of haplotypes with surnames indicating underlying paternal ancestry. Six of the nine repeated haplotypes include individuals from one name only, while there is also locally good phylogenetic coherence between different Y chromosomes of the same name. The position of the Donohoe samples in this local genetic landscape, through repeated haplotypes, is shown in **Figure 4.17**. Considering the two major Cavan lineages of the surname, the Donohoe-B cluster remains distinctive; however the Donohoe-A ancestral haplotype is identical with the modal Maguire type.

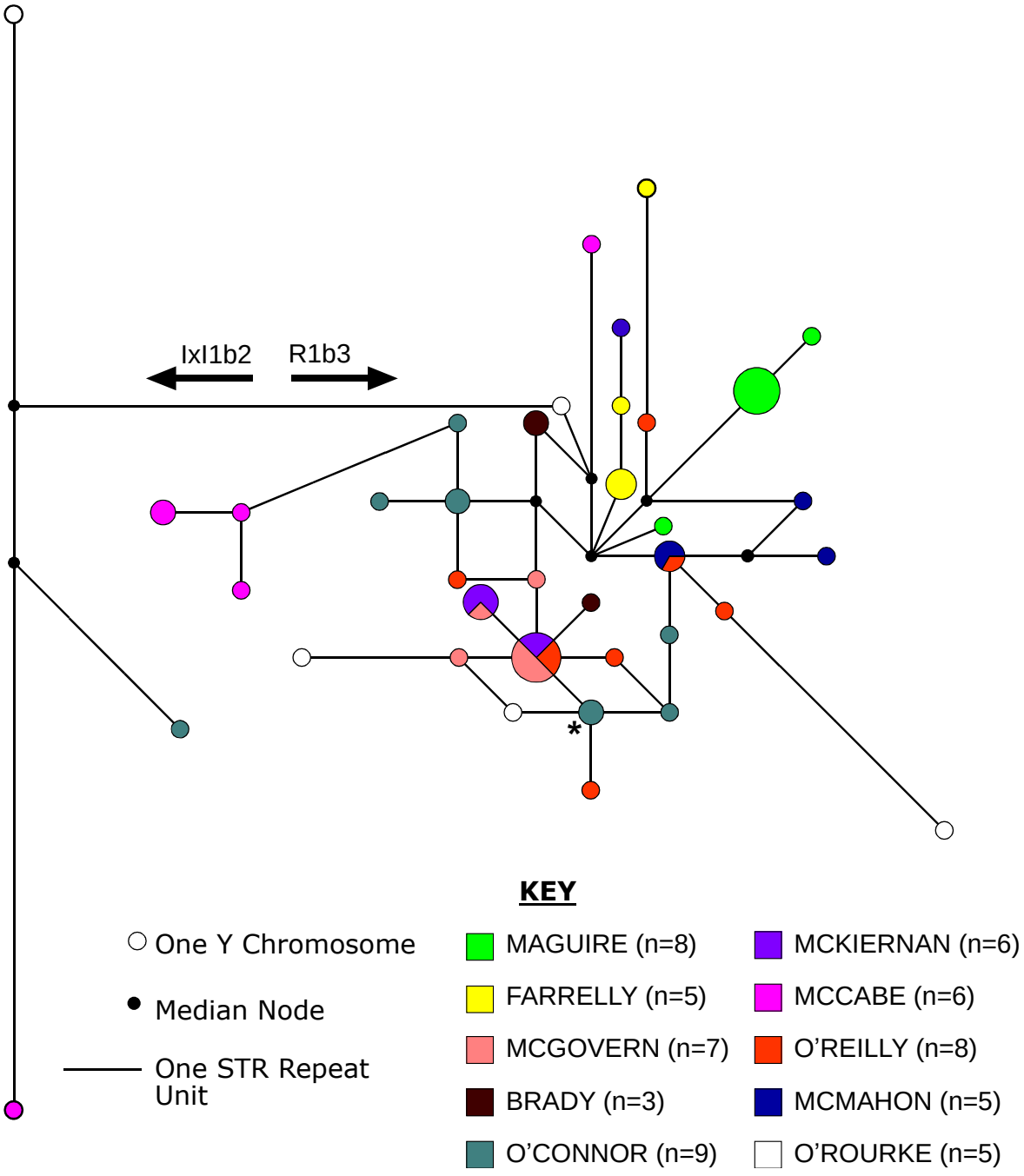


Figure 4.16 *MJ network of Y chromosomes from 10 Cavan area surnames. The 62 individuals are spread across 38 haplotypes. Most repeated haplotypes contain individuals of one name only and there is some evidence for tentative clustering, at least in subsets, of different Y chromosomes within the same name. * = IMH.*

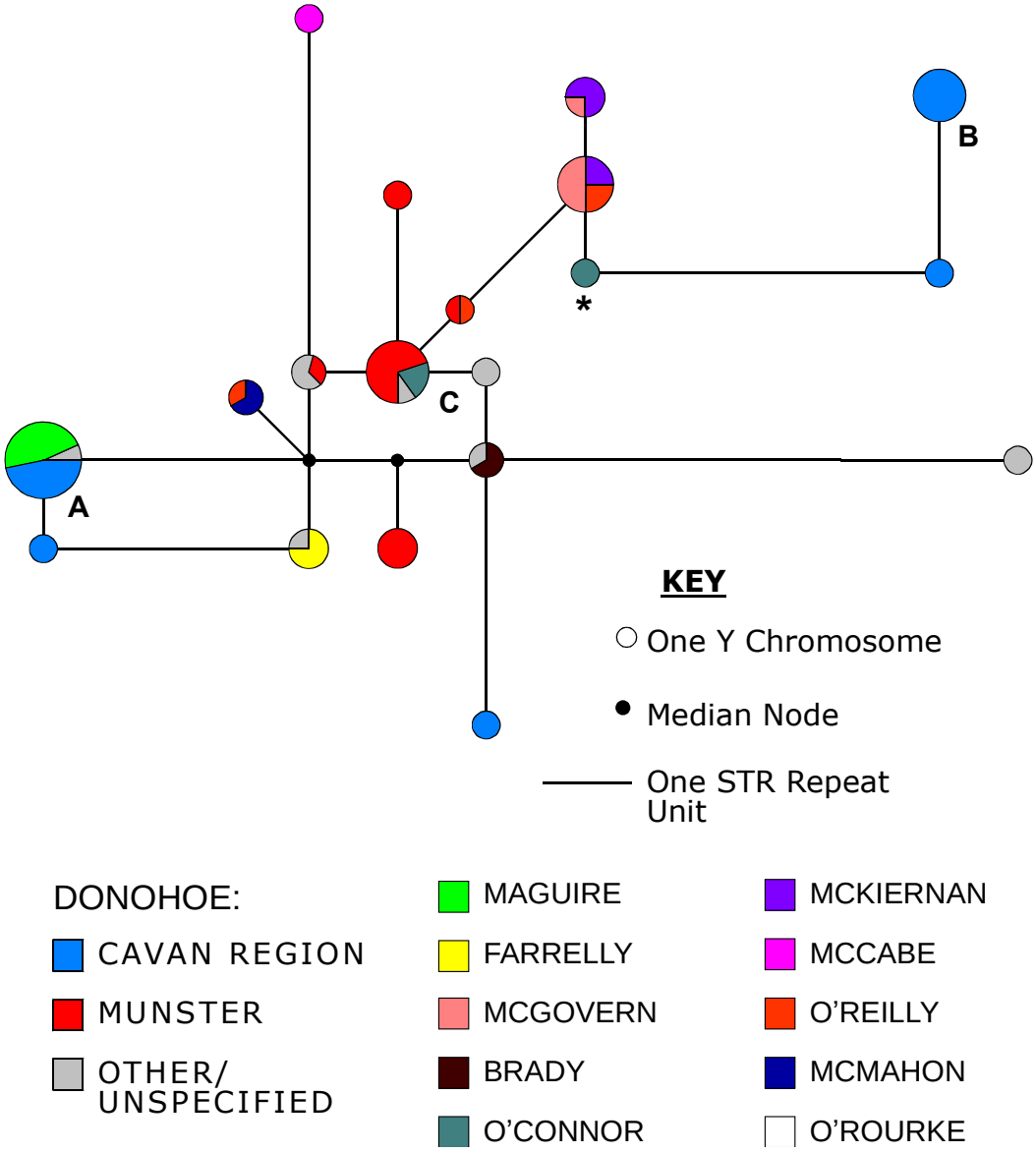


Figure 4.17 *MJ network of repeated Y chromosomes in Donohoe and other Cavan area surnames. While one of the two main Cavan region Donohoe lineages (**B**) is unique in a local genetic context, the presumed ancestral haplotype (**A**) of the other is cognate with the Maguire modal type. The Munster Donohoe cluster is also indicated (**C**). *=IMH.*

In assessing the congruence of paternal ancestry and surname in the Cavan area, samples from the Donohoe group of names with a confirmed Munster origin (n=22) were excluded, as there is ample historical and now genetic evidence to support a separate origin. Within the Cavan area population sample there is 7.2-fold greater tendency for identical haplotypes to be found in men of the same surname. AMOVA confirms that a large (20.4%), and highly significant ($p<0.00001$) proportion of the Y chromosome variance is accounted for by surnames. This demonstrates again, importantly in a second local context, that surnames correspond to real paternal ancestry. Finally a partial Mantel test, based on a binary (match or no match) description of surname and Y chromosome relationships (including Munster Donohoe samples) shows a significant but low positive correlation between the two variables even in the face of any geographic effect ($r=0.096$ $p<0.00001$).

4.4 DISCUSSION

4.4.1 Surnames and Patrilineal Ancestry

The detailed examination of Y chromosome diversity in McGuinness and Donohoe, together with other surnames from their respective local areas of Down and Cavan, demonstrates that Irish surnames reflect recent common paternal ancestry. However, these results simultaneously reveal considerable complexity within individual surname histories. It seems clear that there is no simple correspondence of a single name to one founding male. Instead, there is evidence of multiple paternal ancestries, each of varying modern legacy. Consequently, there is an important role for a molecular Y chromosome approach in the elucidation of Irish surname adoption and history.

4.4.2 Origin of the McGuinness Surname

The Y chromosomes of McGuinness surname bearers record an early and principal foundation of the name, in agreement with historical accounts and timeframes. Presumably this lineage represents the genetic legacy of the 10th century eponymous *Aongus* yet only 28% of the modern surname population appear to descend from him. The vast majority thus have alternative paternal histories indicating the permeability of the surname to other males. The range and number of haplotypes outside the founding cluster suggests these subsequent introgressions were common. Furthermore, the detection of a relatively old lineage from the same area as *Aongus*, supports an early start and long history to this process. However, it is impossible to judge from the genetic data alone the reasons for these other than to conclude that the coinheritance of Y chromosome and surname from the original founding father was broken on multiple occasions. Non-paternity events are an obvious but certainly not the only explanation (see also Chapter 5).

Additional prominent but less frequent lineages, apparently of some antiquity, are both genetically and geographically distinct of the name's County Down historic heartland. Their origins may lie in the same processes that created diversity in East Ulster transposed to different areas but it seems more likely that they represent completely separate historical foundations, which never had anything more than a superficial connection to the main County Down surname. *Aongus* was a relatively common personal name and descendents of different bearers could correctly but independently adopt the hereditary surname. Alternatively, they may represent 'attraction' anglicisation events to the McGuinness name. Although not related to several phonetic possibilities (McGinn, McGinley, and McGinty), this does not preclude such an origin from other names. Finally, the I1a haplogroup of one of these lineages, which is atypical of Gaelic Irish surnames generally, may indicate an exogenous origin. The occurrence of a similar name in Scotland and the 17th century plantation of Ulster together present a possible explanation.

4.4.3 McGuinness Family Reunion

The long duration of Irish surname history together with the process of anglicisation have created considerable opportunity for the diversification of names from a single original version into many different forms today and vice versa. The McGuinness founding lineage itself provides an interesting insight into the haphazard nature of anglicisation with no less than six different spelling versions. However, the broader reconstruction (or deconstruction) of more obscured potential relationships seems to be a powerful genealogical application of the Y chromosome. Suggestions that the McCreesh and Neeson surnames are corruptions of the Irish *Mac Aonghusa* (McGuinness) are apparently confirmed but only with an unexpected twist. Although, they each show clear congruence with a McGuinness lineage, these represent two

distinct and later additions to the name. Therefore and somewhat paradoxically, while both are related to McGuinness, they are not related to each other. In contrast, the exact origin of the Guinness surname remains somewhat enigmatic. Although it is not closely associated with the larger McGuinness clusters, this does not preclude descent from one of the multiple smaller lineages. The presence of a Guinness place name (townland) in Co. Down presents another possibility. Surprisingly, this does not derive from the surname but rather comes the Irish *Gion-ais* meaning 'wedge back' (Dr. Kay Muhr, Queens University, Belfast, personal communication). The surname could have arisen from this, simultaneously explaining the family history link with Co. Down but lack of obvious McGuinness paternal ancestry.

In a display of versatility and range, the Y chromosome also proved informative on pre-surname relationships. The apparent confirmation of genealogical records linking McGuinness and McCartan to a 6th century AD common ancestor (*Mongan*) is remarkable. However, the occurrence of this Y chromosome in the McEvoy and McVeigh surnames, which probably have a common origin (see Chapter 5), cautions of local population stratification effects and is a reminder that multiple foundations by different men with the same or closely related Y chromosomes will be difficult to detect. While a formal genealogical connection between McEvoy and McGuinness is not known or refuted (to the author's knowledge), the McEvoy family were the hereditary 'installers' of the McGuinness chief (Mathews 1968). Therefore, like McGuinness and McCartan, the congruence of Y chromosome with McEvoy may actually represent a real and perceived pre-surname kinship.

4.4.4 Ancient Y Chromosome Diversity in East Ulster

Most individuals, both from McGuinness and other East Ulster surnames, fall into the R1b3-M269 haplogroup. However, the high frequency of haplogroup I-M170 Y chromosomes is somewhat surprising. Its relative rarity in other Gaelic Irish surnames, converse abundance in names with an Anglo-Scottish origin (Chapter 5; Hill 1999; Hill et al. 2000) and high general diversity in Ireland together led to the suggestion that it represented the signature of post-foundation, and perhaps very recent, migration. The phylogenetic coherence of the I1c cluster in East Ulster supports an *in situ* population expansion of some antiquity (perhaps about 2500 YBP), strongly suggesting that some non-R1b3 lineages have a longer continuous history in Ireland than was perhaps previously anticipated. The high frequency of this type in the Gaelic surname population of Leinster relative to Connacht (Hill et al. 2000) might best be interpreted as the legacy of longer term links between this region and Britain or Europe, that pre-date the Anglo-Norman conquest.

In this context, the origin of the major I–M170 types is of interest and importance. Both the I1a-M253 and I1c-M223 sub-groupings display decreasing frequency gradients from peaks in Scandinavia and Germany/Netherlands respectively (Rootsi et al. 2004). Their conveyance West *via* Britain to Ireland is consistent with the continued east–west cline within the island (see Chapter 3). However, the ultimate distribution of the I1c and I1a lineages may be the result of a post-glacial re-expansion from an Iberian refugium. Considering the impact of this process on the genetic legacy of the European Atlantic façade (see Chapter 2), it is at least conceivable that these already diversified lineages (compared to the R1b3-M269 type) were a minor component of initial or early settlement.

Historical evidence provides an interesting angle to these possibilities since the McGuinness and McCartan progenitors arose from a population grouping that is, by some accounts, exogenous. This is based mainly on their name, Cruthin, which is a Q-Celtic term meaning Britons. Given the eastward centre of gravity in the distribution of the I1a and I1c lineages and the close proximity of Scotland, it is tempting to postulate an origin there for these types. Such long-term links between the two countries, also indicated by maternally inherited mtDNA (Chapter 2), could account for the apparent exceptional diversity of this area of the island from an early (pre-surname) stage.

4.4.5 Origin and History of Donohoe Surname

The Y chromosome diversity of the Donohoe surname provides additional support for both the common patrilineal ancestral basis of surnames and the simultaneous complexity of individual surname histories. The distinctiveness of the Munster O'Donoghues is unsurprising given both historical accounts and the discrete geographical distribution of the name. However, the equal genetic legacy of two major founders in the Cavan region was previously unknown. It seems reasonable to conclude that one of these represents the historically-posited early 12th century eponymous *Donnchadh* ancestor. The origin of the second major group is, however, more enigmatic and parallels the situation with McGuinness in East Ulster. The observation that the mode of one of these Donohoe lineages is cognate with that of Maguire suggests a potential explanation. Another *Donnchadh* in the personage of Donnchadh Ceallach Maguire is proposed to represent an independent eponymous ancestor to the Donohoe name (Bell 1988, pg 54). His conquest of parts of Fermanagh on the Cavan border makes the historical geography consistent but his timeframe at *ca*. 1450 AD about 350 years after the putative original founder seems to preclude this explanation given the equivalent date and size of the two clusters. However, such a margin of error could be

comfortably contained in the uncertainty of TMRCA estimates. By way of anecdotal support, the only Donohoes in the sample population with a County Fermanagh origin are of the Donohoe/Maguire modal haplotype. Were this explanation correct then it would suggest that variance in reproductive success could outweigh the age of foundation in determining the number of modern descendents (see Chapter 5). However, it is also possible that the two Donohoe foundations are both early with the Maguire connection a coincidental product of local Y chromosome stratification or earlier genealogy.

4.4.6 Conclusions

The detailed study of Y chromosome diversity in the McGuinness and Donohoe surnames demonstrates two general and notable findings. Firstly, it provides strong evidence of underlying paternal ancestry to Irish surnames in agreement with historical evidence and surname nomenclature. The relationship between the two is significant and robust, despite any local genetic substructure. This suggests Y chromosome diversity was sufficient even 1000 thousand years ago for many or most surname adoptions to be reflected in distinct Y chromosomes. Secondly, and based on the foregoing observations, modern Y chromosome diversity provides a powerful means of unravelling the old and perturbed history of Irish surnames revealing otherwise cryptic complexity even within single surnames. The potential range and finesse of this application is remarkable, exemplified by the McGuinness surname, with insights possible into early population origins, pre-surname genealogical links as well as original foundation(s) event(s) and subsequent surname history right up to the very recent past.

CHAPTER 5

Y CHROMOSOME INVESTIGATION OF IRISH SURNAME HISTORIES

5.1 INTRODUCTION

5.1.1 Gaelic Surname Diversity

A general similarly in cultural and temporal origins links indigenous Gaelic Irish surnames. However, both historical accounts and modern variations in frequency and geographical distribution point to the importance of individual and variable factors in the course of each surname's evolution. In-depth examination of Y chromosome diversity within the McGuinness and Donohoe surnames, together with that found in other names from their respective local areas, demonstrated the patrilineal ancestral basis of Irish surnames. It also illustrated the valuable role for a molecular Y chromosome approach in unravelling and reconstructing surname origin and history (Chapter 4).

In this Chapter, the Y chromosome diversity in eleven additional Gaelic Irish surnames, along with some potential derivatives, is examined in detail. These encompass variability in current frequency, geographic origin and distribution, as well as prior historical information on the nature and number of foundations (monogenic versus polygenic origins). While allowing valuable insight into each individual surname history, they also provide the opportunity to assess the presence or absence of common themes and factors in Irish surname development. A brief summary of the relevant history of each name, focusing on issues that may be addressed in a Y chromosome genetic framework, such as number of founders and potential relationship between names, follows in succeeding sections. Historical narratives are primarily taken from McLysaght (1982, 1985a, 1985b) and these are examined in concert with a novel analysis of the geographic distribution of each name in the mid-19th century using data from 'Griffith's Valuation' (see **Section 4.1.5**).

5.1.1.a Byrne

Byrne is one of the most common surnames in Ireland with an estimated 40,000 bearers. The name derives from the Irish *Ó Broin* meaning 'grandson of Bran', reputedly a mid-11th century King of Leinster from the modern Wicklow area. Some 900 years later, the distribution of the name was still focused on Wicklow, with a gradient of decreasing frequency radiating from the East Coast (**Figure 5.1A**). Bucking this general trend, there are minor increases in frequency in some western Counties of Connacht and also in Donegal.

5.1.1.b Kennedy

As with Byrne, the Kennedy surname has a prominent historical eponymous ancestor in the father of Brian Boru or *Cinnéide* (*ca*. early 11th century), whence the original Irish version *Ó Cinnéide*. Although substantially less frequent than Byrne, it remains one of the most common Irish names with an estimated 20,000 bearers. In the 19th century the name was concentrated on its historical place of origin around modern Tipperary (**Figure 5.1B**) with additional discrete but small frequency foci in the Ulster Counties of Donegal and particularly Antrim. Historical sources suggest a branch of the Tipperary family migrated to Antrim *ca.*1600 AD but Kennedy can also be of Scottish origin and was probably introduced to Ulster during the plantations of the 17th century.

5.1.1.c Ryan

The vast majority of the modern Ryan surname population (estimated at approximately 38,000 people in Ireland) come from the Irish *Ó Maoilriain*, in English O'Mulryan and later further anglicised to Ryan. There are apparently no historical details on this eponym, which means 'grandson [or other male descendent] of a follower of Riain'.

Indeed despite its current high frequency, the name did not become prominent until the 14th century when the family rose to power around modern Tipperary, an area that still represented the striking focus of a limited surname distribution in the 19th century (**Figure 5.1C**). A small number of Ryans centred on County Carlow in Leinster are thought to have an independent origin from the distinct Irish name Ó*'Riain* ('grandson of Riain') also anglicised to Ryan.

5.1.1.d O'Neill

O'Neill comes for the personal name Niall. A similarly named *Uí Néill* dynasty (derived from the putative 5th century AD *Niall Noígiallach* or Niall of the Nine Hostages) was the most powerful grouping in Ireland in the centuries before the adoption of surnames. Although the surname O'Neill is reputed to ultimately descend from this source, the specific eponym for the hereditary surname was *Niall Glún Dubh* (Niall Black Knee), an early 10th century claimant to the high kingship of Ireland from West Ulster. However, the distribution of the name and several variants, whose current Irish population is about 36,000, showed a considerable range in the 19th century (**Figure 5.1D**). There are frequency peaks as expected in Ulster (Tyrone and Antrim) but also at the opposite end of the island in Munster.

5.1.1.e O'Sullivan

The O'Sullivan name displays a sharp geographic specificity to South Munster (Cork and Kerry) (**Figure 5.1E**). An anglicisation of the Irish *Ó Súileabhain* variously translated as 'grandson of the one eyed' or 'hawk eyed', an unusual personal nickname or descriptor. The eponym apparently lived in the mid-10th century, although the surname did not become prominent until after the Anglo-Norman invasion in the twelfth century. It subsequently divided into sub-branches with O'Sullivan-Beare and O'Sullivan-Mór the most prominent of these. In the 16th century, a member of the latter

branch and son of *Gilla Mochuda* O'Sullivan, through the addition of the Mac prefix, started the new hereditary surname McGillycuddy (*Mac Gilla Mochuda*). With about 38,000 bearers it is one of the most common surnames in Ireland.

5.1.1.f McCarthy

Like O'Sullivan, McCarthy shows a strong geographic bias to South Munster (**Figure 5.1F**). Indeed, the two names and many others reputedly stem from the *Eóganachta* (meaning 'people of Eoghan'), a confederation of dynasties that controlled much of Munster from the start of recorded history (*ca*. 450 AD) up until the 10th century. These claimed common descend from *Conall Corc* (*ca*. 400 AD) but took their collective name from his earlier ancestor *Eógan Mór* who lived five or six generations previously (Mac Niocaill 1972, pg 5-9). The origin of the McCarthy (*Mac Cárthaigh*) surname is much later, deriving from *Cárthach*, King of Cashel and head of the *Eóganachta* confederation who died about 1045 AD. One of the most frequent names in Ireland generally with about 24,000 bearers, it is also probably the most common example of the Irish *Mac* surname type.

5.1.1.g O'Gara and O'Hara

By comparison to the previously described surnames, O'Gara is rare with an estimated population of just 700. The name is derived from the Irish *Ó'Gadhra*, claiming descent from an early 10th century chief of an area in North Connacht. The later distribution of the O'Gara name is consistent with this historical monogenic origin (**Figure 5.1G**). The presence of the name in Donegal is reputed to represent a later migration by sea from North Connaught. The surname O'Hara (*Ó hEaghra*) has an origin in the same area, also from an early 10th century eponym (*Eaghra*). Remarkably, given the timeframe involved, there are historical sources that suggest the eponymous ancestors of both O'Hara and O'Gara were paternal uncle and nephew respectively.

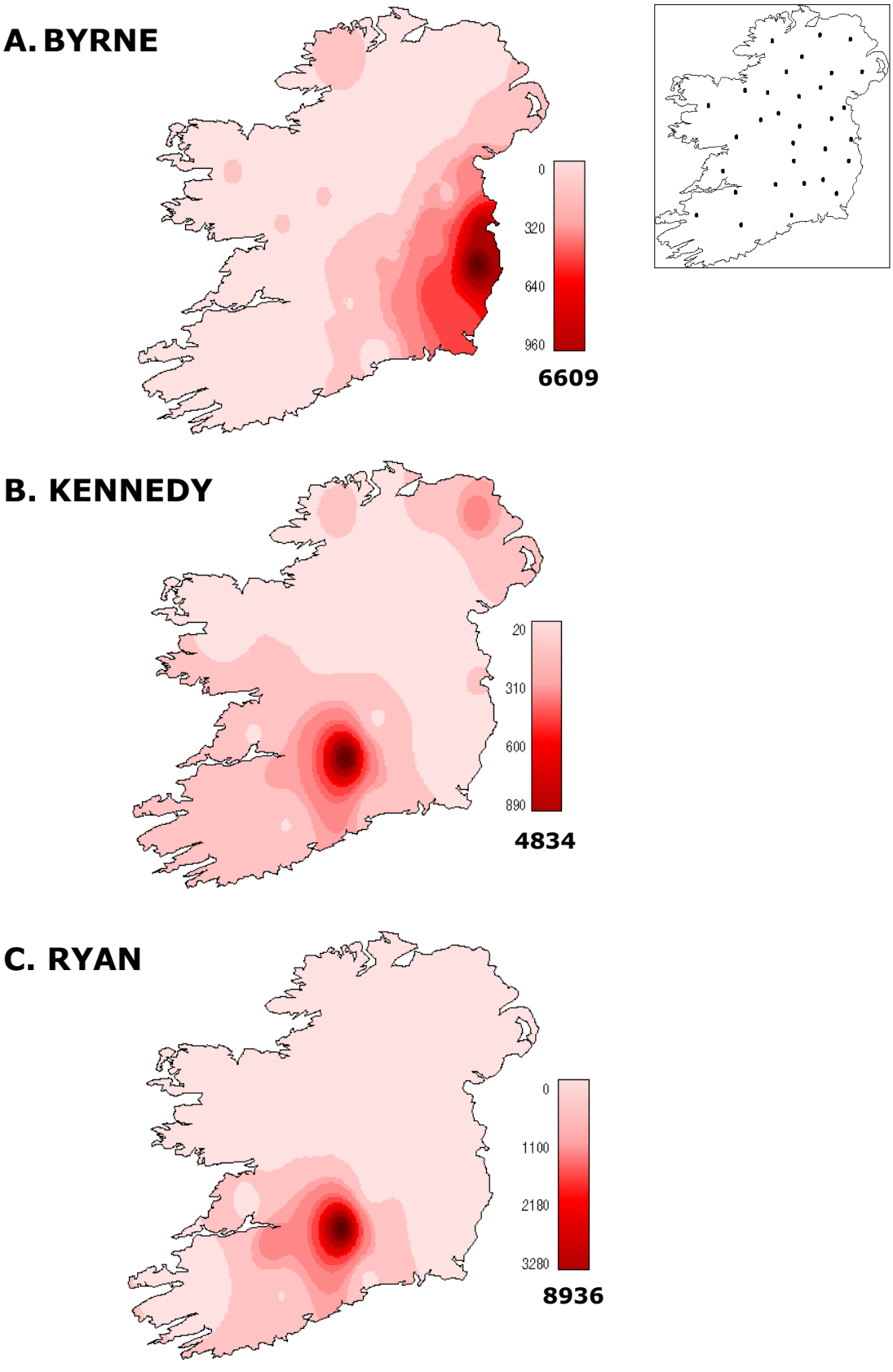


Figure 5.1 *Synthetic surface maps of Ireland showing the geographic distribution of eleven Gaelic Irish surnames in the mid-19th century.* **(Full legend on page 134)**

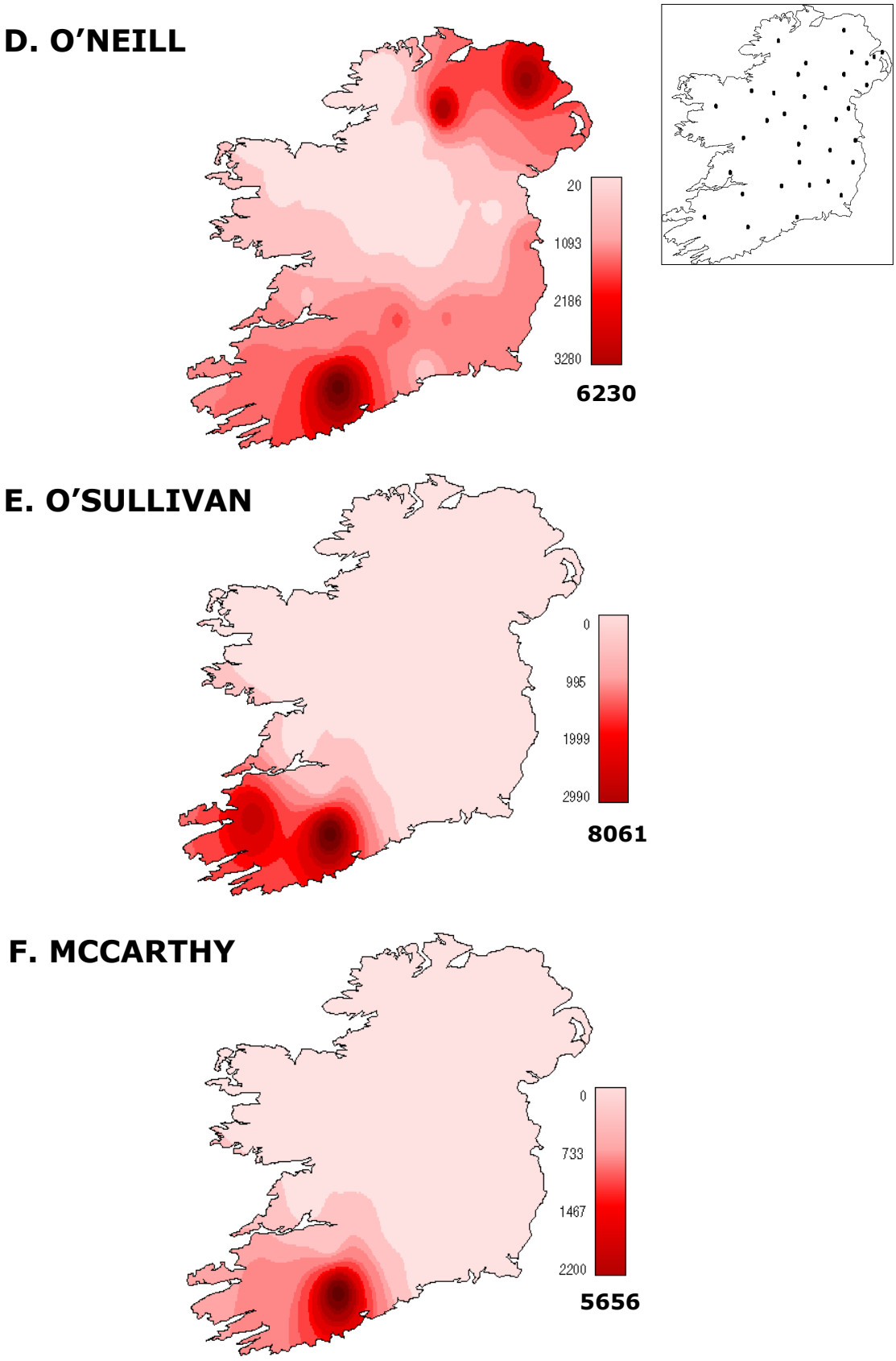


Figure 5.1 *Synthetic surface maps of Ireland showing the geographic distribution of eleven Gaelic Irish surnames in the mid-19th century.* ***(*Full legend on page 134*)***

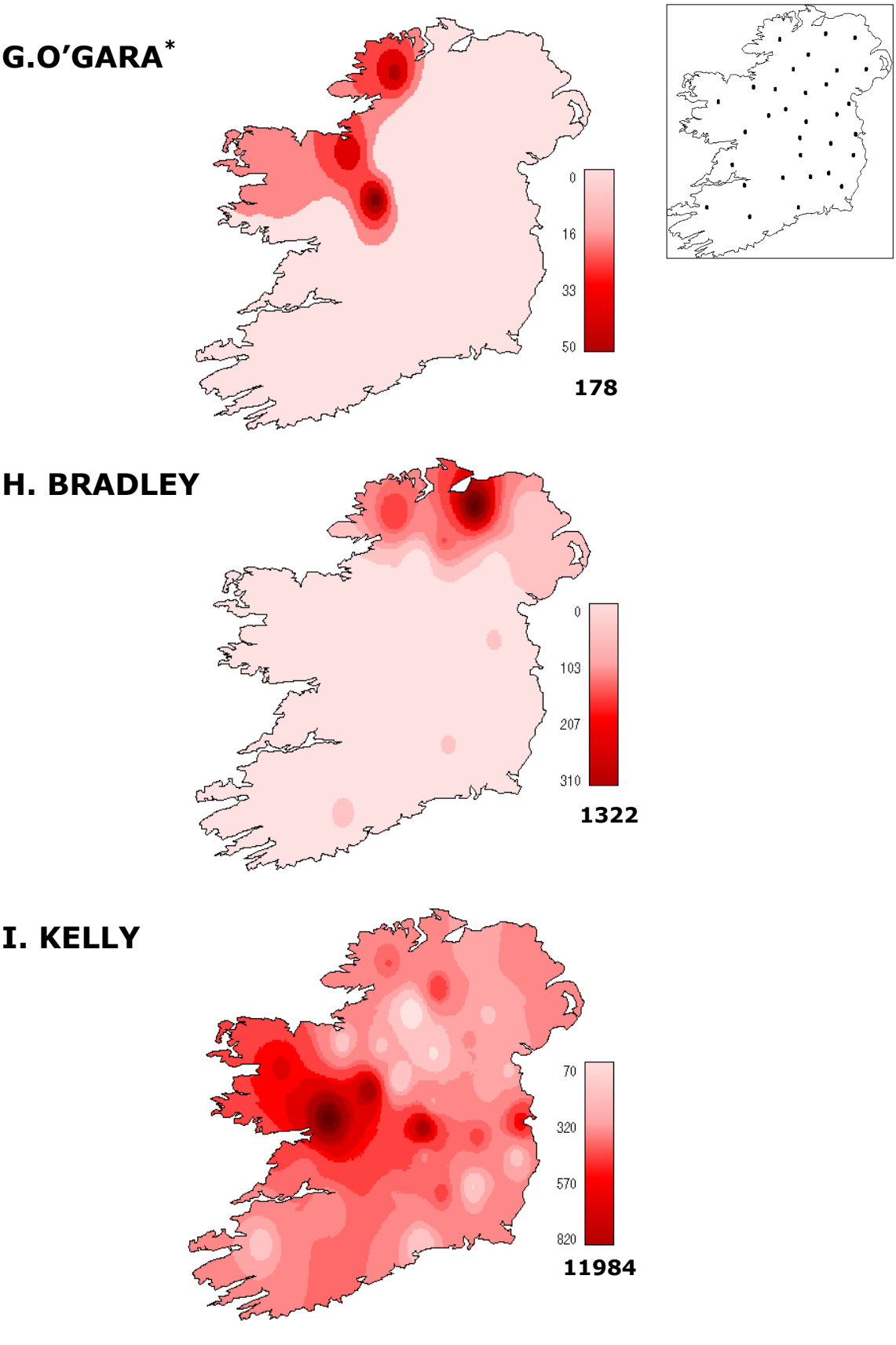


Figure 5.1 *Synthetic surface maps of Ireland showing the geographic distribution of eleven Gaelic Irish Surnames in the mid-19th century.* ***(*Full legend on page 134*)*** ** Map G is based on the number of O'Gara households only and does not include figures for the putatively related O'Hara surname.*

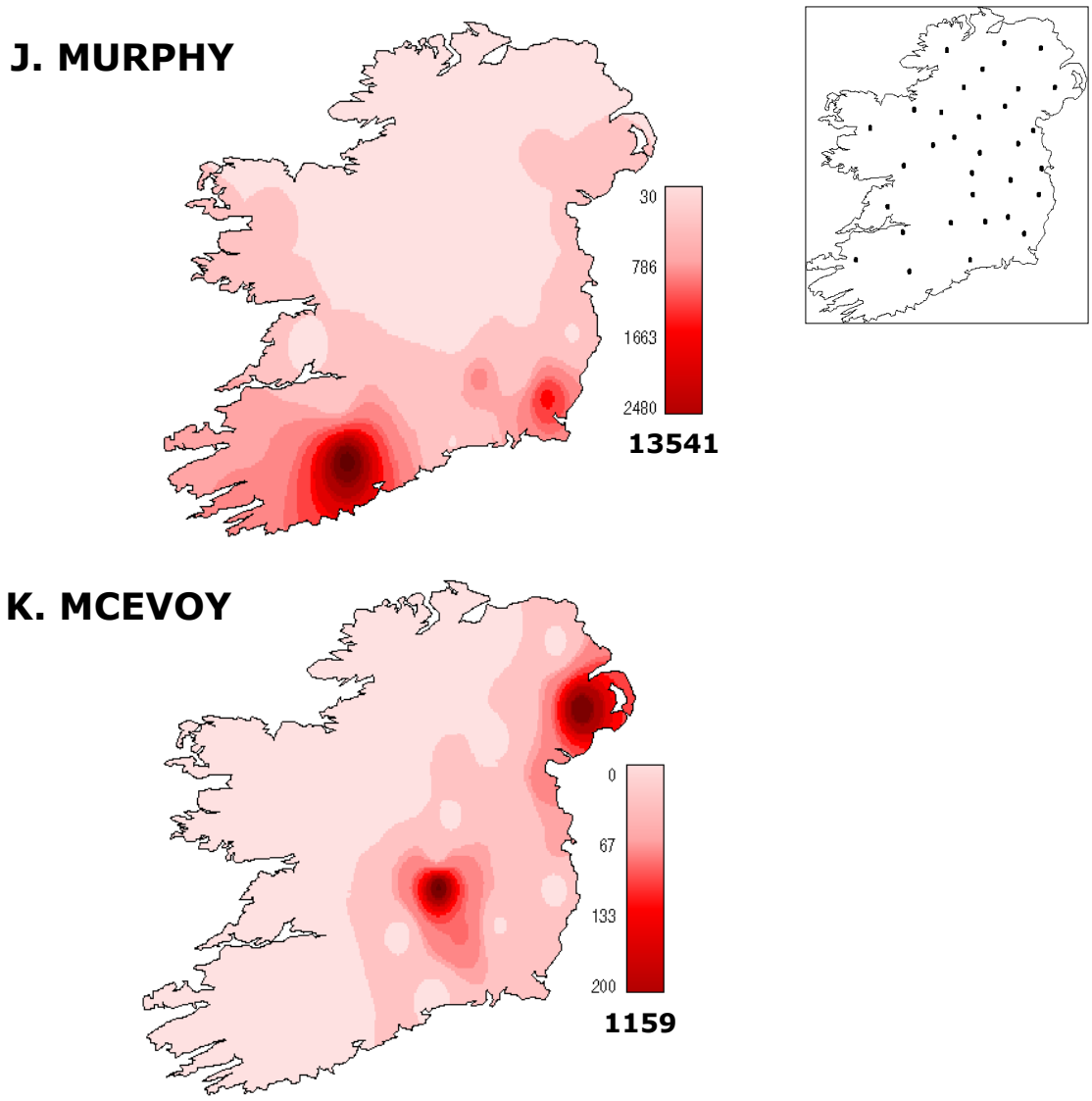


Figure 5.1 A-K (pg 131-134) *Synthetic surface maps of Ireland showing the geographic distribution of eleven Gaelic Irish surnames in the mid-19th century. Data taken from the 'Primary Valuation of Ireland' conducted by Richard Griffith between 1848-1864 (see* ***Section 4.1.5****) and accessed online at http://www.ireland.com/ancestor. The total number of households recorded with each surname (including close variants) is shown below each legend. The synthetic surface maps were generated from the county level distribution of each surname (32 county coordinates are shown inset) using the Spatial Analyst extension of Arcview (Version 3.2, Environmental Systems Research Institute Inc.). The observed values for each county were interpolated to generate 20 equal interval contours.*

5.1.1.h Bradley

Most of the estimated 7,300 Bradley surname bearers in Ireland derive from the Irish name *Ó Brollachain*, an example of anglicisation by substitution with an English surname judged to be phonetically or otherwise equivalent. In turn the English name Bradley is a toponymic derived from several villages in Northern England. The name shows an overwhelming concentration in West Ulster, especially Derry and Donegal (**Figure 5.1H**). Away from here there are minor peaks in frequency in Meath, Kilkenny and Cork. Despite the distance involved, there are suggestions that the presence in Cork is the result of a migration from Ulster. The Kilkenny branch is thought by some to be of English settler origin.

5.1.1.i Kelly

Kelly is the 2nd most common name in Ireland (60,000 bearers) and derives from the historically common personal name *Ceallach*. The name appears to have passed into hereditary surnames on multiple occasions in every part of Ireland from Antrim to Cork and Galway to Wicklow. The geographical distribution reflects this history with a generally high frequency of the name across the island. In a testament to this, dips in frequency are as notable in Kelly as peaks in other surnames (**Figure 5.1I**). The most prominent historical origin is in Galway where the 9th century *Ceallach* eponymous ancestor stemmed from the older population group/kingdom of *Uí Maine* (Hy Many).

5.1.1.j Murphy

Murphy is the most common Irish surname with about 66,000 bearers or more than 1% of the island's entire population. Analogous with Kelly, the name derives from a common Irish personal name (*Murchadha*) and has numerous foundations or a polygenic origin. As a result, and somewhat unusually, both the *Mac* and *Ó* prefixes, as *Mac Murchadha* and *Ó Murchadha*, are found in the original versions of the name. While common throughout Ireland, two peaks in frequency are found in Cork and Southeast Leinster (Wexford and Kilkenny) (**Figure 5.1J**). In the latter area the name may derive from *Murchadha*, grandfather of Dermot McMurrough, King of Leinster, who initiated the Anglo-Norman invasion of 1169 AD. However, corresponding historical detail on an eponymous Murphy ancestor in Munster is apparently absent although a common origin with the Leinster branch is sometimes suggested. In contrast to the southern half of the island, the *Mac Murchadha* version appears to underlie the origin of the Murphy name in Ulster, indicating separate origin(s) here.

5.1.1.k McEvoy

The McEvoy (or McAvoy) surname displays two geographically-distinct foci centred on County Laois in the Midlands/West Leinster and County Down in East Ulster (**Figure 5.1K**). The origins of the current surname population, of about 6,000 people, is split roughly equally between these areas. The name is believed to have arisen twice from the anglicisation of two different Irish surnames to the same English form. In the Midlands, the name derives from the Irish *Mac Fhíodhbhuidhe (*possibly meaing 'son of the woodman'), while in Ulster it is thought to have arisen from *Mac an Bheatha* ('son of life'), sometimes also anglicised to McVeigh.

5.2 MATERIALS AND METHODS

5.2.1 Sample Collection

Samples were collected as described in Chapter 4 (**Section 4.2.1**), primarily through postal requests for participation. The campaign is summarised in **Table 5.1**, which shows samples sizes, sources and DNA retrieval rates.

	SAMPLE COLLECTION					SAMPLE QUALITY			
NAME*	Samples Collected	Samples from Postal Requests	Postal Request Made	Response Rate (%)	Samples from Other Sources[1]	No DNA	Female DNA	Usable DNA	Usable DNA (%)
MURPHY	68	43	125	34.4	25	3	0	**65**	95.6
BYRNE	61	41	128	32	20	1	0	**60**	98.4
KENNEDY	68	65	150	43.3	3	0	1	**67**	98.5
RYAN	64	43	123	35	21	2	0	**62**	96.9
KELLY	60	41	169	24.3	19	4	1	**55**	91.7
O'NEILL	80	55	150	36.7	25	0	0	**80**	100
MCCARTHY	70	52	144	36.1	18	4	0	**66**	94.3
BRADLEY	31	30	91	33	1	0	1	**30**	96.8
MCEVOY [2]	54	53	127	41.7	1	4	0	**50**	92.6
O'SULLIVAN	70	54	209	25.8	16	3	0	**67**	95.7
-MCGILLYCUDDY [3]	21	NA	NA	NA	21	0	0	**21**	100
O'GARA[4]	28	NA	NA	NA	28	1	0	**27**	96.4
-GEARY [4]	9	NA	NA	NA	9	2	0	**7**	77.8
-O'HARA [4]	10	NA	NA	NA	10	0	0	**10**	100
TOTAL	**694**	**477**	**1416**	**33.7**	**217**	**24**	**3**	**667**	**96.4**

* Including trivial spelling variants

[1] Mainly collected through personal contacts and the particpation of staff and students in TCD

[2] Includes the McEvoy samples from East Ulster described in Chapter 4

[3] Sample requests distributed amongst members of the McGillycuddy Clan Association with the assistance of Richard, The McGillycuddy of the Reeks

[4] Samples for the O'Gara , O'Hara and Geary names were collected by post from individuals who had agreed to participate followed initial telephone contact. The role of Mrs. Maura O'Gara-O'Riordan in recruiting these volunteers is acknowledged

Table 5.1 *Summary of sample collection for 11 Gaelic Irish surnames. Potential subsidiary names are shown in italics below the major surname to which they putatively relate.*

A summary of the complete sample collection process for the Irish surname study is shown in **Figure 5.2**. The overall response rate at 31.8% is relatively high considering participation required appreciable effort by volunteers with no direct individual benefit. Furthermore, the high rate of success in obtaining usable male DNA (95.3%) demonstrates that unsupervised self-collection and postal return present no technical barrier to this method of collection. Overall, postal volunteer recruitment appears a feasible, if somewhat inefficient, means of assembling relatively large numbers of DNA samples from dispersed populations.

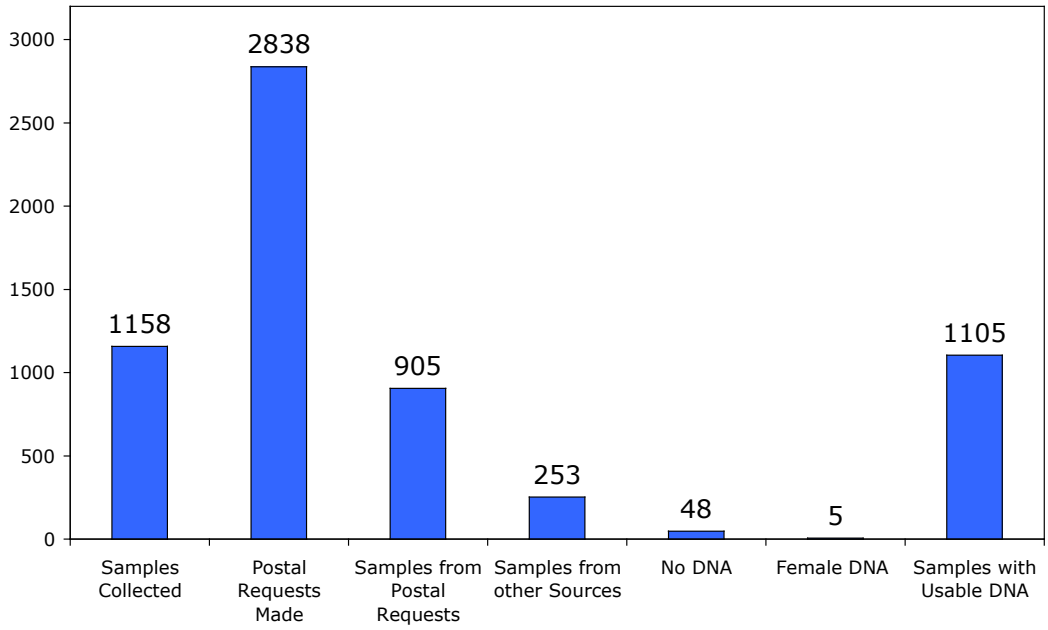


Figure 5.2 *Summary of the sample collection process undertaken for the investigation of Irish surname history.*

5.2.2 Laboratory and Analytical Methods

Each sample was genotyped for 19 Y chromosome STR markers in three multiplex PCRs and up to 6 UEP loci as described in Chapter 4 (**Sections 4.2.5** and **4.2.6**, respectively). Analytical methods including network construction, cluster size estimation and dating were also essentially carried out as outlined in **sections 4.2.7** to **4.2.11** of that Chapter.

5.3 RESULTS

5.3.1 Congruence of Paternal Ancestry and Surname

The methods used at the local level in Chapter 4 were extended to make a broader general assessment of the congruence of Gaelic Irish surnames with paternal ancestry. Using only well-sampled surnames (25 samples or greater) and including McGuinness and Donohoe (Chapter 4) yields a total sample population of 801 Y chromosomes over 13 surnames. Collectively these reflect differences in historic and geographic surname origins as well as current frequencies. Differences in the latter mean that the ratio of each sample size to the total male surname population, indicating sampling coverage, is quite variable (**Figure 5.3**) ranging from nearly 8% in O'Gara to ~0.2% of Murphy and Kelly males, the two most frequent Irish surname.

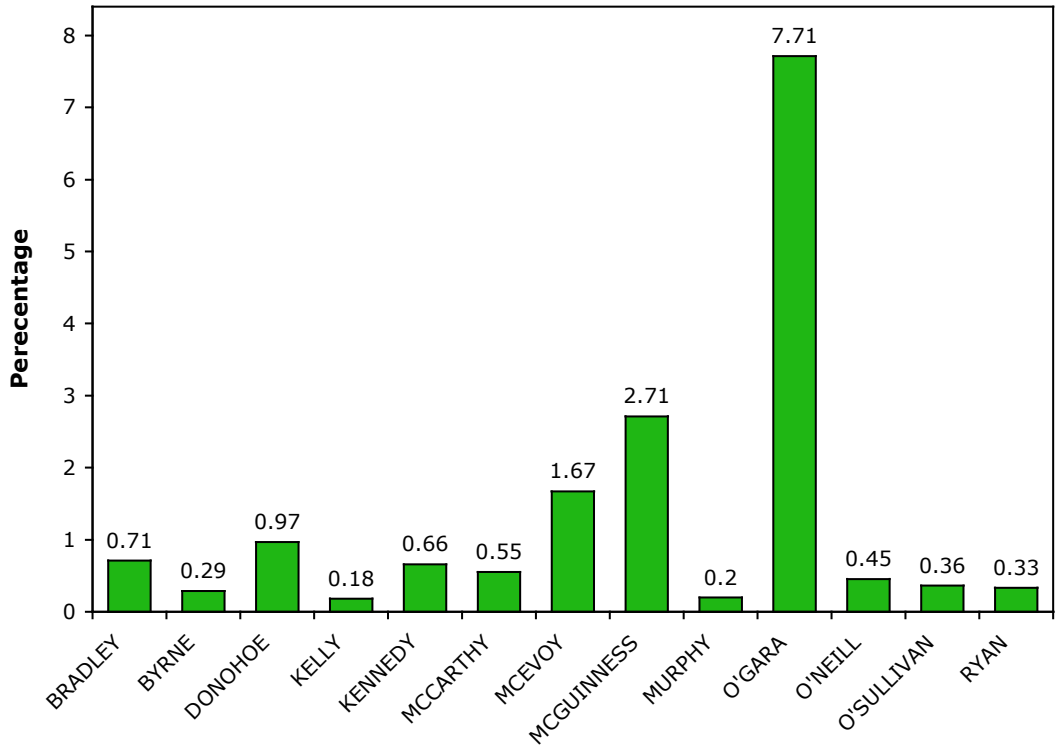


Figure 5.3 *Sample coverage of the major surnames investigated. Sample sizes are expressed as the percentage of the estimated male population of each surname in Ireland.*

Initially, the strength of any common paternal ancestry within each surname was measured through haplotype frequencies and the dependent average probability that any two Y chromosomes drawn from a defined population or surname will match. The values for each name are shown in **Figure 5.4A** alongside background levels calculated from 796 random Irish Y chromosomes, as a whole and for each province individually. Values differ widely between surnames with the lowest found in Kelly and Murphy, consistent with their polygenic origins, and the highest in putative monogenic names like Ryan and O'Sullivan. Notwithstanding this, sharing of identical haplotypes within surnames is always higher than that found in the general Irish population as whole (an average of 19-fold greater across all surnames) or against a provincial mean (average 10-fold increase), which attempts to reflect any geographic stratification (**Figure 5.4B**). Further investigation of haplotype sharing tolerating varying levels of mismatch or mutational divergence from 0 (identical haplotypes as above) to 10 repeat unit differences demonstrates that the strong relationship within surnames is only preserved over short mutational distances and thus recent timeframes, consistent with the 1,000 year old history of Irish surnames (**Figure 5.4C**).

Analysis of Molecular Variance (AMOVA) (**Section 4.2.11**) across the 13 surname populations revealed a highly significant ($p<0.00001$) 10.9% of the total variance was distributed between names, collectively confirming them as genuine Y chromosome genetic divisions. Finally, a partial mantel correlation (**Section 4.2.11**) was carried out using a binary description of surname and genetic relationships, with 1 representing an exact match of haplotype or surname and 0 a mismatch (ignoring trivial spelling differences). Despite this relatively crude division, there is a very significant positive correlation between the two variables, over and above any effect of geographic stratification ($r=0.167$, $p<0.0001$).

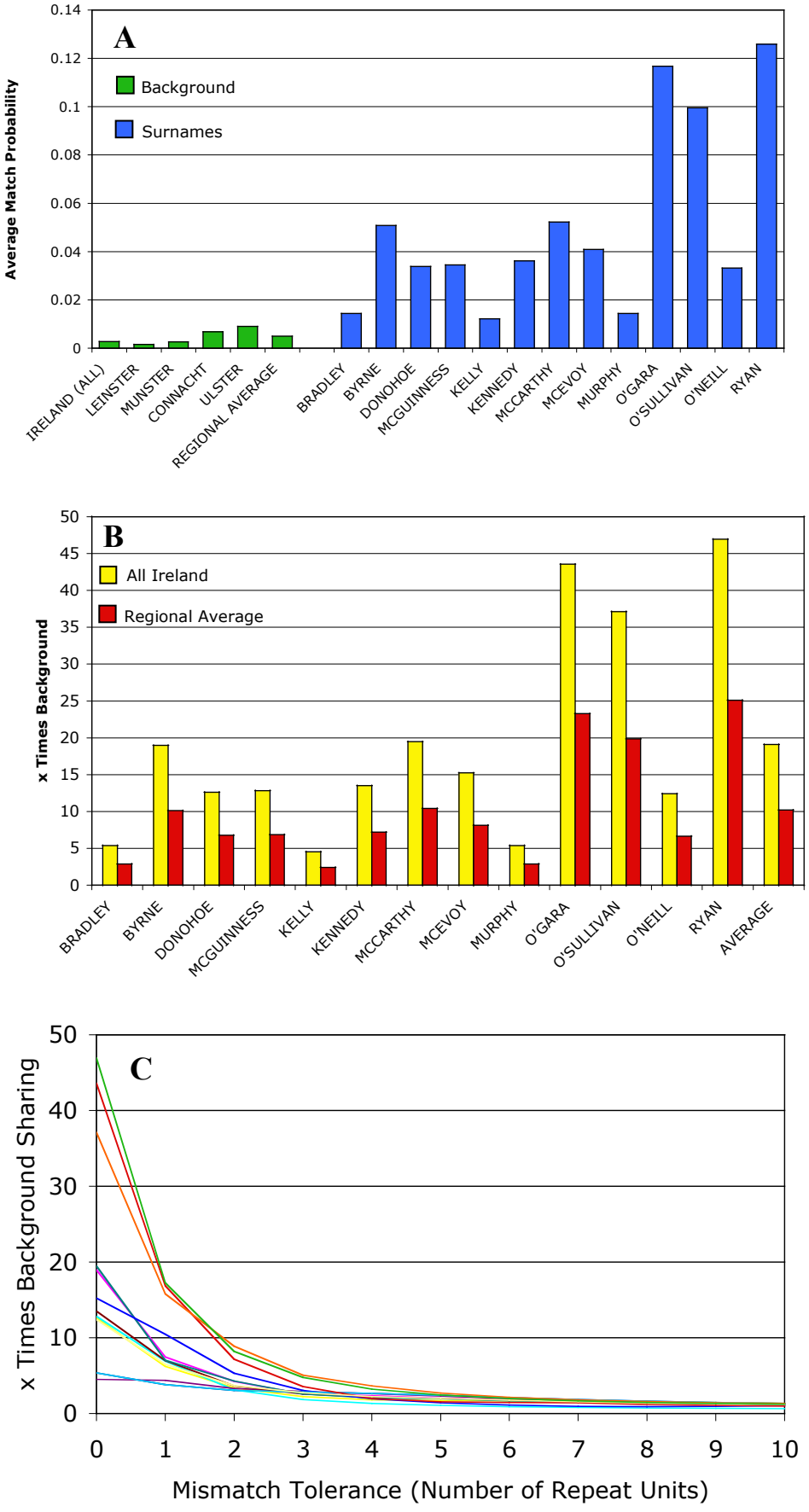


Figure 5.4 *Haplotype sharing in Irish geographic regions and surnames* ***(*Full Legend Overleaf***).

Figure 5.4 (preceding page) *Haplotype sharing in Irish geographic regions and surnames. (**A**) The average probability that two random Y chromosomes (17 STR haplotype) from any Irish region or surname will match exactly. (**B**) Probability scaled against the general Irish or average regional (to account for geographic substructure) background. All surnames have closer ancestry than general Irish groupings but the extent of this varies, reflecting differences in polygenic versus monogenic surname origins. (**C**) The close relationship within surnames (each represented by a separate line), relative to general Irish background levels, decays rapidly as the stringency for the definition of a haplotype match decreases (increased mismatch tolerance) indicating that the relationship is recent. The data for 796 Irish Y chromosomes were taken from Moore (2004) and divided into four provincial samples (sample size): Leinster (315), Munster (184), Connacht (144) and Ulster (153).*

5.3.2 Y Chromosome Diversity in Individual Surnames

Median-joining (MJ) networks (**Sections 4.2.7** and **4.2.8**) were constructed for each surname and colour coded to indicate the geographic origin of each sample, generally to provincial level. Although Dublin is part of Leinster, it is identified separately as a paternal origin here will often reflect a lost knowledge of an earlier ancestral location. Prominent lineages in each surname were identified and TMRCA estimated (see **Section 4.2.9** and **4.2.10**). These ages (including 95% CIs) and other relevant details of each cluster are summarised in **Table 5.2.**

5.3.2.a Byrne (Figure 5.6)

A predominant lineage, accounting for 31.7% of the sample population and centred on a frequent modal haplotype provides strong evidence for one early and principal founding male in the Byrne surname. The contrast to the phylogeny of a similar number of random Irish Y chromosomes is obvious (**Figure 5.5**). Both the geography and age (790 YBP) of this cluster are consistent with historical accounts suggesting a mid-11th century origin of the name in Leinster. Although there are no other obvious clusters, West Ulster (Donegal and Tyrone) samples tend to group together away from the core Leinster lineage. Similarly, samples from Connacht fall outside the major founding Y chromosome cluster suggesting that distinct geographic peaks in frequency also reflect distinct genetic origins.

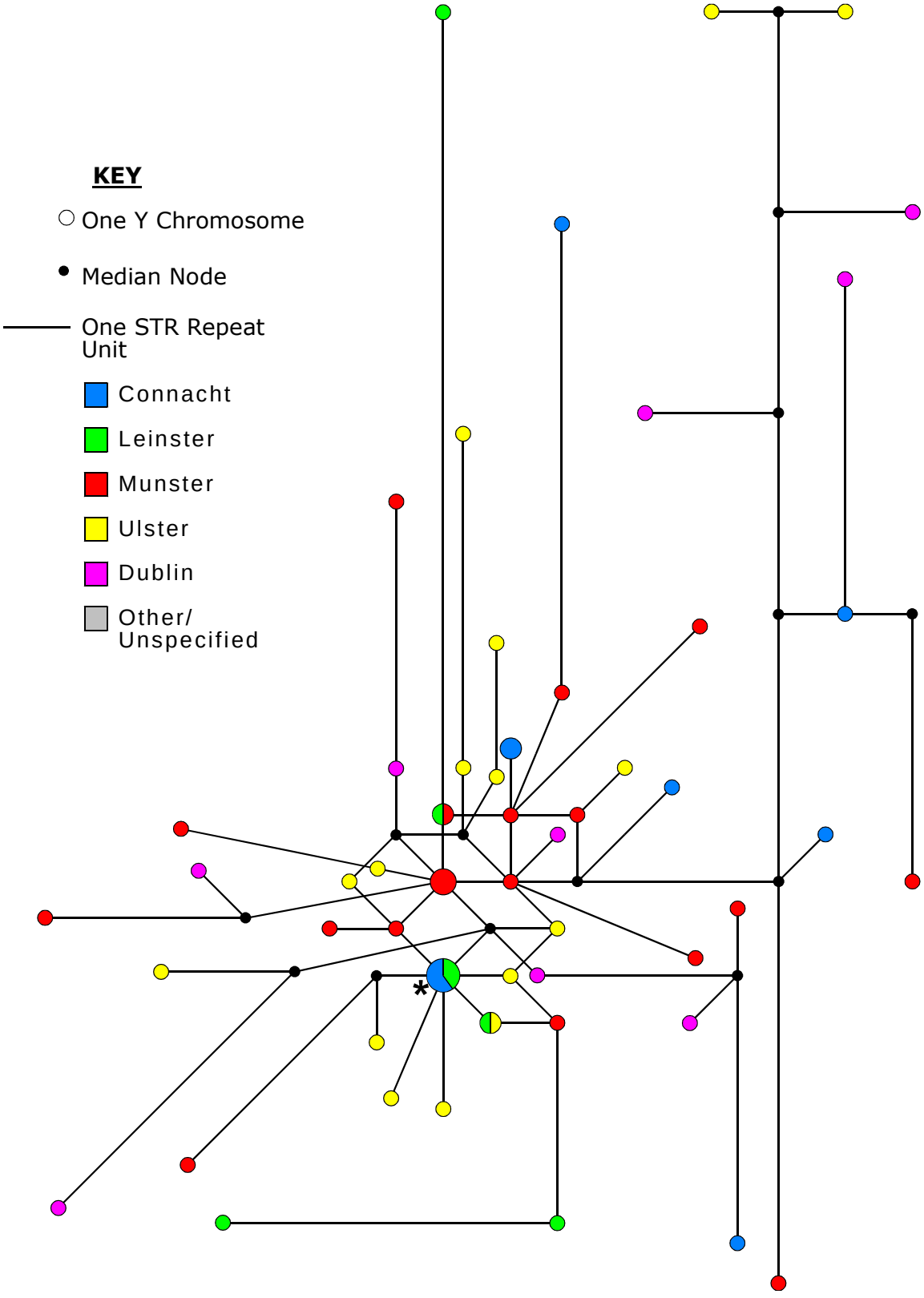


Figure 5.5 *Median-Joining (MJ) network of 62 random Irish Y chromosomes (17 STR marker haplotypes). This number is the mean sample size of the major surnames examined. There is high diversity with only 4 haplotypes repeated. The most frequent of these (*) occurs 5 times and is also the most common haplotype found in Ireland generally at 3.8% (Irish Modal Haplotype or IMH). Y chromosomes were randomly taken from 796 seventeen-marker haplotypes described by Moore (2004).*

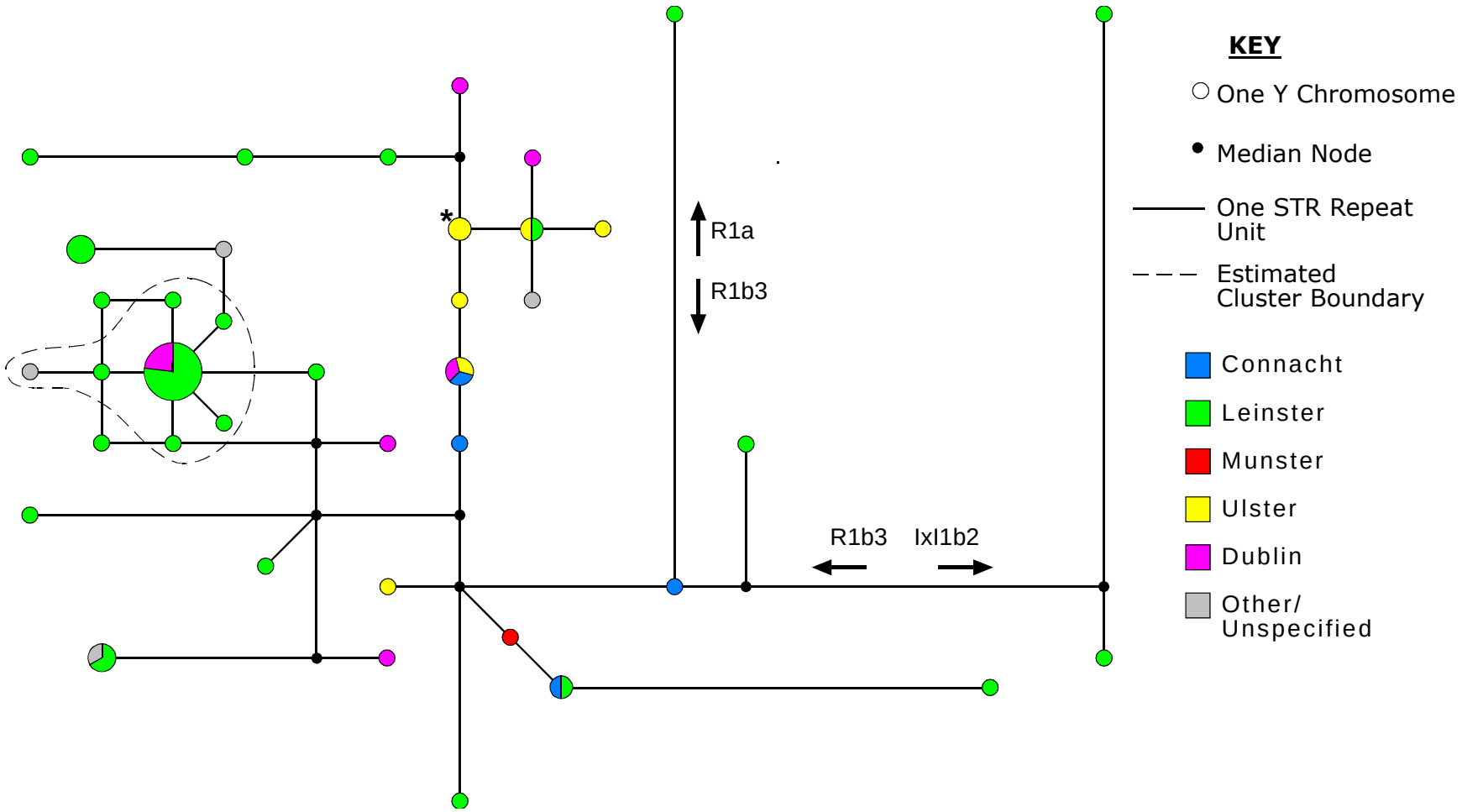


Figure 5.6 *MJ network of Y chromosomes in the Byrne surname. The 60 samples are represented across 39 haplotypes. A prominent diversified lineage (cluster) (n=19), centred on the modal haplotype (n=13), supports an early single ancestor for the name in Leinster. The distinction of samples from Connacht and Ulster from this lineage suggests multiple distinct origins of the name in these areas. *= IMH.*

Lineage	Haplogroup	Geography	Cluster Size		Time to Most Recent Common Ancestor (TMRCA) Years Before Present (YBP)						
			n	% Surname	Evolutionary Mutation Rate [A]				Directly Observed Mutation Rate [B]		
					95 +	CI[C]	95 -		95 +	CI[C]	95 -
Bradley B	R1b3	West Ulster	11	36	2740	1300 **1710** 2250	1240	95 + CI[D] 95 -	1500	560 **720** 930	470
Byrne	R1b3	Leinster	19	31.7	1260	620 **790** 990	570		690	260 **330** 420	220
Kennedy	R1b3	North Munster	24	35.8	2280	1180 **1420** 1720	1030		1250	500 **600** 710	390
McCarthy A	R1b3	South Munster	17	25.8	600	270 **380** 520	270		330	110 **160** 210	100
McCarthy B	R1b3	South Munster	10	15	2050	960 **1280** 1710	930		1120	410 **540** 710	350
McEvoy A	R1b3	Leinster/ Midlands	18	36	3420	1760 **2130** 2610	1550		1870	740 **900** 1090	590
McEvoy B	I1c	East Ulster	17	34	2210	1100 **1380** 1710	1000		1210	470 **580** 720	380
Murphy	R1b3	Munster	16	24.6	2780	1400 **1730** 2140	1260		1520	590 **730** 900	480
O'Neill	R1b3	Ulster	26	32.5	1970	1020 **1230** 1470	890		1080	430 **520** 620	340
O'Sullivan	R1b3	South Munster	35	52.2	1860	990 **1160** 1350	840		1020	420 **490** 570	320
O'Gara / O'Hara	R1b3	North Connacht	22	50	2640	1380 **1650** 1990	1200		1450	580 **690** 840	450
Ryan	R1b3	North Munster	33	53.2	1350	700 **840** 1000	610		740	290 **350** 420	230
'Eóganachta'*	R1b3	South Munster	60	NA	1940	1080 **1210** 1360	880		1060	450 **510** 570	330

[A] Mutation rate of 2.76 x 10^{-5} per locus per year (Zhivotovsky et al. 2004)
[B] Mutation rate of 6.57 x 10^{-5} per locus per year; calculated from the data presented in Dupuy et al. (2004)
[C] 95% Confidence Intervals based on mutation rate uncertainty
[D] 95% Confidence Intervals associated with sampling variance
* Composed of McCarthy-B, O'Sullivan, Bradley-A, and Donohoe-C (Chapter 4) lineages

Table 5.2 *Summary of major Y chromosome surname lineages, including TMRCA estimates and their 95% confidence intervals (CI) associated with mutation rate uncertainty (blue) and sampling variance (red).*

5.3.2.b Kennedy (Figure 5.7)

Analogous with Byrne, a single and diverse lineage, centred on the surname modal haplotype (marked **A** in **Figure 5.7**), is prominent amongst Kennedy Y chromosomes and constitutes 35.8% of the sample population. Nearly all members of this cluster trace an origin to North Munster (principally Tipperary), coincident with the historical place of origin for the eponymous Kennedy ancestor. The TMRCA date of 1420 YBP is also not inconsistent with his putative late 11th century timeframe considering the uncertainties involved in such estimates (see **Table 5.2**). However, even in the Munster area, there has been introgression of multiple additional men including one haplotype from Cork/Kerry only (**B**), which is repeated five times indicating some antiquity. Meanwhile, three samples from Connacht are identical (**C**) and distinct from the main founding lineage supporting a separate origin here, while the name also appears to have multiple diverse origins in Ulster.

5.3.2.c Ryan (Figure 5.8)

The strong signature of single expanded ('Star like') lineage is obvious in the Ryan surname and encompasses over half (53.2%) of the sample population. The ancestral (modal) haplotype alone of the cluster (**A** in **Figure 5.8**) accounts for ~35% of the 62 samples. Unsurprisingly, the geography of this lineage is predominantly North Munster (Tipperary) in line with the highly focused distribution of the name. Although there is no specific historical detail on an eponymous ancestor, the TMRCA of 840 YBP is consistent with the time depth of early Irish surname foundation. A minor (in modern frequency terms) separate origin in the Carlow area of Leinster (from the Irish *Ó Riain* as opposed to *Ó Maolriain* in Tipperary) is tentatively supported with the observation of two repeated and closely related haplotypes (**B**), whose members have an origin in this area.

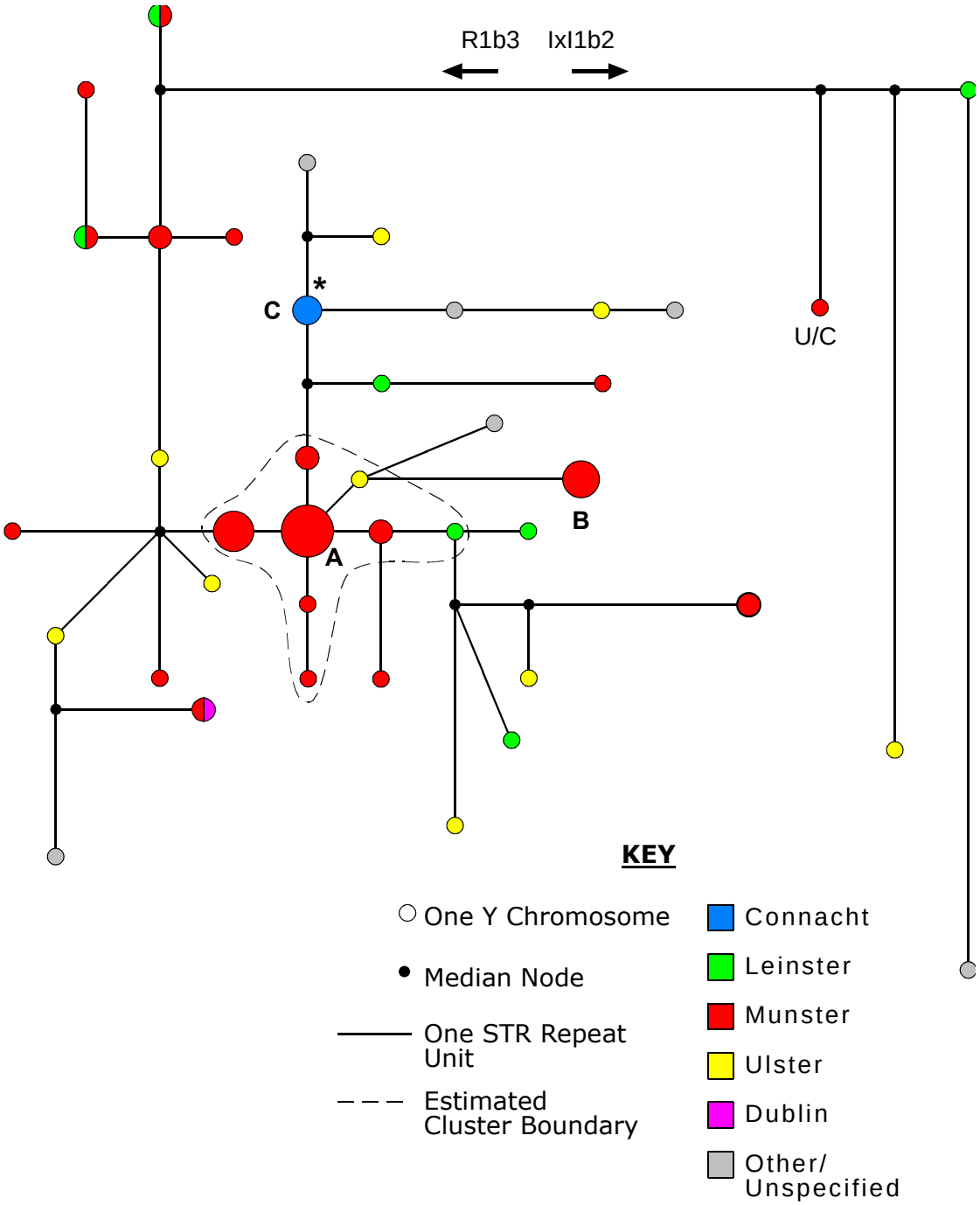


Figure 5.7 *MJ network of Y chromosomes in the Kennedy surname. The 67 individuals are represented by 40 distinct haplotypes. A major lineage (n=24), centred on the modal Kennedy haplotype (**A**) (n=10) is proposed to represent the signature of the eponymous founder of the surname. However, the impact of other men both in Munster and other regions is apparent. Some time depth to these events is suggested by the relatively high frequency of haplotype **B** in South Munster (n=5) and **C** from Galway (n=3). *= IMH. U/C= Unclassified Haplogroup.*

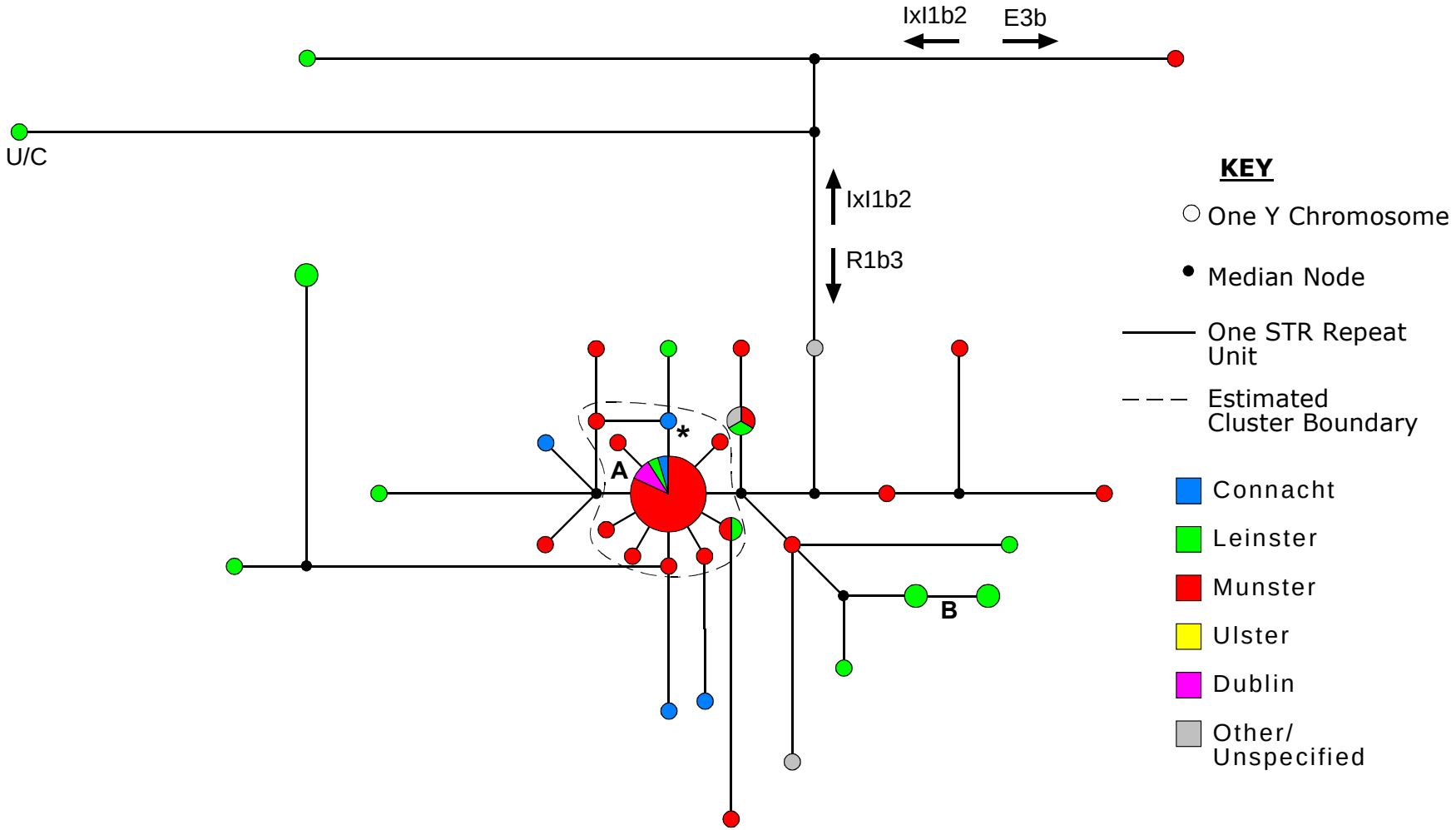


Figure 5.8 *MJ network of Y chromosomes in the Ryan Surname, showing the 35 haplotypes found in 62 individuals. A large diversified single lineage accounts for a majority (53.2%) of the sample population with the ancestral haplotype (**A**) alone recorded 22 times. The TMRCA and geography of this cluster is consistent with a single early founding male in the North Munster area. A separate origin of Ryan in South Leinster from the Irish Ó Riain (rather than Ó Maoilriain in Munster) is tentatively supported by the presence of repeated and related haplotypes (**B**) from Carlow and surrounding counties. U/C= Unclassified Haplogroup. *=IMH.*

5.3.2.d O'Neill (Figure 5.9)

Once again, evidence for a single prominent founder is apparent with a frequent lineage (32.5% of the sample population) centred on the O'Neill modal haplotype. From a geographic perspective, this cluster is dominated by Ulster (particularly Tyrone), the historical place of origin, with proximate counties in North Connacht also represented. However, the modal haplotype also contains Y chromosomes from several Munster O'Neills. This is somewhat surprising given the distance involved and additional indications of multiple independent origins of the name in Munster. The TMRCA of the major lineage at 1230 YBP is consistent with the timeframe of the O'Neill eponym (*Niall Glún Dubh*, died 919 AD). Nonetheless, it is clear from the numerous other haplotypes, in Ulster alone, that introgression by other males has been a common occurrence. The high frequency of the Irish Modal Haplotype (IMH) in Northwest Ireland is proposed to represent the biological legacy of the hegemony of the *Uí Néill* ruling dynasty in the 2nd half of the first millennium AD (Moore 2004). However, despite reputed descent from the 5th century AD founder of this group (*Niall Noígiallach*), the ancestral haplotype of the O'Neill surname is clearly distinct from the IMH.

5.3.2.e O'Sullivan (Figure 5.10)

The familiar prominent single lineage is especially obvious amongst the Y chromosomes of O'Sullivan surname bearers (**Figure 5.10A**). The ancestral haplotype alone is repeated 21 times in 67 samples, while the entire lineage constitutes 52.2% of the surname population. The TMRCA of the cluster (1160 YBP) is consistent with the early surname foundation period. The McGillycuddy surname is historically reputed to have arisen from a 16th century member of the O'Sullivan family. A MJ network of 21 McGillycuddy Y chromosomes (**Figure 5.10B)** is consistent with a single recent

common origin for most of these samples. Furthermore, a single network of the McGillycuddy and O'Sullivan surnames together (**Figure 5.10C)** shows the modes (presumed ancestral haplotypes) of both names are identical and there is consequently tight phylogenetic coherence between the founding lineages, consistent with the historically suggested common origin.

The unusual nickname from which O'Sullivan derives together with the specific geographic distribution of the name, suggest that the surname arose only once, providing the opportunity to estimate a general introgression rate less affected by the confusion of multiple 'legitimate' founders or anglicisation events. Assuming all those outside the main cluster (32 individuals or 47.8% across 28 haplotypes) represent the input of post-foundation males over 30 generations of the surname history (generation time of 35 years since 950AD), then the rate of decay from the founding Y chromosome is 2.14% of the surname population per generation. (Decay rate $=1-\sqrt[n]{x}$, from a standard compound calculation formula, where n is number of generations and x is the percentage still derived from the original founder). However, this only corresponds to an event rate if repeated haplotypes were introduced multiple independent times. Rescaling under the assumption that each haplotype represents a separate event the rate falls slightly to 1.65% per generation.

5.3.2.f McCarthy (Figure 5.11)

Although also a prominent and frequent surname of the Cork/Kerry area, the McCarthy Y chromosome phylogeny is quite distinct from that of O'Sullivan. There is an obvious modal type (**A** in **Figure 5.11**) but this has little close subsidiary diversity either in the number or frequency of one-step neighbour haplotypes. Those that are observed are defined by the relatively rapidly mutating DYS439 and DYS390 loci (see **Section 4.2.8**). Consistent with these observations, the cluster TMRCA at 380 YBP is younger

than those in other names and no confidence interval (**Table 5.2**) stretches to the putative 11th century foundation, despite the lineage accounting for 25.8% of the sample population. A second diversified lineage (15% of the total) based on a less frequent central haplotype (**B**) is older (1280 YBP) than McCarthy-A but similar to other founding surname lineages Additional small potential clusters centred on haplotypes (**C**) and (**D**) are indicative of further antiquity to the McCarthy diversity. Notwithstanding one of the strongest cases for a single historical ancestor, the paternal history of McCarthy is complex with indications that the most likely founding lineage is no longer the most frequent in the surname population.

5.3.2.g O'Gara and O'Hara (Figure 5.12)

While O'Gara has the smallest sample size, the low general frequency of the name means that it enjoys the best sampling ratio of any surname under study (**Figure 5.3**). In addition, all O'Gara and O'Hara samples are of no known relation to each other. The work of Mrs. Maura O'Gara-O'Riordan in achieving this is gratefully acknowledged. Early sources name single ancestors, living in the 10th century, for both the O'Gara and O'Hara surnames. They also indicate a common paternal ancestor for these founders in the previous two generations. In full agreement with this, there is evidence for major single founders in each name with both lineages displaying a common ancestral haplotype (**A** in **Figure 5.12)**. Suggestions of a common origin of O'Gara with the Munster name Geary are refuted as the modal Geary Y chromosome (**B**) is distinct from the O'Gara/O'Hara ancestral haplotype. As with all names, the impact of additional male introgression is evident. One of these lineages (**C**) falls into the R1a haplogroup, an otherwise rare grouping in Ireland (Hill et al. 2000; Moore 2004). A single haplotype (**D**) encompasses all O'Gara samples from Donegal suggesting a relatively recent and distinct origin of the name here.

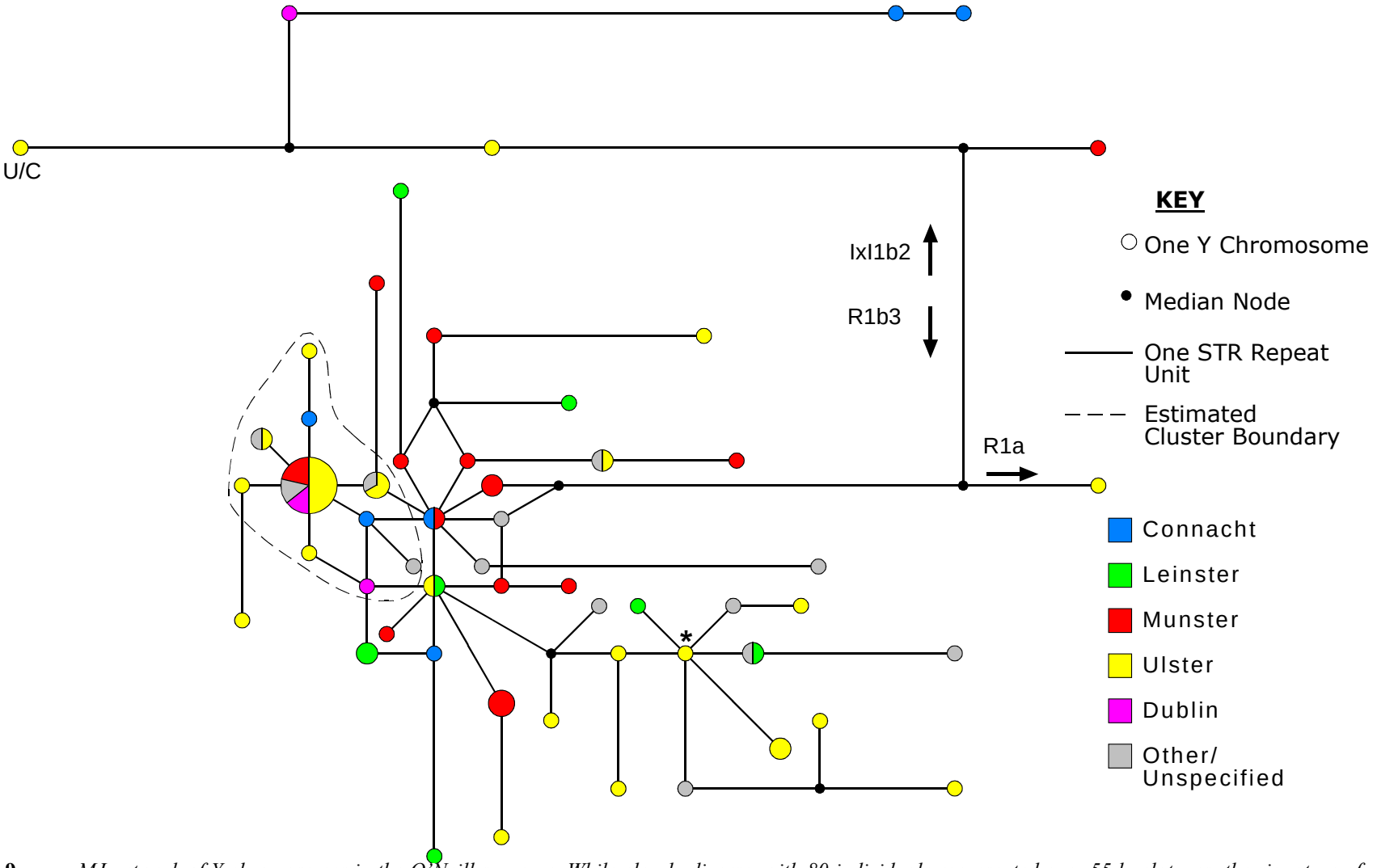


Figure 5.9 *MJ network of Y chromosomes in the O'Neill surname. While clearly diverse, with 80 individuals represented over 55 haplotypes, the signature of a major single foundation is nonetheless evident with a large cluster (n=26) centred on the modal haplotype (n=14). Although predominantly of Ulster provenance, the mode of this lineage also includes Munster O'Neills despite evidence of multiple independent foundations in this region. The main ancestral lineage is also distinct from the proposed Y chromosome legacy (*IMH) of the Uí Néill dynastic hegemony despite reputed decent from this grouping. U/C=Unclassified haplogroup.*

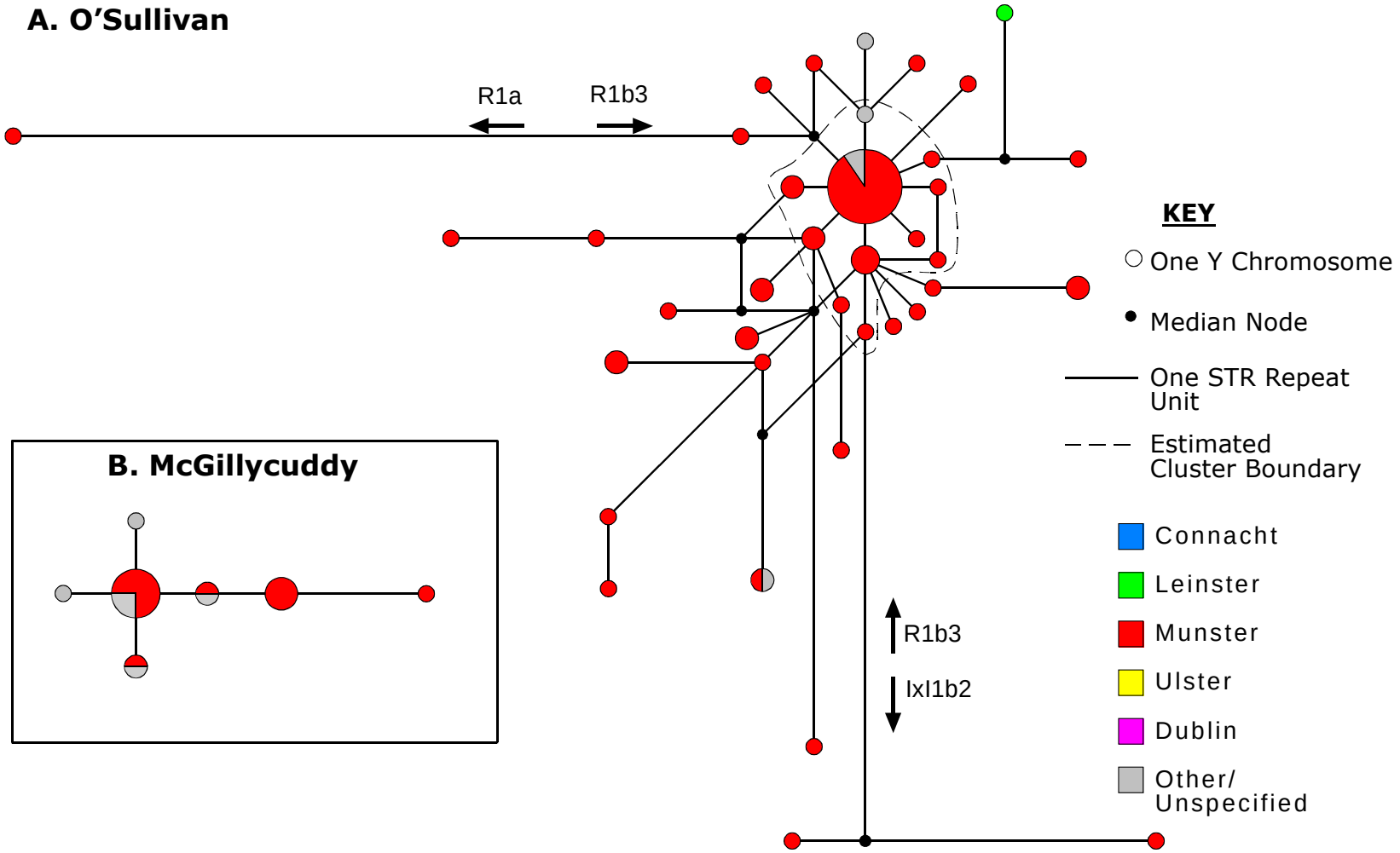


Figure 5.10 *MJ networks of Y chromosomes in the O'Sullivan and McGillycuddy surnames. (**A**) The network of 67 O'Sullivan Y chromosomes shows 38 distinct haplotypes with a prominent single lineage (n=35), centred on an obviously modal haplotype (n=21), supporting a monogenic early origin for the surname. (**B**) A network of 21 Y chromosomes from men with the McGillycuddy surname is also largely consistent with a historically suggested single 16th century founder.* **(Figure and legend continued overleaf)**

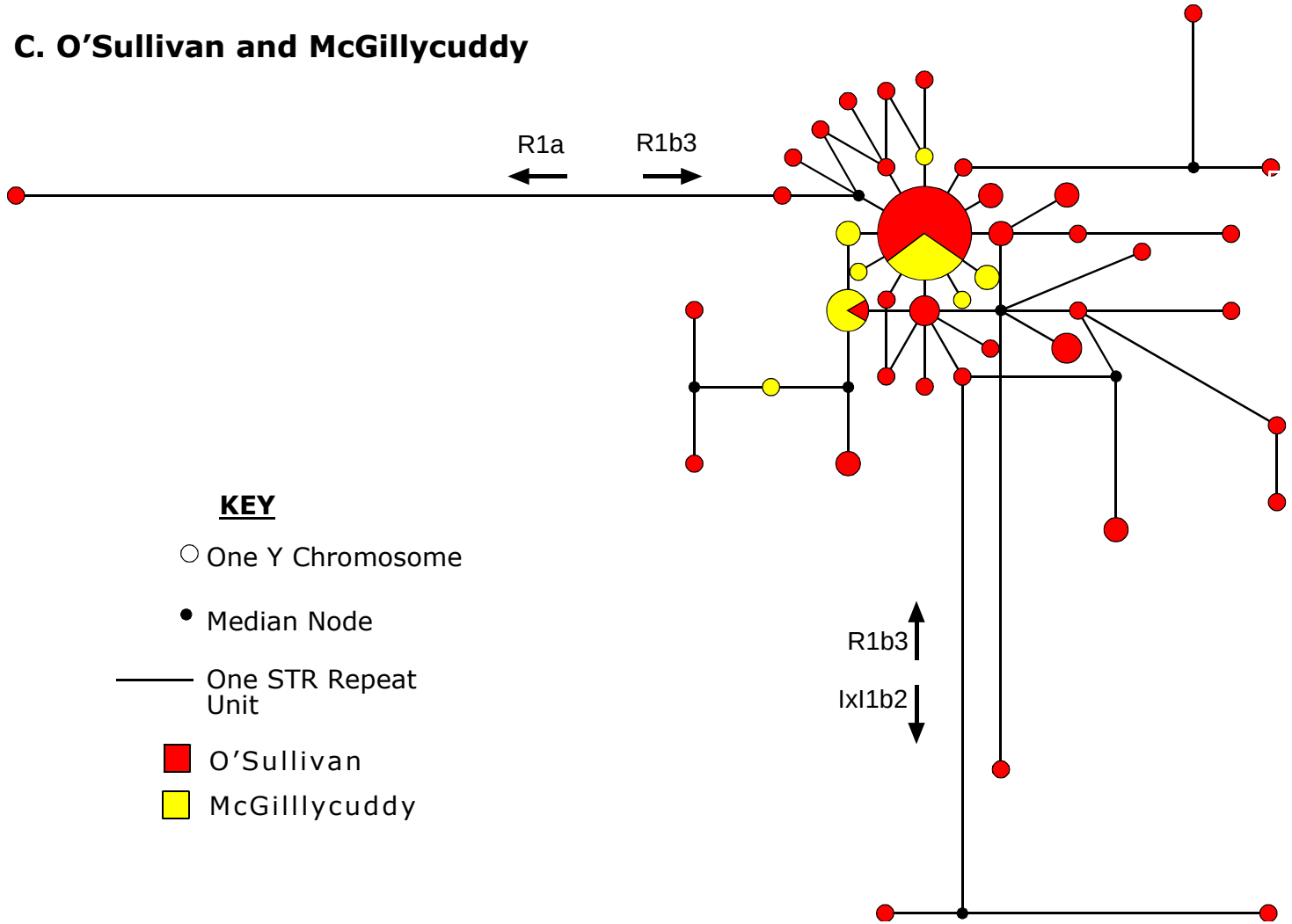


Figure 5.10 (Continued from pg 153) *MJ networks of Y chromosomes in the O'Sullivan and McGillycuddy surnames.* ***(C)*** *A joint network of O'Sullivan and McGillycuddy Y chromosomes supports the putative origin of the latter name from a member of the O'Sullivan family with tight phylogenetic coherence, including a shared ancestral (modal) haplotype, between the main lineages in each name.*

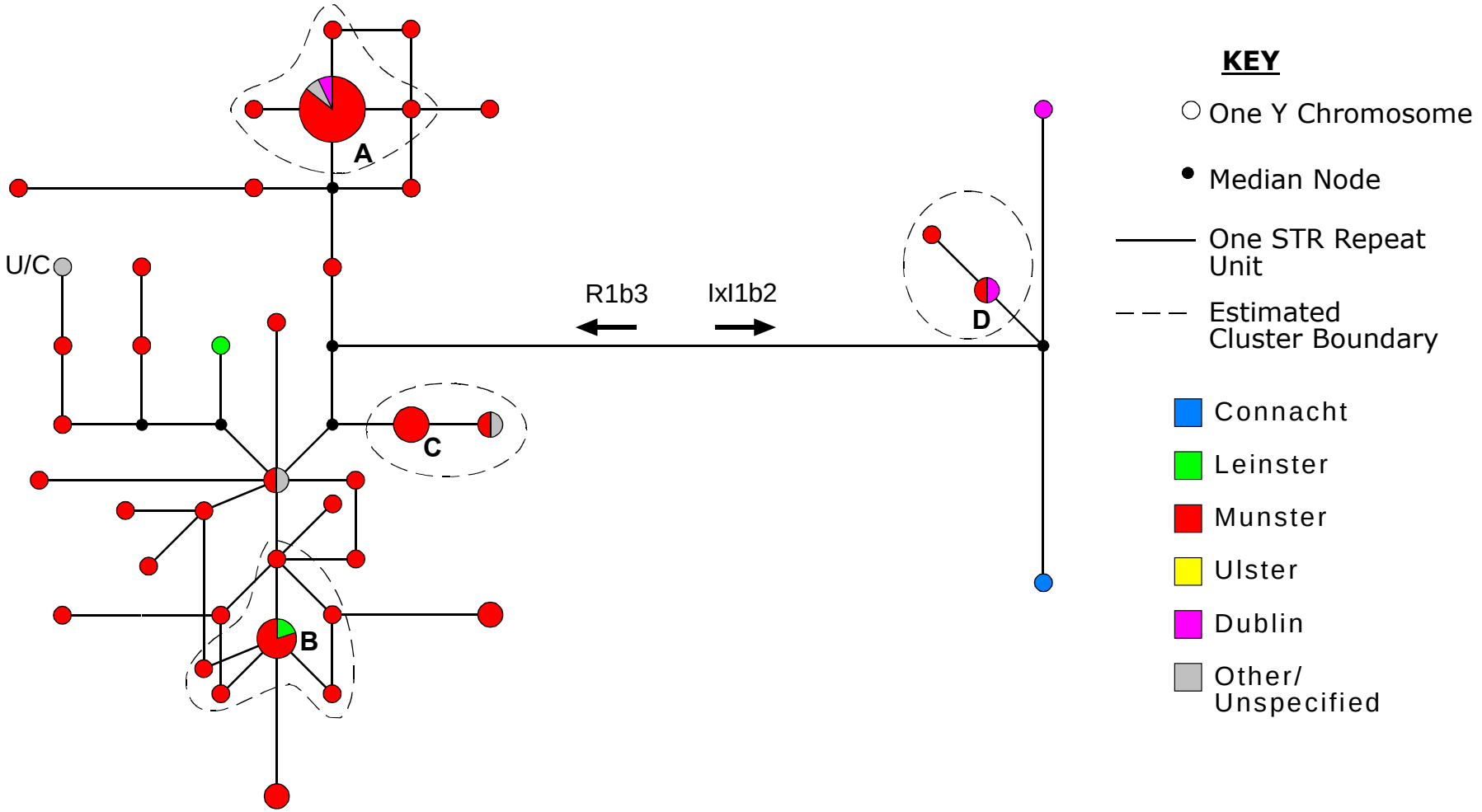


Figure 5.11 *MJ network of Y chromosomes in the McCarthy surname. The 41 haplotypes amongst 66 samples indicate a complex history. The diversity (and thus TMRCA) of the largest and most obvious lineage (n=17), centred on the modal haplotype (**A**) (n=14), appears inconsistent with early surname foundation. A second, less frequent lineage (n=10) based on the ancestral haplotype (**B**) (n=5) is older and may represent the legacy of the earliest foundation. Further small potential clusters (**C** and **D**) indicate additional antiquity to the diversity in the McCarthy name. U/C= Unclassified Haplogroup.*

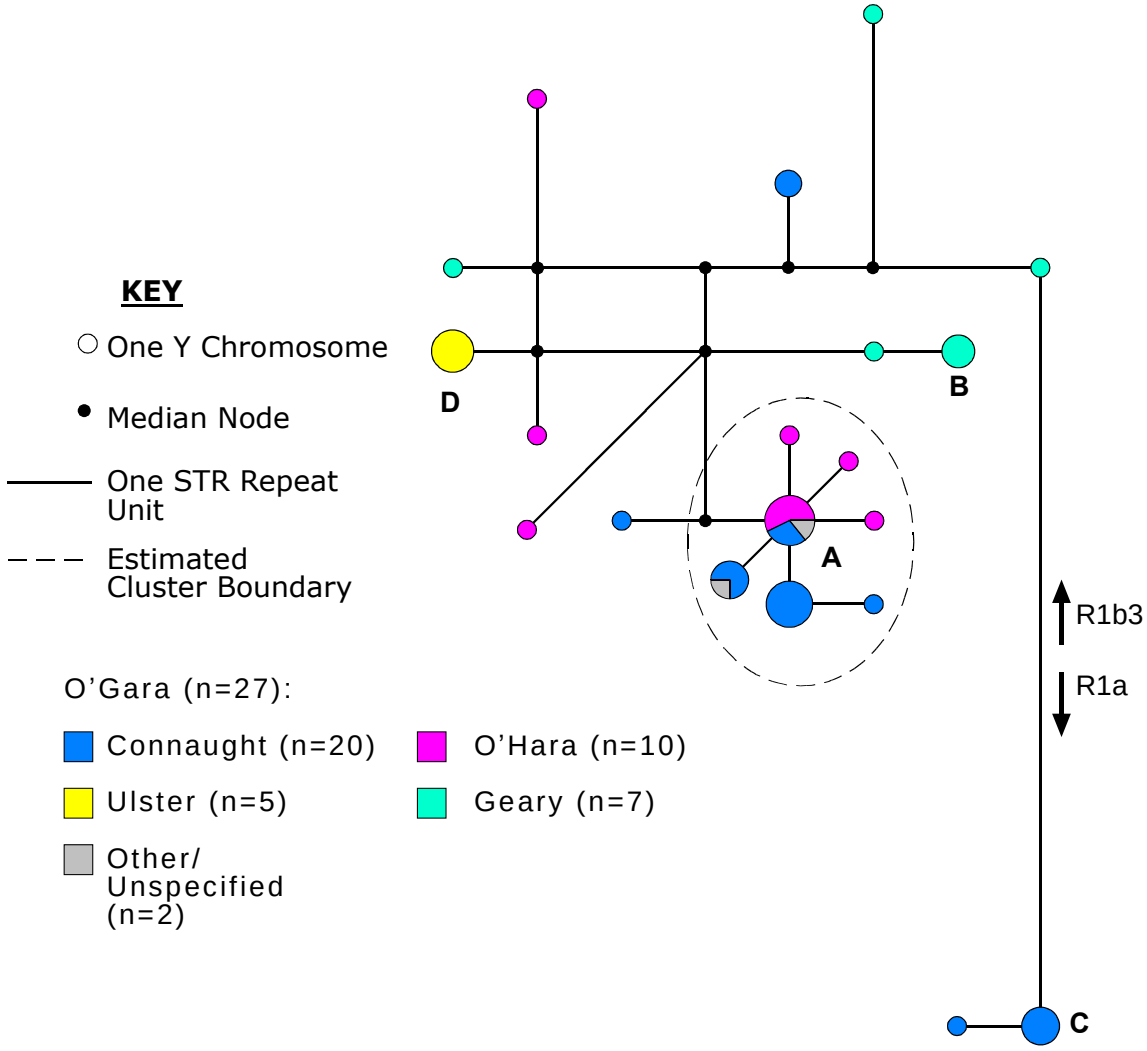


Figure 5.12 *MJ network of Y chromosomes in the O'Gara, O'Hara and Geary surnames. Suggestions of single early founders for the O'Hara and O'Gara names, as well as the close paternal relationship of these progenitors, appear to be confirmed by the observation of major lineages in each name centred on a shared ancestral haplotype (**A**). In contrast, a common origin of O'Gara and Geary is refuted with the observation that the latter's modal type (**B**) is distinct from that of O'Gara/O'Hara. Unusually, another O'Gara lineage (**C**) falls within the otherwise rare R1a haplogroup, while the occurrence of a single haplotype (**D**) in all O'Gara samples from Donegal points to a separate and recent origin of the name here.*

5.3.2.h Bradley (Figure 5.13)

The Bradley surname is predominantly associated with West Ulster, with additional small peaks in frequency found in Kilkenny and Cork. The phylogeography of Bradley Y chromosomes supports the genetic distinctiveness of the name in different areas implying multiple independent origins. While the modal type (**A** in **Figure 5.13**) is associated with Cork, a slightly less frequent haplotype (**B**) forms the core of a more numerous diversified lineage in Ulster. The TMRCA of this was estimated at 1710 YBP, somewhat older than other surnames but still not inconsistent with the surname foundation period. Finally, the IxI1b2 haplogroup of the modal Leinster type (**C**) is consistent with suggestions of a relatively recent origin from English settlers in the Kilkenny area. IxI1b2 is found at higher frequencies in England (Capelli et al. 2003) and is uncharacteristic of most Gaelic surnames (Hill et al. 2000; present study).

5.3.2.i Kelly (Figure 5.14)

The phylogenetic relationship of Kelly Y chromosomes offers a refreshing contrast to most of the surnames examined thus far, which have shown evidence for a limited number of founding males. In the Kelly surname, however, few haplotypes are repeated and there is little evidence of genuine clusters, indicated by geographic coherence. Indeed, the network displays *prima facie* greater similarity to a random sample of Irish Y chromosomes (**Figure 5.5**) than it does to most other surname populations. The large diversity and lack of any outstanding single founder supports the proposition that the historically common personal name from which Kelly derives (*Ceallach*) passed into otherwise independent hereditary surnames on numerous occasions. However, the faint echo of some common paternal ancestry still exists with haplotype sharing within the name over two-fold greater than background regional levels (**Section 5.3.1** and **Figure 5.4B**).

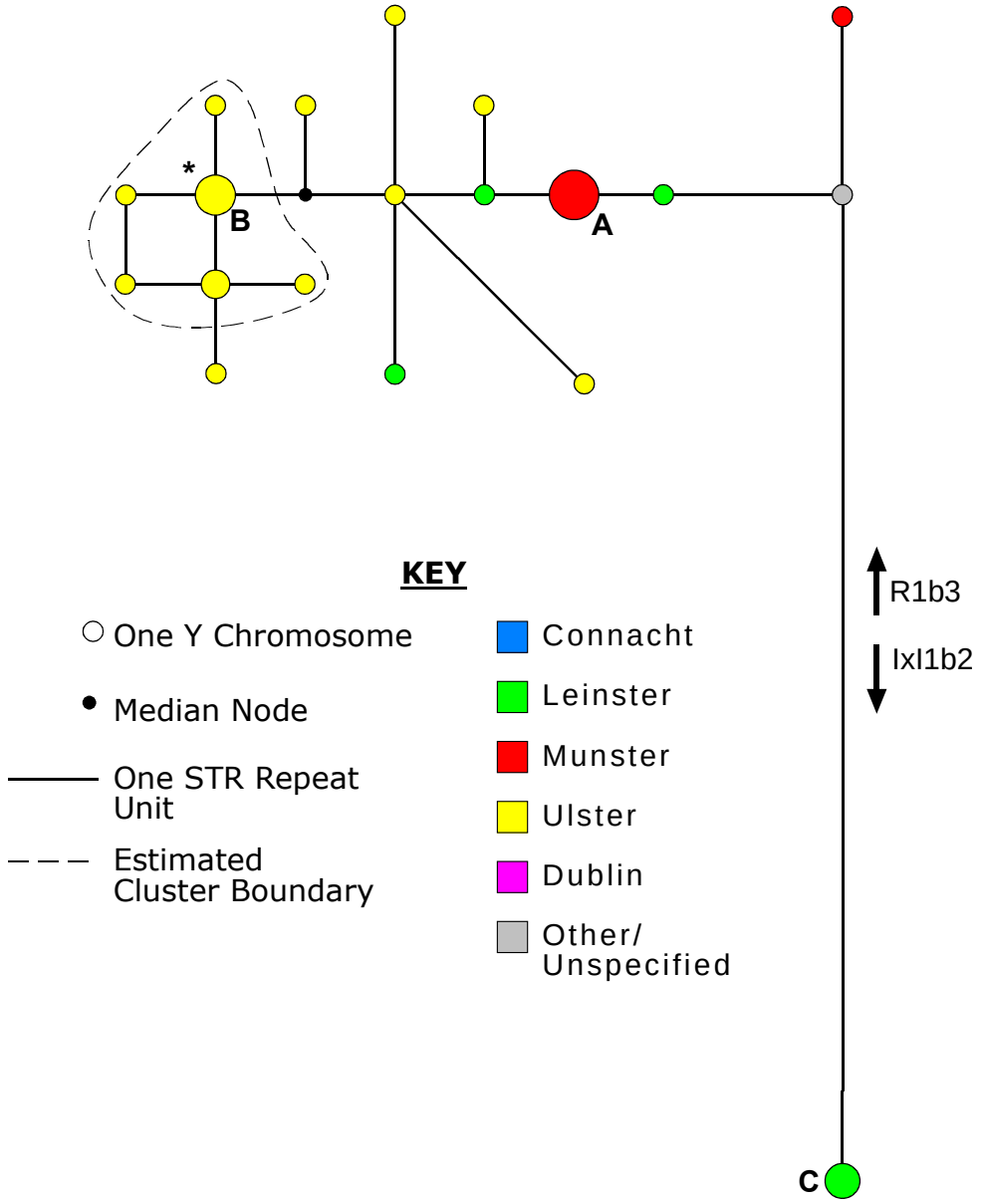


Figure 5.13 *MJ network of Y chromosomes in the Bradley surname. The 30 individuals represented in 19 haplotypes show evidence for geographically distinct foundations of the name in Ulster, Munster, and Leinster. Although the modal haplotype is composed purely of Munster samples (**A**), a marginally less frequent Ulster haplotype (**B**) forms the core of a diverse lineage, suggesting the earliest foundation was in this area. The occurrence of haplogroup IxI1b2 in a Leinster (Kilkenny) lineage (**C**) is consistent with a reputed English origin of the name here. *=IMH.*

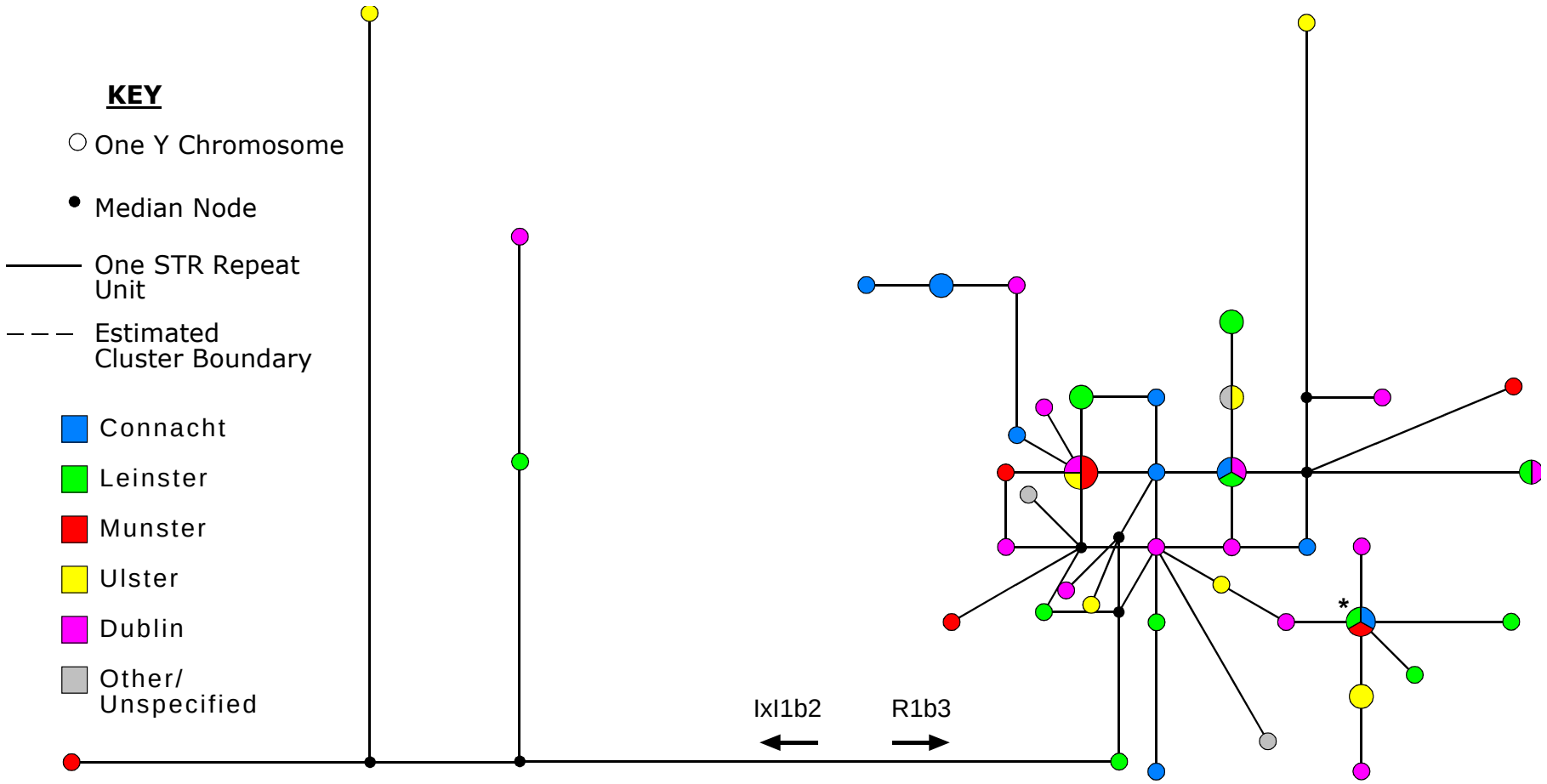


Figure 5.14 *MJ network of Y chromosomes in the Kelly surname. Kelly displays the highest haplotype (42) to sample size (55) ratio of any major surname examined, reflecting the substantial diversity within this single name. The absence of any predominant founding lineage is consistent with suggestions that the historically common personal name underlying Kelly (Ceallach) passed into hereditary names on numerous occasions throughout Ireland. *= IMH.*

5.3.2.j Murphy (Figure 5.15)

Murphy Y chromosomes display a similar, if somewhat less extreme, version of the diversity seen in Kelly, probably also reflecting the historic frequency of the personal name (*Murchadha*) whence the surname derives. While Murphy is common throughout the island it does display two distinct frequency foci in the Southwest and Southeast. The modal Murphy type (**A** in **Figure 5.15**) forms the centre of the only obvious cluster, which has a strong Munster geographic provenance and a TMRCA estimate of 1740 YBP. Although Southeast Leinster (Wexford, Carlow, Kilkenny) samples are distinct from this cluster, they do not form a single group suggesting multiple origins in this area alone. The low sampling ratio for the Murphy name (see **Figure 5.3**) means power to detect diversified clusters is also low. In this context, the presence of two repeated haplotypes (**B** and **C**) tentatively hint at two potentially old foundations in the Southeast.

5.3.2.k McEvoy (Figure 5.16)

McEvoy shows a strong dual foundation (centred on haplotypes **A** and **B** in **Figure 5.16**), cleanly divided by UEP defined haplogroups and underlying STR diversity. The divergence is reflected in the distinct Midland and Ulster geography of the two lineages respectively. These observations support suggestions of separate origins from the anglicisation of two phonetically similar Irish names to the same English version. Y chromosomes from the surname McVeigh are closely related to the Ulster McEvoy-B lineage (see **Section 4.3.4**) strengthening this conclusion as McVeigh represents a second anglicisation from the same Irish name underlying McEvoy in Ulster. The TMRCA (1380 YBP) of the McEvoy-B lineage is consistent with the early surname foundation period while that of McEvoy-A (2130 YBP) is substantially older, possibly an artefact of the apparent early splitting of the ancestral haplotype by the hyper-mutable DYS439.

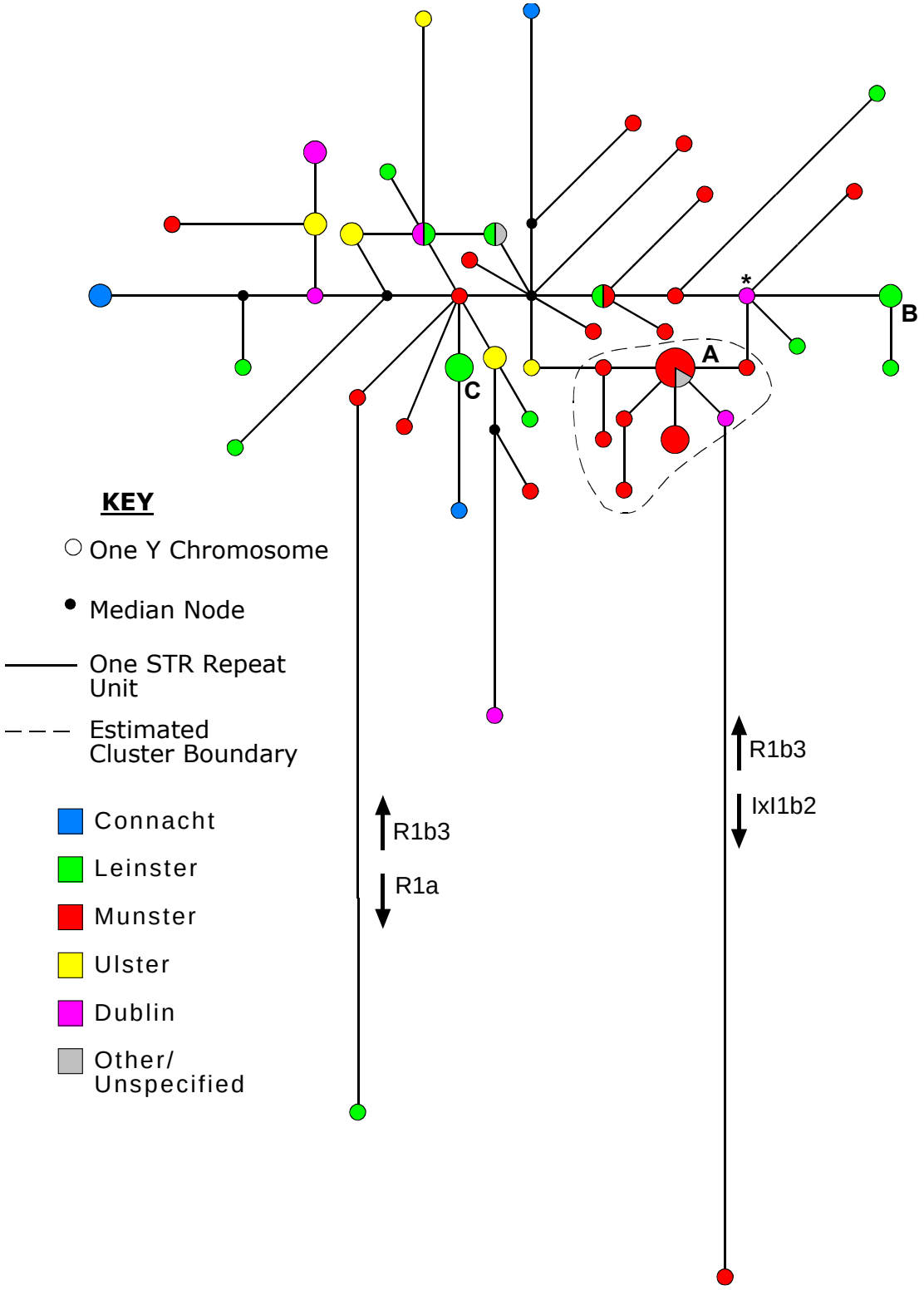


Figure 5.15 *MJ network of Y chromosomes in the Murphy surname. The polygenic origin of the name is reflected in high Y chromosome diversity with 47 haplotypes observed in the 65 individuals shown above. However, the legacy of a single major founder in Munster is apparent and centred on the ancestral haplotype **A**. While there is no corresponding predominant lineage in the Southeast (South Leinster), the repeated haplotypes **B** and **C** tentatively hint at the expected antiquity of the name here. *=IMH.*

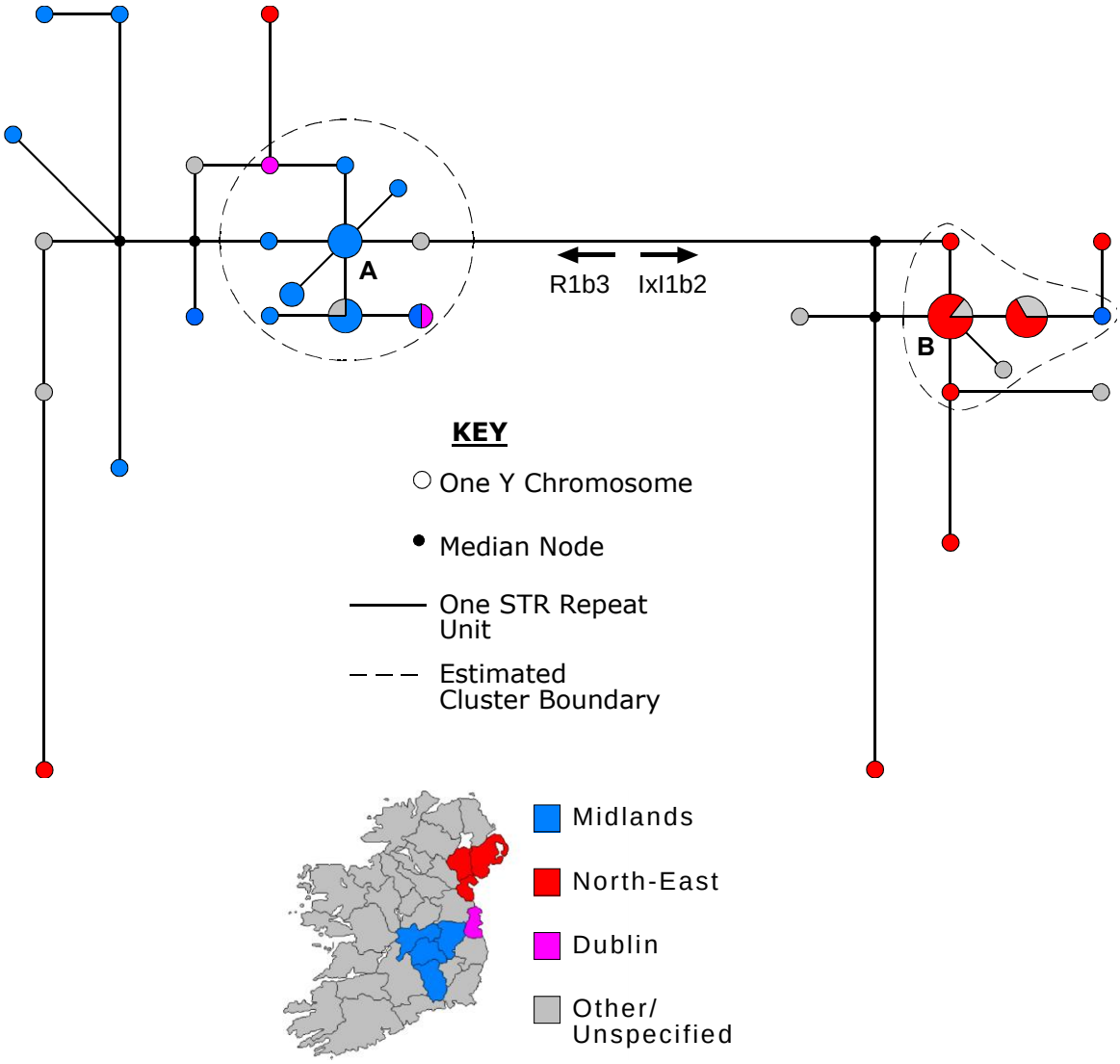


Figure 5.16 *MJ network of Y chromosomes in the McEvoy surname. The 31 haplotypes observed in 50 men reveal two prominent founding lineages centred on ancestral haplotypes **A** and **B**, which fall into different haplogroups. The clusters strongly respect the geographic distribution of the surname and apparently confirm the dual origin of McEvoy in the Northeast and Midlands from the independent anglicisation of two distinct, but somewhat phonetically similar, Irish language surnames.*

5.3.3 Inter-relationship of Major Surname Lineages

The relationships of prominent lineages identified within individual surnames were examined by constructing a single median-joining network of their ancestral (modal) haplotypes (**Figure 5.17;** full STR haplotypes for each lineage given in **Table 5.3**). In line with earlier evidence of good congruence between Y chromosomes and surnames (**Section 5.3.1**), only 2 of the 14 haplotypes are shared across surnames. In the first of these instances, the relationship between the founding McGuinness lineage (McGuinness-A) and the Ulster McEvoys (McEvoy-B) may represent a real genealogical link (discussed in Chapter 4). The second shared haplotype includes the South Munster branches of Bradley (Bradley-A) and O'Donoghue (Donohoe-C) as well as the presumed founding lineages of the Cork/Kerry names, O'Sullivan and McCarthy (McCarthy-B). The common geographic theme suggests a locally frequent Y chromosome in Munster. However, this haplotype was observed only 4 times in 184 random Munster Y chromosomes (2.2%) (data from Moore 2004), arguing instead that the observed lineage sharing represents a relationship independent of simple geographic coincidence. The founders of the O'Sullivan, McCarthy and O'Donoghue names (there is no information for Bradley to the author's knowledge), are reputed to stem from the pre-surname *Eóganachta* ruling Munster dynasty, whose members traced an origin to a common 5th century AD ancestor *Conall Corc*. (Ó Corráin 1972, pg 175). The TMRCA for the 60 individuals across these four surname lineages was estimated at 1210 YBP, consistent with a relatively recent common origin.

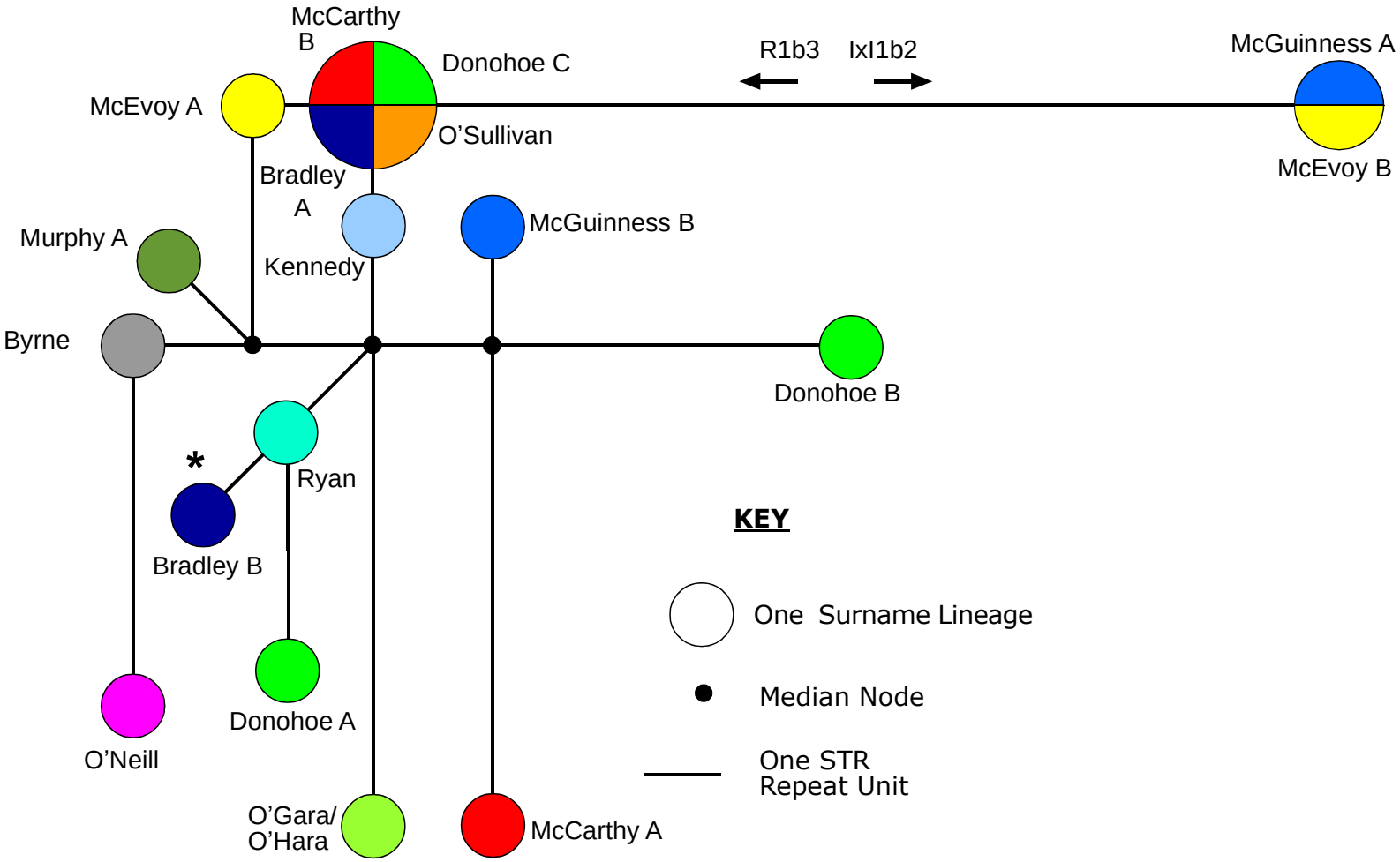


Figure 5.17 *MJ network of the ancestral Y chromosome haplotypes from major surname lineages. Only two of the 14 haplotypes are shared between surnames indicating a generally distinctive patrilineal ancestry underlying surnames. The occurrence of the same ancestral haplotype across four South Munster surname lineages is a major exception, and it may reflect the legacy of a pre-surname genealogical connection through the Eóganachta dynasty and its reputed common ancestor 'Conall Corc'. *= IMH. Full 17 marker STR haplotypes for these lineages are given in* **Table 5.3**.

	Y CHROMOSOME STR HAPLOTYPE																
SURNAME LINEAGE	DY S19	DY S388	DY S389I	DY S389IB	DY S390	DY S391	DY S392	DY S393	DY S434	DY S435	DY S436	DY S437	DY S438	DY S439	DY S460	DY S461	DY S462
MCGUINNESS A	15	13	13	16	24	10	12	15	9	11	12	15	10	11	11	10	12
MCEVOY B																	
MCGUINNESS B	14	12	13	16	24	11	13	13	9	11	12	15	12	14	11	10	11
BRADLEY B	14	12	13	16	25	11	14	13	9	11	12	15	12	12	11	10	11
BYRNE	14	12	14	16	24	11	13	13	9	12	12	15	12	12	11	10	11
O'NEILL	14	12	13	17	24	11	13	13	9	12	12	15	12	11	11	10	11
RYAN	14	12	13	16	25	11	13	13	9	11	12	15	12	12	11	10	11
KENNEDY	14	12	13	16	24	11	13	13	9	11	12	15	12	11	11	10	11
O'SULLIVAN	14	12	13	16	24	10	13	13	9	11	12	15	12	11	11	10	11
DONOHOE C																	
MCCARTHYB																	
BRADLEY A																	
MCEVOY A	14	12	14	16	24	10	13	13	9	11	12	15	12	11	11	10	11
O'GARA/O'HARA	14	12	13	16	24	11	13	12	9	11	12	15	12	12	10	11	11
MURPHY A	14	12	14	16	24	11	13	13	9	11	12	15	12	12	10	10	11
MCCARTHY A	14	13	13	15	23	11	13	13	9	11	12	15	12	13	10	10	11
DONOHOE B	15	12	12	17	24	11	13	13	9	11	12	15	12	13	11	10	11
DONOHOE C	14	12	13	16	25	11	13	13	9	11	12	16	12	12	11	11	11

Table 5.3 *Full ancestral STR Y chromosome haplotypes for major surname lineages. A MJ network of these is shown in* ***Figure 5.17****.*

5.3.4 TMRCA Estimates and STR Mutation Rate

TMRCA estimates for each major surname lineage are inevitably approximate given departures from the stepwise mutation model assumed in the dating method as well as uncertainties in cluster boundaries and especially in mutation rate. Estimates of the latter based on directly observed events in father-son pairs or longer pedigrees and those calculated from accumulated diversity in isolated populations over an evolutionary time scale (hundreds of years) are quite different (**Section 4.2.10**). The uncertainty in this factor alone is illustrated in **Figure 5.18** with TMRCA point estimates for 16 presumed founding or early surname lineages calculated using both rates. Although there is variation in estimates, when confidence intervals are considered most dates are not inconsistent with the major period of Irish surname foundation *ca*. 900–1200 AD. Indeed, it is noteworthy that this timeframe encompasses extensive overlap in the ranges of the two sets of estimates. While acknowledging the uncertainty in individual estimates, this argues that the age of prominent surname lineages collectively is most consistent with the historical period of surname adoption.

Assuming these prominent lineages (see legend, **Figure 5.18**) are each the product of about 1000 years of evolution from a single origin, then it is possible to estimate the average mutation rate for the 17 Y chromosome STR loci used in this study. From the relationship applied in TMRCA dating (**Section 4.2.10**) where Average Square Difference in repeat size between all current Y chromosomes and the ancestral type (ASD_0) is equal to μt (with μ is mutation rate and t is TMRCA), it is trivial to deduce that $\mu = ASD_0/t$. Accordingly, with an average ASD_0 per locus equal to 0.03671, the mutation rate is estimated at approximately 3.7×10^{-5} per locus per year. This value is similar, if somewhat faster, to the 2.76×10^{-5} evolutionary-based rate deduced by Zhivotovsky et al. (2004).

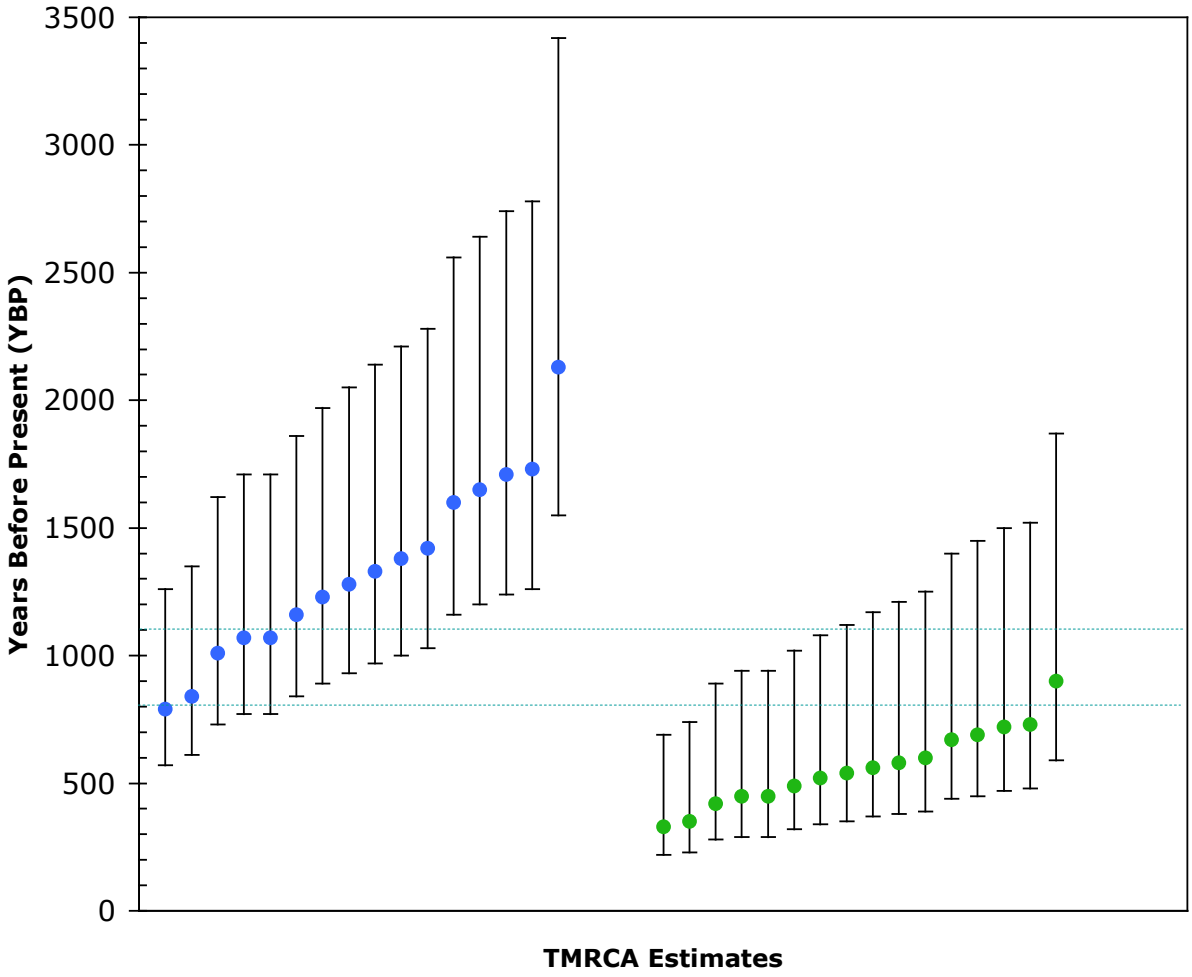


Figure 5.18 *TMRCA estimates for founding or early surname lineages. Dates for these 16 lineages are based both on evolutionary (blue) and direct observations/pedigree (green) derived mutation rates. The lineages are from left to right: Byrne, Ryan, McGuinness-B, Donohoe-A, Donohoe-B, O'Sullivan, O'Neill, McCarthy-B, Donohoe-C, McEvoy-B, Kennedy, McGuinness-A, O'Gara/O'Hara, Bradley-B, Murphy-A and McEvoy-A). While estimates vary and confidence intervals are large, most are not inconsistent with the main period of Gaelic surname foundation ca. 900 – 1200 AD indicated between the broken lines. Indeed, this period encompasses a great deal of overlap between the ranges in TMRCA estimates from both rates. Despite individual age uncertainty, major surname lineages collectively are consistent with an origin in this timeframe. The average evolutionary rate estimate for these 16 lineages is 1340 YBP (2140 – 970 YBP) while that for the faster direct observation rate is 560 YBP (1170-370 YBP).*

5.4 DISCUSSION

5.4.1 What is in an Irish Surname?

It seems clear, from both investigations conducted at the local level (Chapter 4) and in a general island-wide survey, that surnames mark genuine Y chromosomes divisions. The genetic distinction of prominent lineages between surnames as well as age estimates consistent with the timeframe of early Irish surname adoption, strongly suggests that Irish surnames collectively reflect real and recent kinship through a common paternal ancestor. Such a conclusion is in agreement with both recorded history and indeed the very fabric of Irish surnames, which nearly always emphasise an ancestral bond through the prefixes *Mac* and *Ó*.

While this is a robust general truism, it is equally clear that the extent and nature of the relationship between individuals of the same name is a variable quality, reflecting differences in the foundation and later histories of each surname. Some are monogenic, dominated by the Y chromosome legacy of a single and early founding male and these include O'Sullivan, Ryan and Byrne. Others are characterised by multiple foundations, most notably the highly diverse Kelly and Murphy surnames. However, regardless of early origins, the Y chromosome diversity of all surnames invariably bears witness to the introgression of numerous post-foundation males betraying the considerable complexity of subsequent surname history and evolution.

5.4.2 Agreement of Y Chromosomes and History

The general agreement of intra- and inter- surname Y chromosome diversity with expectations from historical and genealogical information was notable and indicates the remarkably detail preservation and time range to these sources. They have faithfully transmitted a record of events over the past 1500 years that includes the derivation of McGillycuddy from O'Sullivan in the 16th century, a 10th century common origin of O'Hara and O'Gara, and the 6th century common ancestor of McGuinness and McCartan. In the broader realm of Irish history, the knowledge preservation this represents suggests a degree of general reliability to early accounts and records that are otherwise difficult to confirm.

There is also more tentative evidence for the survival of a pre-historic genealogical record in the *Eóganachta* dynasty of Munster. The recent common ancestor of several surname lineages with reputed *Eóganachta* origins is at least consistent with both the idea of single dynastic ancestor and the suggested timeframe (*ca.* 400 AD). It is a hypothesis easily tested by comparing the Y chromosome diversity of other putative *Eóganachta* derived names, such as O'Callaghan, O'Keefe and O'Mahoney, to that found in surnames originating from other Munster population or tribal groupings. In the 10th century, one such group, the *Dál Cais*, took political control of Munster and genealogies were apparently forged to link them to the *Eóganachta* as a means of validating the change of power (Ó Corráin 1972, pg 9). In this regard, it is interesting to note that the primary lineage in the Kennedy surname, which arose from the *Dál Cais* leaders, does not show the putative *Eóganachta* haplotype although, intriguingly, it is only one mutational step removed. The idea of a biologically-defined ruling caste also appears to have parallels in the Northern Ireland equivalent of the *Eóganachta*. Here the *Uí Néill* grouping also seems to have reflected or supported their political control with a large kin group (Moore 2004).

5.4.3 Surnames and Geography

Even in the absence of any specific history, the geographic distribution of surnames alone seems to be an informative source on their origins. Geographically distinct peaks in surname frequency generally indicate separate origins. The occurrence of the Bradley surname in Derry and Cork, for example, against a low frequency background is best explained by independent origins rather than a migration from one end of the island to the other. The intense geographic localisation of many monogenic names right up to 19th century (sometimes 900 years after foundation) and the apparent lack of gene flow between multi-centric surnames like Donohoe, together suggest a low male migration rate over this period. The presence of the otherwise predominantly Ulster O'Neill founding lineage in bearers of the surname in Munster is a notable exception and may be explained by the equally exceptional prominence of the family in Irish history. The southern link may be the result of settlement in Munster by members of the main lineage after the defeat of a large Ulster force, led by Hugh O'Neill and Hugh O'Donnell, at the Battle of Kinsale (County Cork) in 1601.

While discrete geographic frequency foci generally represent distinct origins, it does not necessarily follow that a single focus represents a single foundation. Donohoe in Cavan and McGuinness in Down (Chapter 4) each feature the prominent genetic legacies of two men within the same general area. Overall, while both the history and geography of surnames are important sources of information, the molecular Y chromosome approach has a powerful dual role in confirming (and refuting) putative genealogical links and also providing a fresh perspective to reveal unanticipated aspects of surname history.

5.4.4 The Complexity of Surname History

Many surnames showed evidence of a single major founding male. However, even in these cases the once total linkage between surname and original Y chromosome has been eroded by subsequent history. All names appear to be highly permeable to later male introgression. Even in O'Sullivan, the closest ideal of a monogenic name encountered in this study, only about half the current surname bearers are estimated to descend from the original single founder. Although the Y chromosome is effective at recording breaches of coinheritance with surname, it can offer little direct evidence on the exact cause of each event.

The most obvious explanation is simply the cumulative effect of non-paternities. Current estimates of this vary hugely but recent analysis of genotypic data in Iceland indicated a rate of 1.49% per generation (Helgason et al. 2003). The 1.65% per generation surname/Y chromosome decay rate observed in O'Sullivan is very similar. However, as the underlying events were not observed this represents a maximum rate of non-paternity, only realised if all the accumulated diversity was accounted for by this explanation. In pre-1600 Gaelic society, an unusual expression of non-paternity could result from the practice of 'naming'. A woman could attribute her child, often many years on, to a true biological father leading to name changes and sometimes dramatic advances in status for the child. Mathew O'Neill, a 16th century Earl of Dungannon, for example, lived as Matthew Kelly, son of Dundalk blacksmith until his teens when he was attributed by his mother to an O'Neill Earl of Tyrone. (Nicholls 2003, pg 88-90). Understandably, this may have attracted new Y chromosomes to pre-existing surnames, particularly those associated with power and influence. However, the same net effects can also be explained by adoptions or maternal transmission of surname. As an example of the latter, when the male line of one O'Neill branch died out in the late 18th century the name continued through a daughter and her sons disregarding the father's surname

(Chichester). Their descendents, including Terence O'Neill former Prime Minister of Northern Ireland, incongruously carried the archetypal surname of the Gaelic order but 'Chichester' Y chromosomes (Bell 1988, pg 210).

Alternatively, rather than reflecting decaying vertical transmission, the diversity may represent, in part, the horizontal absorption of independent but phonetically-similar names. Exemplifying Byrne, the transmutation of the similar Connacht surname 'Beirne', derived from the Norse personal name Bjorn (MyLysaght 1985b, pg16) (see also Chapter 6), may best explain the genetic distinctiveness of the name in the Northwest away from the monogenic historic Leinster heartland. Numerous other such possibilities exist both for Byrne and most surnames, a product of their perturbed development over the past millennium.

5.4.5 Variation in Surname Frequency

Massive disparity in frequency is one of the most obvious differentials between modern Irish surnames. The most common Irish surname, Murphy, constitutes over 1% of the entire island's population while many others number in the low hundreds and sometimes less. The confirmation of polygenic origins for Murphy and Kelly help to explain their positions as the most numerous surnames in Ireland. However, the high frequency of essentially monogenic names raises a new puzzle of why qualitatively similar origins can now display such dramatic quantitative differences. For example, there is ample historical and now genetic evidence to suggest that both O'Sullivan and O'Gara each had a single 10th century founder. Yet whereas *Gadhra* (the O'Gara eponym) has several hundred descendents left in 21st century Ireland, his rough contemporary in Munster *Súileabhain* has some 20,000.

Surnames, by virtue of their patrilineal inheritance, theoretically simulate neutral alleles of the Y chromosome with frequencies (and extinctions) influenced by the forces of

neutral evolution including genetic drift (Manni et al. 2004 and references therein). Although these alone could create modern frequency variation, differences are also probably influenced by social/cultural environmental selection leading to reproductive variance. Ireland in the Middle Ages was essentially a polygamous society. Particularly for the 'upper classes', multiple wives and concubines were accepted, while divorce was easily attained leading to serial monogamy. Archbishop Lanfranc of Canterbury writing in 1074 condemned the loose Irish marriage practises as "a law of marriage which is rather a law of fornication" (Duffy 1997, pg 22-27). Furthermore, as witnessed in the 'naming' practice, Irish law did not distinguish legitimacy in succession matters between children from any of these relationships or any casual encounters, all of whom were entitled to carry their biological father's surname and to share inheritance rights (Nicholls 2003, pg 83-90).

There was an obviously potential for this to result in high reproductive variance or as the 17th century genealogist Dualtagh Mac Firbisigh put it "as the sons and families of the rulers multiplied, so their subjects and followers were squeezed out and withered away". History records the consequences, including the 15th century Turlough O'Donnell, Lord of Tirconnell, who had 18 sons by ten different women and 59 direct male line grandsons (Nicholls 2003, pg 11-13). The attribution of 16 million male-line descendents to Genghis Khan (1162-1227) and his kin group is a testament to the potential impact of social driven selection over relatively short timeframes (Zerjal et al. 2003). In a slightly less dramatic example the most frequent lineage in the McCarthy surname does not seem to correspond with that of the oldest foundation but rather is the result of a rapid expansion from a more recent introgression. A similar effect is potentially evident in the Donohoe surname of Cavan (Chapter 4), with the possibility that the modern Y chromosome legacy of a 15th century adoption is as great as that associated with the original 12th century foundation. While powerful and prominent

founders may have led to more frequent and fragmented modern surname populations, this does not seem to be the whole story. Both O'Sullivan and Ryan (the most 'intact' names with over 50% of the current population descended from the original foundation) appear to have risen to later prominence from humbler ancestors. This suggests a converse effect where a large kin group provided the basis for an ascent to power and influence. Finally, it is also conceivable that direct natural selection of Y chromosomes could alter the frequency of linked surnames. Although Y-linked genes can affect fertility (Jobling and Tyler-Smith 2000; Jobling and Tyler-Smith 2003), there is no specific evidence here that could prove or disprove such a proposition. Overall the dramatic variance found over the short history of surnames points to the importance of male behaviour in shaping longer-term and wider population Y chromosome differentiation.

5.4.5 Conclusions

Irish surnames collectively are real and robust markers of common patrilineal ancestry. However, the exact extent of this depends on the specific name in question and the nature of its foundation. Some names have numerous early origins, while most have a defined and focused early genesis. In either case, it is clear that subsequent events of the thousand-year long history of Irish surnames have been a substantial force in shaping the genetic diversity of a modern surname population. The frequency today may be influenced by the power and prestige associated with the name in the past, while intra-surname Y chromosome heterogeneity bears witness to the impact of a complex history of multiple events, arising from the life and times of succeeding generations. Precisely because of this complexity, the Y chromosome is a useful and important new source in the investigation of surname adoption and history.

CHAPTER 6

SCANDINAVIAN ADMIXTURE IN AN IRISH NORSE SURNAME POPULATION

6.1 INTRODUCTION

6.1.1 The Vikings and Ireland

The term 'Viking' describes sea-faring raiders and later settlers, originating in Scandinavia from the late 8th century AD. Over the following 400 years or so, Viking expeditions affected much of Europe and beyond, from the Caspian Sea in the East to North America in the West. This great expansion was facilitated by their superior maritime technology, including the fast, flexible 'long-ship', and spurred on by increasing population pressure on limited resources within Scandinavia (Haywood 1995).

Three raids on monastic sites in 795 AD were Ireland's introduction to the Vikings. The following decades saw a gradual increase in both the frequency and geographic extent of the raids to include inland sites along major rivers as well as coastal areas. The precise origin of these Vikings is somewhat enigmatic, especially given the difficulty of equating the labels of historical writers to modern national boundaries and ethnicities (Davies 2000, pg 209-211). However, it is generally agreed that modern Norway represents the homeland for the majority of Viking raiders and settlers and the term 'Norse' is therefore frequently used to describe their activity in Ireland. There is also some evidence for a Danish element. For example, Dublin passed between the control of *Finngaill* ('blond foreigners') and *dubh gaill* ('dark [haired] foreigners') in the ninth century AD; descriptions traditionally thought, but not universally agreed, to represent Norwegians and Danes respectively (Ó Cróinín 1995, Chapter 9; MacShamhraín 2002, pg 47-48).

In 841 AD, less than fifty years after the first raids, a new phase in Viking or Norse activity began with the foundation of their first settlements in Ireland. Initially, these were crude military encampments or 'longphorts', but later grew into the first recognisable Irish towns including Dublin and later Cork, Limerick, Wexford and Waterford. Norse power and influence reached its height in the early to mid 10th century but declined soon after and the Norse colonial kingdoms increasingly fell under the over-lordship of local Irish kings. The Norse towns retained varying degrees of autonomy under local Scandinavian dynasties until the Anglo-Norman conquest in 1169 and some communities endured as late as the 14th century (Ó Cróinín 1995, Chapter 9; Bradley 1988).

6.1.2 Viking Settlement in Ireland

Despite their prominence in Irish history there is little direct evidence for large scale Scandinavian immigration. Even at its height, Norse power was geographically quite restricted, centred on towns and their immediate hinterlands. Furthermore, these small urban kingdoms appear to have been largely independent of each other. In contrast, other Viking domains, such as the Danelaw in England and Normandy in France, were extensive and unified. The scarcity of Norse place names in Ireland and paucity of word borrowing from Norse into Irish also suggests a limited number of Scandinavian settlers (Ó Cróinín 1995, Chapter 9).

It appears that after the initial shock of Norse raiding and settlement, contacts between them and the native Irish became commonplace. There was inter-marriage from an early stage and the small Norse kingdoms quickly became involved in the fractious Irish politics of the time, forming alliances with the native Irish leaders. From the early tenth century, such connections are reflected in the lack of distinction between the material

culture of Norse towns and surrounding Irish settlements. Both Scandinavian and Irish influences had become fused in new styles, for example in metal work, that were neither purely one nor the other. The term 'Hiberno-Norse' or 'Hiberno-Scandinavian' is employed to describe this culture and its population (Bradley 1988). Nonetheless, the inhabitants of Hiberno-Norse towns were distinctively referred as *Ostmen*, old Norse for 'men of the east'.

6.1.3 The Genetic Legacy of the Viking Era

The activities of the Vikings have left a substantial genetic legacy in some parts of North Atlantic Europe, particularly Iceland (Helgason et al. 2000a; Helgason et al. 2000b) and the Orkney Islands (Wilson et al. 2001). A small number of previous molecular genetic studies have also directly attempted to assess the impact of Norse introgression on the Irish population. 6.1% of the mutations in the Phenylalanine Hydroxylase gene found in Ireland (and which are responsible for the Hyperphenylalanininaemia group of autosomal recessive disorders) are of a type predominantly associated with Scandinavia (O'Donnell et al. 2002). This fraction was attributed to Viking activities although the most common Norwegian mutation was not observed in Ireland. A wider assessment of the Viking genetic legacy in the British Isles using the Y chromosome produced mixed results for Ireland (Capelli et al. 2003). Principal component analysis (PCA) of Y chromosome allele (haplogroups) frequencies suggested little or no Viking introgression in population samples from Roscommon and, more surprisingly, Fingal, an area that owes its very name to Norse settlement. In contrast, admixture calculations from the same study using a maximum likelihood method (Chikhi et al. 2001) indicated a substantial 34.8% Norwegian ancestry in the Fingal sample. However, confidence intervals on this estimate were wide and encompassed the range 1.9% to 93.8%.

6.1.4 Norse Surnames and Paternal Ancestry

Hereditary surname adoption (see Chapters 4 and 5) began in Ireland *ca*. 900 AD and continued for several centuries. This roughly coincides with the Norse period in Ireland (*ca*. 800 AD to 1200 AD) when numerous Scandinavian personal names such as Swen, Torcall, Sitric and Olaf, were introduced onto the island. These and others subsequently passed into hereditary surnames by the prefixing of *Mac* and *Ó*. Some surnames also developed from nicknames or other descriptions of the Norse, including Doyle, the most numerous putative Norse surname today, which is derived from the Irish *Ó dubh ghaill* or 'descendent of the dark foreigner' (Woulfe 1923, pg xxiv - xxvii; Ó Cuív 1988)

Surnames often display a specific cultural or population origin and can therefore be used as proxy identifiers of ethnic ancestry. Furthermore, the theoretical paternal coinheritance of surnames with Y chromosomes allows the latter to investigate the congruence of this putative ancestry with genetic affinities. Hill et al. (2000) used this approach to identify significant differences in the Y chromosomes of Irish men with indigenous Gaelic surnames compared to those of exogenous origin. A small sample of 18 Y chromosomes from Irish men with Norse and Norman surnames (the latter are putative descendents of Vikings in France) showed some increased diversity though not to the same level seen in Irish Anglo-Scottish surname bearers. However, the extent of Viking paternal ancestry in Irish men with putative Norse surnames remains unknown.

6.1.5 Estimating Population Admixture

The contribution of different source populations to a hybrid or admixed group can be roughly gauged through genetic distances. However, a quantative estimate of the proportion of the hybrid's ancestry derived for each parent population can also be made under certain assumptions. Most attempts to calculate admixture percentages are based on the model illustrated in **Figure 6.1**. In this scenario, an initial ancestral population

diverges into two or more parental populations, which evolve independently for a certain time. Later, a hybrid population is created by instantaneously combining contributions from each parent in a certain percentage or admixture proportion (*m*). After formation, the new hybrid and parental populations once again evolve in isolation with no further genetic exhange. In practise many of these assumption will be violated in real human situations. For example, gene flow may continue between parental populations even after they 'split' from a common ancestor and the creation of human hybrid populations is rarely instantaneous but rather might involve contact and gene flow over several generations.

Early attempts, and indeed most current methods, use allele frequencies as the basis for quantitative admixture estimates. In a hybrid population these are expected to be a linear combination of the allele frequencies in the parental populations. In mathematical terms, $P_H=P_1m+P_2(1-m)$, where *m* is the admixture proportion derived from the first parental population and P_H, P_1,and P_2, are the allele frequencies in the admixed and two parental populations respectively (see **Figure 6.1**). This model is simplistic as allele frequencies can fluctuate after the admixture event from random genetic drift, migration, mutation, and potentially selection. Several different admixture estimation methodologies are commonly used, many of which try to incorporate some of these potential complications (reviewed by Choisy et al. 2004). However, no current admixture method can fully account for all potential confounding factors.

Figure 6.1 (overleaf) *The Admixture Model. Most admixture calculations assume the simplified scenario outlined in the above diagram. An ancestral population P_0 splits into two parental populations P_1 and P_2 at time T_2. These evolve independently during Period B, until at a later time (T_1) a new hybrid population (H) is instantly created with* ***m*** *proportion of its genes derived from P_1 and the remainder (1-* ***m****) from the second parental populations P_2. The three populations then evolve with no further contact during Period A. Finally, at the present time (T_0) sample populations are drawn and admixture proportions calculated. Figure redrawn with modification from Choisy et al. (2004).*

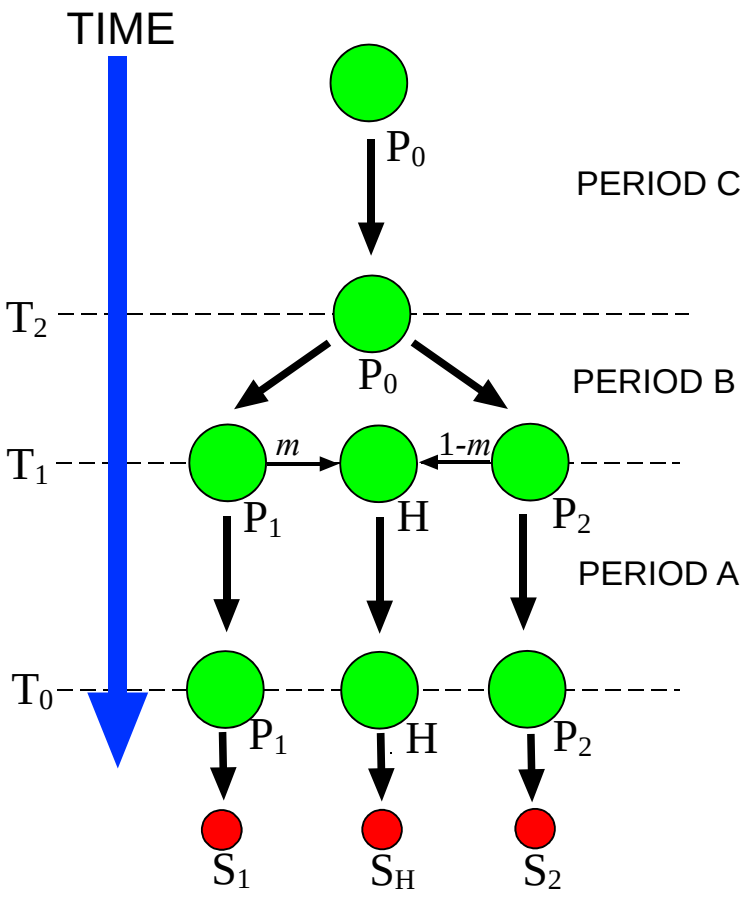


Figure 6.1 *The Admixture Model* **(Full legend on the preceding page)**

6.1.6 Present Study

UEP and STR Y chromosome variation in 47 Irish people with putative Norse surnames is examined in this Chapter. Under the paternal coinheritance of surname and Y chromosome, they are hypothesised to represent an admixed group combining Irish and Scandinavian genetic contributions in unknown proportions. The extent of any Norse paternal ancestry in this group can be used to gain insight into the scale and nature of Viking settlement in Ireland. Their general affinity to potential source populations in Ireland and Scandinavia is initially investigated and this is followed by the calculation of quantitative admixture proportions using a range of current estimators. Although these display different strengths and weakness, together they allow the formation of broad based conclusions and provide the parallel opportunity for a comparative assessment of each method's efficacy.

6.2 MATERIAL AND METHODS

6.2.1 Sample Collection

Putative Viking surnames were identified according to McLysaght (1985b). Samples were collected in the period 1999-2001 through a mixture of personal contacts and volunteer recruitment through postal request. Ms. Nessa O'Sullivan assembled 17 samples in 1999, while a further 10 were collected in 2000, mainly through personal contacts and in co-operation with Ms. Claire Brady. The remaining samples were collected in 2001 by postal recruitment as described in Chapter 4 (**Section 4.2.1**) with the assistance of Ms. Luzveminda O'Sullivan. Request were sent to 77 bearers of putatively Viking/Norse surnames residing primarily in counties Cork, Limerick, Dublin (particularly Fingal), Waterford and Wexford where Norse activity and settlement is thought to have been greatest. A total of 26 replies were received, a response rate of 33.8%. A total of 53 samples were therefore available for this study.

6.2.2 Y Chromosome Genotyping

DNA was extracted using a standard phenol/choroform protocol described in Chapter 2 (**Section 2.2.2**). Samples were genotyped for 13 STR loci (DYS19, DYS388, DYS390, DYS391, DYS392, DYS393, DYS385A, DYS385B, DYS389I, DYS389II, DYS460, DYS461, DYS462) in two multiplexes (MS1 and EBF) as described in **Section 4.2.5**. Several UEPs, which divide samples into broader haplogroups, were also examined in a hierarchal manner (see **Section 4.2.6)**. STR profiles were generated for 47 samples (88.7% of the total sample archive) but complete UEP typing was only possible for 44 of these. The haplogroups of the remaining three samples were inferred from their STR profiles.

The 47 usable DNA samples were distributed over 26 names as follows (*number*): Arthur (*2*), Beirne (*2*), Bligh (*1*), Boland (*3*), Caskey (*1*), Coll (*1*), Coppinger (*2*), Doyle (*4*), Gohery (*2*), Hanrick (*2*), Harold (*1*), Hendrick (*1*), Higgins (*2*), Kells (*2*), Kettle (*1*), Loughlin (*1*), McGetrick (*3*), McLoughlin (*3*), Nelson (*1*), Norris (*1*), O'Higgins (*1*), Sugrue (*2*), Sweetnam (*1*), Thunder (*1*),Toner (3), Tormey (*3*).

6.2.3 Analysis of UEP Variation

All analysis was carried out separately for UEP and STR markers to allow an assessment of the relative information content of markers with very different mutational dynamics, which can be difficult to analyse together. Norse surname group (NSG) Y chromosomes were divided into haplogroups as outlined in Capelli et al. (2003), in turn based on Y Chromosome Consortium (2002) nomenclature. The SNPs M173 and M17 used in the former study were not typed in the NSG but M269 and SRY1532 respectively were taken as phylogenetic equivalents for the purposes of comparable haplogroup definition. The general affinity of the NSG using UEP variation was investigated in concert with the data of Capelli et al. (2003). Regional population groupings were constructed from this as follows (*sample size*): Scottish Isles (*272*); Scotland (*178*); Ireland (*119*); Wales (*196*); England North (*231*); England Midlands (*205*); England Southwest (*125*); England Southeast (*256*). Potential Viking homelands were represented by Norway (*201*) and Denmark/North Germany (*190*). Population genetic distances as linearised Φ_{ST} values were calculated from the mean pairwise difference between Y chromosome UEP haplotypes as implemented in the ARLEQUIN package, Version 2.000 (Schneider et al. 2000). Φ_{ST} significance was ascertained by randomly permuted individuals across populations over 10,000 replicate analyses. The inter-population Φ_{ST} matrix was summarised in two-dimensional space using the ALSCAL program in the SPSS package (Version 11.0, SPSS Inc.).

6.2.4 Analysis of STR Variation

An independent comparative population dataset was assembled to investigate the affinities of the NSG through STR variation. Samples were sourced as follows (*sample number*): Ireland (*221*) (Hill et al. 2000); Norway (250) [composed of 112 samples from Helgason et al. (2000b), 45 from Zerjal et al. (2001) and 83 from Weale et al. (2002)]; Denmark (*74*) [including 12 from Helgason et al. (2000b) and 62 from Bosch et al. (2003)]; Sweden (*110*) (Zerjal et al. 2001); England (*215*) (Weale et al. 2002); Scotland (*61*) (Helgason et al. 2000b); Iceland (*181*) (Helgason et al. 2000b) and Wales (*98*) (Weale et al. 2002). Although 13 STR markers were typed in the NSG only DYS19, DYS390, DYS391, DYS392, DYS393 were considered in most analysis to allow maximum comparability between datasets. Genetic distances as linearised R_{ST} values were calculated using AREQUIN. R_{ST} is an analog of F_{ST} that takes account of the molecular basis of STR evolution (under the simple stepwise mutation model) by relating haplotypes through the sum of the squared repeat size difference between them (Slatkin 1995). Significance assessments and MDS analysis were carried out as with UEP variation (previous section).

6.2.5 Admixture Estimation

There are currently several different methods for quantitatively estimating admixture proportions (*m*) from genotyic data in a hybrid group. All of these use allele frequencies as the basis for admixture calculations but differ in the way this information is applied in computations. As a result, they display different strengths and weaknesses but together provide a broad-based foundation for assessing admixture. Six of the most commonly used methods were applied to the question of Irish and Scandinavian ancestry in the NSG. Estimates were carried out using UEP and STR data separately because of difficulties in combining both for some methods (see Helgason et al. 2000b).

UEP alleles are simply represented by the 11 haplogroups defined in Capelli et al. (2003), while Y chromosome STR alleles are compound haplotypes of the five loci DYS19, DYS390, DYS391, DYS392 and DYS393. This dual approach allows investigation of the efficacy of methods based on a small number of discrete alleles (haplogroups) versus those from a higher number of alleles (multi-STR loci haplotypes), which have a greater 'noise' risk. A brief overview of each methodology and its implementation is given in the subsequent subsections.

6.2.5.a m_R

The most basic estimator m_R (Roberts and Hiorns 1965) calculates m directly from allele frequencies using a least-squares regression line through points defined by P_1-P_2 versus P_H-P_2 (see **Figure 6.1**). The slope of this line gives the admixture proportion derived from the first parental population. m_R was calculated using ADMIX 2.0 (Bertorelle and Excoffier 1998; Dupanloup and Bertorelle 2001) (http://web.unife.it/progetti/genetica/Isabelle/admix2_0.html). Confidence intervals were calculated as the standard deviation over 10,000 random bootstrap samples.

6.2.5.b m_C

A later variation on m_R, the m_C estimator (Long 1991; Chakraborty et al. 1992) also uses a least squares regression method. In addition, it incorporates the effects of drift in the hybrid population since its inception and stochastic sampling error in the hybrid and parental populations. m_C values were calculated using the program LEADMIX ('Likelihood Estimation of Admixture') Version 2.0 (Wang et al. 2003) (http://zoo.cam.ac.uk/ioz/software.html). The software does not return any confidence intervals as estimates are based on a single locus (the Y chromosome).

6.2.5.c m_Y

The novel admixture estimator m_Y (Bertorelle and Excoffier 1998) not only takes into account the frequency of each allele but also their molecular distances using a coalescence theory framework. In doing so, it incorporates the effect of mutation since the admixture event. The method requires the molecular relationship between alleles to be defined. For DNA sequence or UEPs, this is simply the number of pairwise differences. The sum of the squared difference in allele sizes, which is related to coalescence time under the SMM (Slatkin 1995), can be used for STR loci. In addition, the age of the admixture event and mutation rate must be specified. The former was set at 1050 YBP (950 AD, midway through the Viking period in Ireland). The effect of mutation in UEP markers is effectively zero in this timeframe while a rate of 1.38×10^{-5} per year for STR alleles (based on 5 loci haplotypes) was adopted from Zhivotovsky et al. (2004). Confidence intervals are obtained as the standard deviation of 10,000 random bootstrap samples.

6.2.5.d m_L

The foregoing are moment estimators; alternative maximum likelihood approaches are potentially more powerful as they can incorporate and explore a variety of other parameters (such as the effect of drift) when estimating admixture proportions. The method of Chikhi et al. (2001), here designated as m_L, estimates both the admixture proportion and the effect of drift on the parental and hybrid populations using a coalescent based likelihood method. Ancestral allele frequencies compatible with the observed data are reconstructed and the probabilities of obtaining the current frequencies in the hybrid are calculated for different parental population contributions.

Calculation of m_L was carried out using the software LEA ('Likelihood-based Estimation of Admixture') (http://www.rubic.rdg.ac.uk/~mab/software.html). A new version of the program for this study was compiled by Martin O'Hely (Université Paul Sabatier, Toulouse) to allow admixture estimates for loci with more than 100 alleles. Admixture proportion solutions were explored through runs of 55,000 Markov chain Monte Carlo (MCMC) iterations with the first 5,000 discarded as burnin. The median value returned was taken as the estimate of the admixture proportion derived from the first parental population. The 97.5 and 2.5 percentiles provided the 95% credible interval for this estimate.

6.2.5.e m_W

A second maximum likelihood method (Wang et al. 2003) emphasises the effect of genetic drift and models this in all populations (parental and hybrid) both before the admixture event (the extent of differentiation between parental populations) and subsequently in the parental and admixed populations. It also incorporates the effect of sampling error for each population. All parameters are considered jointly to maximise the likelihood of admixture proportion estimates. The calculation of this estimator, referred to here as m_W, was carried out in the program LEADMIX (Version 2.0) (see **section 6.2.5.b)**. m_W values were obtained using 500 'integration points' in computing the likelihood function. High values (such as 500) lead to a more accurate approximation of this function but were also found to acutely increase the computation demand when more than two parental populations were allowed. In the latter circumstances, 100 integration points were used to facilitate a tractable analysis. m_W point estimates are returned bounded by 95% credible intervals.

6.2.5.f m_ρ

Y chromosomes defined by highly variable compound STR haplotypes begin to approximate an infinite allele system. Furthermore, as a single locus, the Y chromosome has a relatively small effective population size making it more susceptible to genetic drift. Together these features can lead to the admixed population sample containing numerous 'private lineages' or alleles not observed/sampled in either parental population. Estimators like m_R cannot incorporate these because their frequencies in the parental sample populations are zero (Helgason et al. 2000b). To deal with the specific case of Y chromosome STR admixture calculation, Helgason et al. (2000b) developed the estimator m_ρ, which uses a heuristic iterative approach to find the parental contributions to the hybrid population that best fit the observed allele frequencies. It identifies the closest match (or matches) as an ancestral haplotype(s) for private lineages permitting their incorporation in calculations and incidentally allowing for the effect of mutation since the admixture event. m_ρ and its 95% confidence intervals were calculated using software provided by Agnar Helgason (deCODE Genetics, Reykjavík). Distances between alleles, for the identification of potential ancestor(s) of private lineages, were defined as the sum of repeat unit difference between haplotypes.

6.2.5.g Parental Populations

The Irish population was held constant in all calculation as one presumed parental population. The uncertainty in the origin of Irish Vikings was reflected and explored using various Scandinavian parental populations permutations as follows: (1) a single combined Scandinavian population sample; (2) individual Scandinavian national (country) samples considered in turn (3) multiple national samples considered as discrete units but simultaneously (only possible for the m_R, m_Y, and m_W estimators).

Irish and Scandinavian parental population samples for UEP and STR admixture analysis are as described in **sections 6.2.3** and **6.2.4** respectively.

6.3 RESULTS

6.3.1 Phylogeography of Norse Surname Y Chromosomes

Previous studies have identified Y chromosome lineages that are thought to be indicative of Viking contact and introgression. Two such UEP/STR defined haplotypes and their one step STR neighbours are found at high frequency in Norway and other areas affected by Viking migrations (Wilson et al. 2001). These types, termed 2.47 and 3.65, are either absent or at low frequency in the NSG (**Figure 6.2**). In contrast the Atlantic Modal Haplotype (AMH) within haplogroup R1xR1a (R1b3), typical of Ireland and Western Britain, is very common in the NSG. An additional lineage (Helgason et al. 2000b), centred on the STR motif 13, 12, 23, 10, 12, 13, 10, 17, 13*17 for DYS19, 390, 391, 392, 393, 389I, 389II, 385A*B respectively, is found in Scandinavia and Iceland but not in the British Isles and this was also not observed in the NSG.

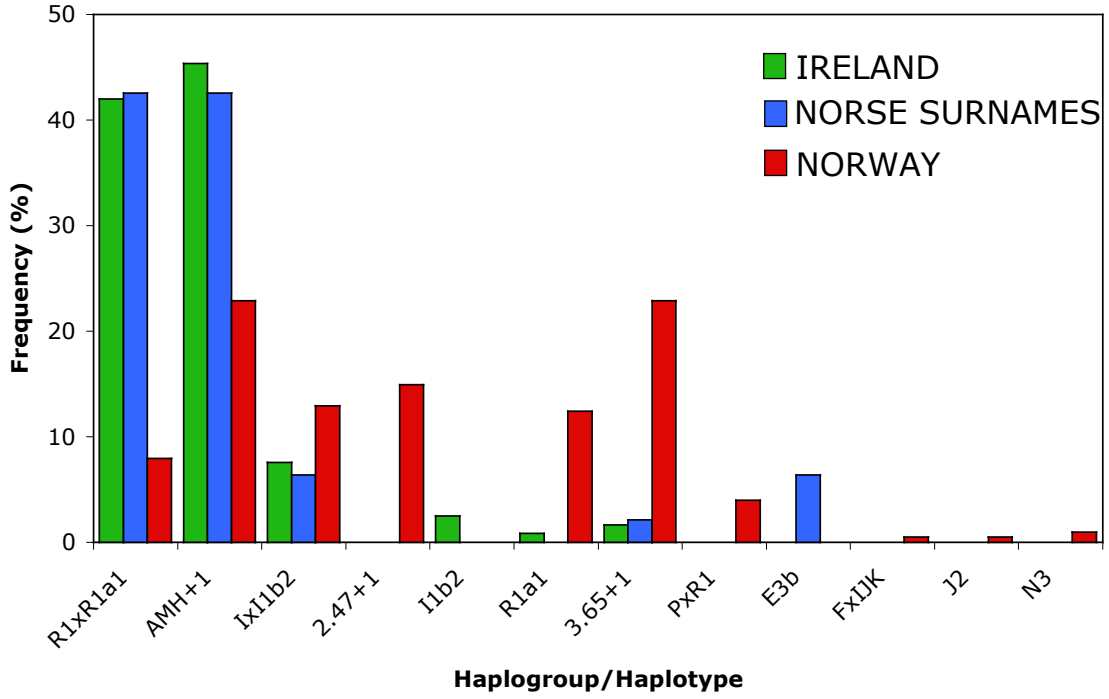


Figure 6.2 *Y chromosome haplogroup frequencies in the Irish Norse surname group, Ireland and Norway. The latter two are potential parental populations for the NSG. Many common Norwegian haplotypes and haplogroups are absent or infrequent in the NSG. Conversely, R1xR1a (R1b3) and its AMH subtype are frequent, indicating greater NSG affinity to Ireland. Comparative Irish and Norwegian data from Capelli et al. (2003). The modal STR profiles for specific lineages are (DYS19, DYS388, DYS390, DYS391, DYS392, DYS393): AMH (14, 12, 24, 11, 13, 13); 2.47 (14, 14, 22, 10, 11, 13); 3.65 (16, 12, 25, 11, 11, 13).*

6.3.2 Norse Surname Group Population Affinities

The general population affinities of the NSG were examined in the context of potential parental populations and regional geographic variation. Analysis was carried out separately for UEP and STR Y chromosome diversity through the calculation of inter-population genetic distances, which were then summarised using MDS (**Figure 6.3**). Despite independent datasets, broadly similar pictures emerge from both UEP and STR variation. The Irish and Scandinavian populations are strongly differentiated with the NSG displaying an obvious greater affinity to Ireland compared to anywhere in Scandinavia. In contrast, the Scottish Isles in UEP variation (**Figure 6.3A**) and Iceland in STR variation (**Figure 6.3B**), both areas of well-attested Viking activity, are shifted toward the Scandinavian pole. This suggests that, at least substantial, Norse introgression should by readily detectable using this approach.

In addition to the genetic distinctiveness of Ireland from Scandinavia, there is also considerable heterogeneity within Scandinavia and the British Isles. Considering STR diversity (**Figure 6.3B**), Sweden and Norway are not significantly different from each other but both differ significantly from their close neighbour Denmark. Interestingly, while the Danish Y chromosome pool is distinct from its near neighbours, it is not so from England. Consistent with this observation, English populations (particularly Eastern areas) occupy intermediate positions between groupings from the Western/Northern British Isles and those from continental Europe (see also Chapter 2).

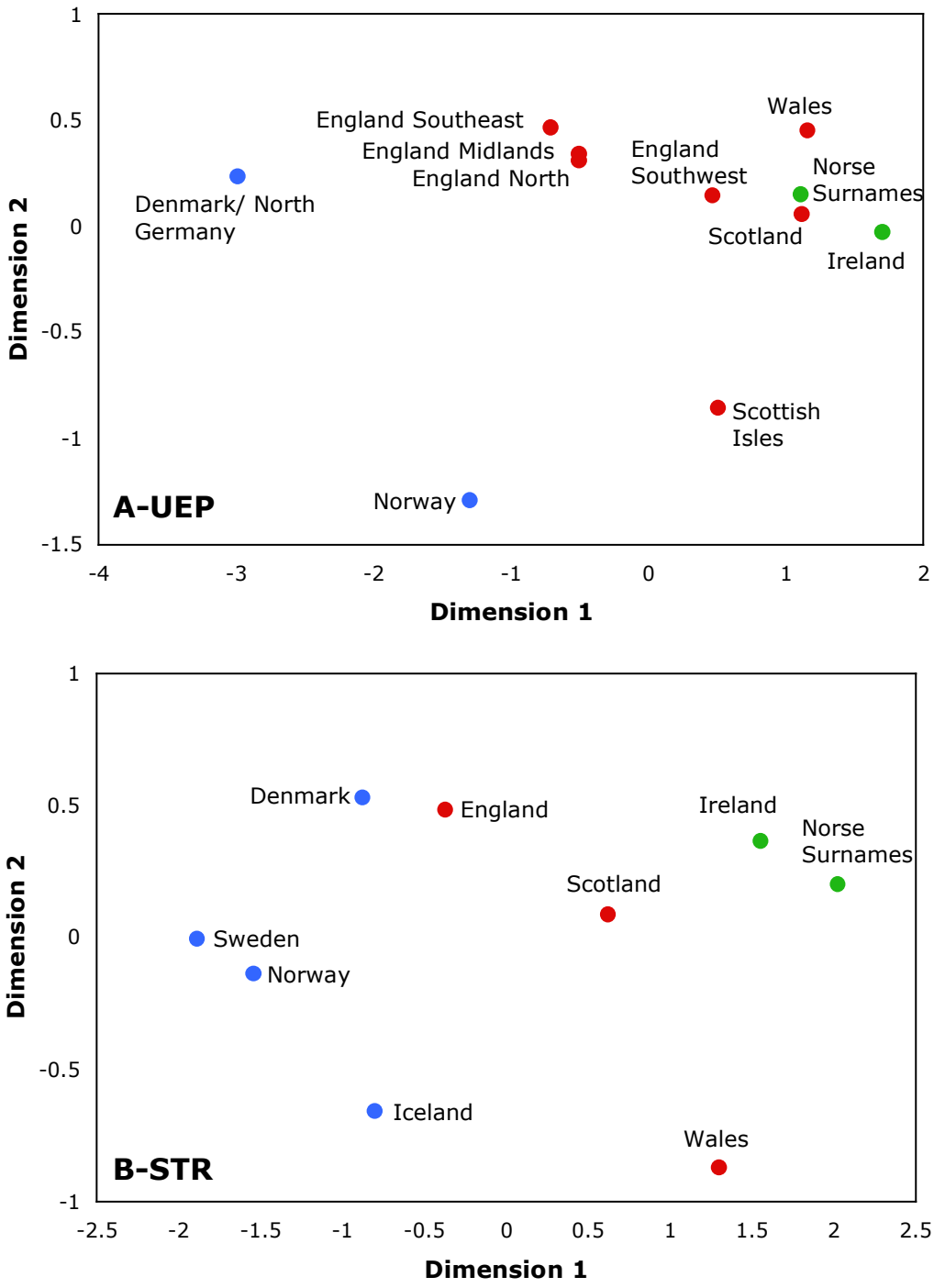


Figure 6.3 *MDS plots showing the relationship of the NSG to potential parental populations and other areas of the North Atlantic region, using (**A**) UEP and (**B**) STR Y chromosome diversity. There is a clear separation across Dimension 1, in both cases, of most British Isles populations (shown in red) and those from Scandinavia (blue). The close affinity of the NSG and Ireland (indicated in green) compared to Scandinavia is also apparent and suggests predominantly Irish ancestry in this surname group. Each plot preserves ~99% of the original variation.*

6.3.3. Scandinavian Admixture Estimates

The proportion of NSG ancestry derived from Ireland was quantified and results compared and contrasted using six different admixture estimators, each of which considers different factors and complexities in its execution. Quantative admixture analysis was carried out independently for UEP and STR Y chromosome variation, allowing a parallel investigation of the relative information content of different marker types. A summary of the admixture calculations is presented graphically in **Figure 6.4**.

The overall trend is clear and decisive with nearly all methods, whether using UEP or STR markers, supporting an overwhelming majority contribution from Ireland to the NSG. By corollary, the Scandinavian influence is minor and estimates rarely exceed 10%. Even then, when confidence intervals are considered most estimates are not inconsistent with complete Irish ancestry. Indeed, the unconstrained estimators (m_R, m_C and m_Y**)** often returned notional Irish ancestry proportions greater than 1. Interpreting these as a 100% Irish contribution, the median estimate across the full permutation of variables (method, marker and parental population), some 51 scenarios, is 97.5% Irish ancestry with a standard deviation (SD) across estimates of 13.3%.

Figure 6.4 (overleaf) *Quantitative admixture estimates for the Norse surname group. Analysis was carried out for* ***(A)*** *UEP and* ***(B)*** *STR Y chromosome diversity separately. Each section shows the estimated Irish (first parental population or P_1) ancestry proportion and associated confidence intervals in up to six estimators using different Scandinavian parental populations (P_2). In one STR case (*), the general Irish population was replaced as P_1 by a Gaelic Irish surname group, while a combined Scandinavian population served as P_2. NA is entered where the method was not appropriate or possible. The number of alleles involved in each calculation is given in brackets after the Scandinavian P_2. Most estimates provide consistent support for an Irish ancestry proportion of 90% or more. However, UEP based estimates for m_L suggest a minority Irish contribution and are clearly exceptional in* ***(C)*** *the distribution of estimates over the 51 permutations of method, parental population and marker type. Overall, the median estimate is 97.5% Irish ancestry (interpreting estimates greater than 1 as 100%) (SD of 13.3%) or 98.8% when the m_L outliers are excluded (SD 5%).*

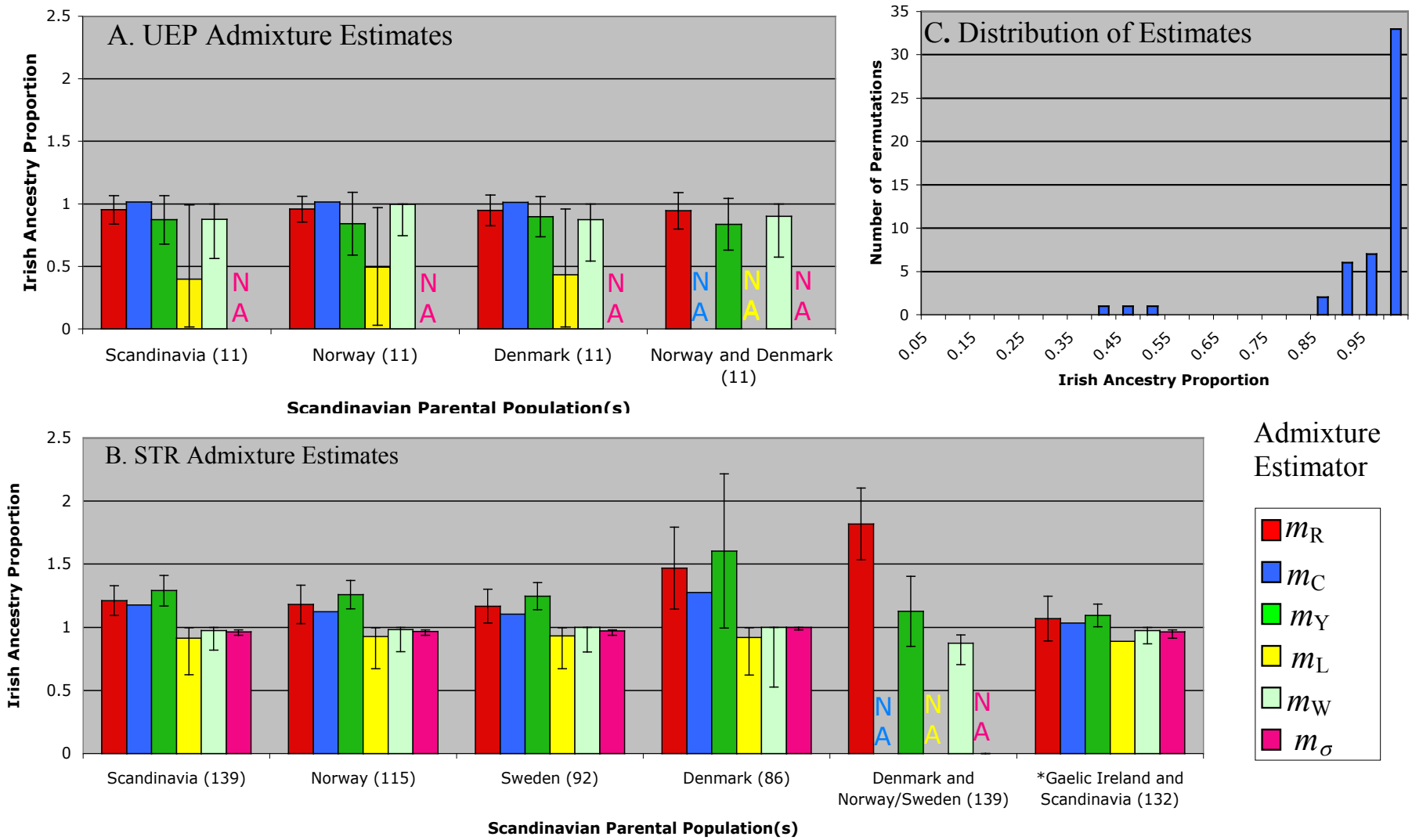


Figure 6.4 *Quantitative admixture estimates for the Norse surname group.* (**See preceding page for full legend**)

6.3.4 Parental Population in Admixture Estimates

For each method and marker type, Ireland was held constant as the first parental population. Given the heterogeneity of Scandinavia, incorrect parental population specification from here could confound accurate estimation of Viking introgression. This uncertainty was extensively explored using sample populations from different areas of Scandinavia either individually, combined, or simultaneously (when method allowed). However, Scandinavian parental population choice did not substantially alter the outcome with most permutations indicating overwhelming Irish ancestry.

None of the admixture methods can account for changes in allele frequency due to migration into the parental population after the hybridisation event. Using information on surname origin, the Irish STR parental population was refined and reduced to include only those with a Gaelic Irish origin. As these surnames developed in Ireland mainly in the 10th to 12th centuries, they may better reflect the Y chromosome composition of the Irish population at the time of Viking incursion, prior to the large-scale historical Anglo-Scottish migrations. However, this adjustment made little impact, with all estimators returning Irish ancestry of at least 89% (**Figure 6.4**)

6.3.5 Marker Type and Method Choice in Admixture Estimates

Overall, there is little variation in results from different methods and markers. However, the UEP based maximum likelihood m_L estimates incongruously indicate minority Irish ancestry in the NSG and are clearly extreme outliers in the distribution of estimates (**Figure 6.4C**). Excluding them increases the median Irish ancestry estimate to 98.8% and tightens the SD considerable from 13% to 5%. The same method using the more numerous STR defined alleles returns estimates of ~90% Irish ancestry, in full agreement with other methods. This finding was replicated in a separate analysis of STR variation in the same samples used for UEP based estimates (data from Capelli et al. 2003).

Median point estimates may give a poor representation of the results over the entire MCMC run of m_L calculation (Chikhi et al. 2001). This observation is reflected here in the huge confidence intervals for the UEP based estimates, which encompass virtually the entire range of 0 to 100% Irish ancestry. The complete distribution of results over 50,000 iterations was therefore considered in the UEP and STR based estimates using exactly the same sample populations for both sets of calculations (**Figure 6.5**). It is clear that UEP defined alleles could not reach convergence or settle on a consistent estimate. In stark contrast, corresponding STR defined alleles show an unambiguous tendency toward a high Irish ancestry proportion estimate. Significantly longer runs (up 250,00 MCMC iterations) were not found to aid UEP estimate convergence.

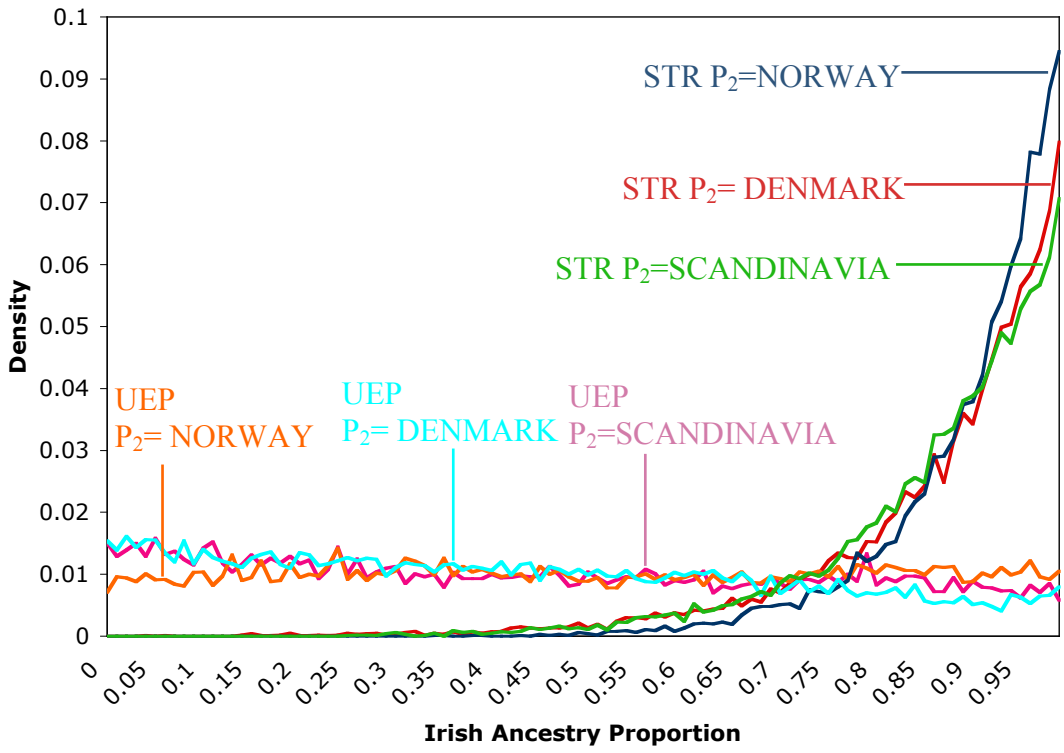


Figure 6.5 *Markov chain Monte Carlo (MCMC) convergence of the m_L admixture estimator. Each line shows the posterior probability density function for a run (over 50,000 iterations) of different marker types and Scandinavian parental populations (P_2). Sample population difference has been removed as a variable by using identical datasets for UEP and STR estimates (data from Capelli et al. 2003). All three UEP (11 allele) based estimates show little convergence. In contrast, STR based calculations (109 to 160 alleles depending on the P_2) show an unambiguous tendency towards high Irish ancestry proportions.*

6.4 DISCUSSION

6.4.1 Nature of Hiberno-Norse Society

Despite the putative Viking or Norse origin of the surnames examined, there is little evidence for a substantial Scandinavian genetic contribution in the Y chromosomes of this sample. Whether through phylogeography of individual lineages or analysis of group affinity, the results emphasize the continuity of Norse surname bearers with the general Irish population. These indications were confirmed and quantified through an extensive exploration of admixture estimates incorporating different marker types, different Scandinavian and Irish parental populations and six methodological frameworks. Indeed some of these returned notionally negative Scandinavian contributions conservatively interpreted as zero. However, strong negative findings for the estimator m_Y may sometimes be interpreted as a rejection of the initial hypothesis (Bertorelle and Excoffier 1998); here that the NSG represents a Scandinavian/ Irish hybrid. A substantial Irish component was not unexpected given historical evidence of assimilation between native Irish and Norse settlers in Hiberno-Norse towns. However, the Norse genetic component of the Hiberno-Norse cultural equation is absent.

The carriage of a Norse-derived surname suggests an ancestral paternal cultural influence, however it seems that this was not accompanied by and therefore unexplainable through a corresponding paternal ethnic bond. It follows, despite Viking prominence in Irish history, that the number of Scandinavian migrants to Ireland was small compared to the population of the settlements. This is consistent with an elite dominance model of settlement where the leaders and upper echelons of Norse colonial society were Scandinavian in origin or ancestry (indicated by historical sources) but most of the inhabitants were native Irish. These adopted aspects of Norse culture and fused it with their own to form distinctive Hiberno-Norse communities.

Such a conclusion is in agreement with historical evidence from, for example, the limited impact of Norse on the Irish language (Ó Cróinín 1995, Chapter 9). However, it is also founded on the assumption that Norse names were restricted to Hiberno-Norse settlements. There is some evidence to suggest they had a degree of popularity in a native Irish context (O Cuív 1988). Such exchange may mean that Norse names were no longer primarily associated with colonial societies by the time of widespread surname adoption. Alternatively, the original personal names/nicknames could have been introduced through a Norse female or indeed any other non-patrilineal ancestor leaving no Y chromosome trace when these names eventually became incorporated into hereditary surnames. Finally, the perturbed nature of Irish surname history means that similar or identical names can have more than one source. For example, the name Sweetman can be of Norse origin or an English toponymic from Swettenham (McLysaght 1985b). The multiple occurrence of E3b Y chromosomes in the sample population hints at this possibility since it is more typical of England than either Ireland or Scandinavia (Capelli et al. 2003). However, the variety of names used and the inclusion of multiple bearers should minimise the impact of this source of error. The foregoing explanations are not mutually exclusive and all may contribute to the overwhelming Irish ancestry observed in the putative Norse surname group.

6.4.2 Quantitative Admixture Estimators

The admixture estimators examined here encompassed a range of considerations and computational frameworks but none alone accounts for all potential confounding factors. The choice of the most appropriate single estimator will depend on the expected relative importance of these factors in the circumstance under consideration. For example with recent admixture or UEP data, the effect of mutation is of less importance than drift. Used together multiple estimators provide a means of circumventing potential limitations to provide broad based foundation for conclusion.

Some current admixture calculation methods provide a challenge in meaningfully combining UEP and STR variation (see Helgason et al. 2000b). Despite the potential limitations, of a relatively small number of alleles (haplogroups) with UEPs or the 'noise' associated with highly variable multiple STR haplotypes, both marker types generally return consistent and sensible results. The maximum likelihood method of Chikhi et al. (2001) using UEP Y chromosome diversity is a notable exception in indicating majority Scandinavian ancestry in the Norse surname group. Considering the incongruity of these finding with 48 other permutations and the absence of MCMC convergence, it seems reasonable to assume that this method was incorrect and using the median value strongly underestimated the proportion derived from the first parental population, in this case Ireland.

UEP diversity is often favoured in population comparisons because of stability and obvious structure. However, on the basis of these results, compound STR haplotypes are at least as informative, if not more so, in discerning population affinities. This is hardly surprisingly given that distinct UEP lineages are often well reflected in underlying STR diversity (Bosch et al. 1999; and see also Chapter 4 and 5). In addition, STRs have potentially wider application in measuring diversity and finer-scale intra-haplogroup relationships. The use of a small number of relatively stable STRs, which can be conveniently typed in a single multiplex, may represent an under utilised tool in macro-Y chromosome population genetics.

6.4.3 m_L and the Neolithic Revolution

The poor resolution of the maximum likelihood method m_L has ramifications in the continuing debate into the genetic legacy modern Europeans derive from migrants associated with the spread of agriculture. Competing theories either emphasize the genetic continuity of modern Europeans with their Palaeolithic ancestors (cultural

diffusion Model) (Zvelebil 2000) or support substantial replacement of the pre-existing population by Neolithic newcomers (demic diffusion model) (Ammerman and Cavalli-Sforza 1984). mtDNA and Y chromosome studies have each estimated the Neolithic component at approximately ~20% (Richards et al. 2000; Semino et al. 2000). However, subsequent analysis of the same Y chromosome data using the m_L estimator supported a majority Neolithic ancestry (Chikhi et al. 2002). In common with the present study, this estimate was based on UEP defined Y chromosome haplogroups, with the Basque Country taken to represent the 'Palaeolithic' first parental population (P_1). The dramatic variation in estimated Neolithic contributions from the same data may be explained by a tendency of m_L to underestimate the P_1 contribution using UEP defined haplogroups. Similarly, the large point estimate and wide confidence intervals for Viking ancestry in the general Irish population (Capelli et al. 2003) may be resolved as an artefact of the same method.

6.4.4 Conclusions

Broad-based analysis of Y chromosome diversity detected little evidence of paternal Scandinavian ancestry in Irish men with a putative Norse cultural connection through their surname. In this instance surnames were a poor guide to a corresponding paternal genetic ancestry, suggesting in turn that Norse culture was adopted by indigenous Irish in Hiberno-Norse settlements and perhaps more widely in native Irish society. This is consistent with a model of Viking Scandinavian settlement in Ireland that was culturally quite strong and influential in the absence of a large number of migrants. The contrast may be explained, in part at least, by an elite dominance model of settlement where the upper echelon of colonial society, which was of Norse origin or ancestry, presided over a community that was primarily composed of indigenous Irish.

REFERENCES

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogvali EL, Kivisild T, Bandelt H, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A (2004) The Molecular Dissection of mtDNA Haplogroup H Confirms That the Franco-Cantabrian Glacial Refuge Was a Major Source for the European Gene Pool. Am J Hum Genet 75:910-918

Adams GB (1979) Prolegomena to the Study of Surnames in Ireland. Nomina 3:81-94

Ammerman AJ, Cavalli-Sforza LL (1984) The Neolithic Transition and the Genetics of Populations in Europe. Princeton University Press, Princeton, New Jersey

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457-65

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 28:e8

Baasner A, Schafer C, Junge A, Madea B (1998) Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. Forensic Sci Int 98:169-78

Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, Papiha SS, Villems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, Jorde LB (2001) Genetic evidence on the origins of Indian caste populations. Genome Res 11:994-1004

Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BV, Rasanayagam A, Hammer MF (1998) Female gene flow stratifies Hindu castes. Nature 395:651-2

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37-48

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743-53

Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V (2002) The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet 71:1150-60

Barbujani G, Bertorelle G, Chikhi L (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. Am J Hum Genet 62:488-492

Bardon J (1992) A History of Ulster. Blackstaff, Belfast

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. Nat Rev Genet 5:251-61

Beddoe J (1885) The Races of Britain: A Contribution to the anthropology of Western Europe. Arrowsmith, Bristol

Behar DM, Thomas MG, Skorecki K, Hammer MF, Bulygina E, Rosengarten D, Jones AL, Held K, Moses V, Goldstein D, Bradman N, Weale ME (2003) Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. Am J Hum Genet 73:768-79

Bell R (1988) The Book of Ulster Surnames. Blackstaff, Belfast

Belledi M, Poloni ES, Casalotti R, Conterio F, Mikerezi I, Tagliavini J, Excoffier L (2000) Maternal and paternal lineages in Albania and the genetic structure of Indo-European populations. Eur J Hum Genet 8:480-6

Bendall KE, Sykes BC (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. Am J Hum Genet 57:248-56

Bennett P, Segurado R, Jones I, Bort S, McCandless F, Lambert D, Heron J, Comerford C, Middle F, Corvin A, Pelios G, Kirov G, Larsen B, Mulcahy T, Williams N, O'Connell R, O'Mahony E, Payne A, Owen M, Holmans P, Craddock N, Gill M (2002) The Wellcome trust UK-Irish bipolar affective disorder sibling-pair genome screen: first stage report. Mol Psychiatry 7:189-200

Bertorelle G, Excoffier L (1998) Inferring admixture proportions from molecular data. Mol Biol Evol 15:1298-311

Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P, Comas D (1995) Human mitochondrial DNA variation and the origin of Basques. Ann Hum Genet 59: 63-81

Bosch E, Calafell F, Santos FR, Perez-Lezaun A, Comas D, Benchemsi N, Tyler-Smith C, Bertranpetit J (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. Am J Hum Genet 65:1623-38

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

Bosch E, Rosser ZH, Norby S, Lynnerup N, Jobling MA (2003) Y-chromosomal STR haplotypes in Inuit and Danish population samples. Forensic Sci Int 132:228-32

Bradley J (1988) The interpetation of Scandinavian Settlement in Ireland. In: Bradley J (ed) Settlement and Society in Medieval Ireland. Boethius Press, Kilkenny

Brown K (2002) Tangled roots? Genetics meets genealogy. Science 295:1634-5

Byrne FJ (2001) Irish Kings and High-Kings. Four Courts Press, Dublin

Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjieva L (1996) From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. Ann Hum Genet 60:35-49

Cali F, Le Roux MG, D'Anna R, Flugy A, De Leo G, Chiavetta V, Ayala GF, Romano V (2001) MtDNA control region and RFLP data for Sicily and France. Int J Legal Med 114:229-31

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

Carvajal-Carmona LG, Soto ID, Pineda N, Ortiz-Barrientos D, Duque C, Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, Ruiz-Linares A (2000) Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. Am J Hum Genet 67:1287-95

Cavalli-Sforza LL, Menozzi P, A. P (1994) The history and geography of human genes. Princeton University Press, Princeton NJ

Cavalli-Sforza LL (1998) The DNA revolution in population genetics. Trends Genet 14:60-5

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

Chakraborty R, Kamboh MI, Nwankwo M, Ferrell RE (1992) Caucasian genes in American blacks: new data. Am J Hum Genet 50:145-55

Chikhi L, Bruford MW, Beaumont MA (2001) Estimation of admixture proportions: a likelihood-based approach using Markov chain Monte Carlo. Genetics 158:1347-62

Chikhi L, Nichols RA, Barbujani G, Beaumont MA (2002) Y genetic data support the Neolithic demic diffusion model. Proc Natl Acad Sci U S A 99:11008-13

Choisy M, Franck P, Cornuet JM (2004) Estimating admixture proportions with microsatellites: comparison of methods based on simulated data. Mol Ecol 13:955-68

Colantonio SE, Lasker GW, Kaplan BA, Fuster V (2003) Use of surname models in human population biology: a review of recent developments. Hum Biol 75:785-807

Coles B (1998) Doggerland: a speculative survey. Proc Prehist Soc 64:45-81

Connolly SJ (ed) (2001) The Oxford Companion to Irish History. Oxford University Press, Oxford

Cordaux R, Aunger R, Bentley G, Nasidze I, Sirajuddin SM, Stoneking M (2004) Independent origins of Indian caste and tribal paternal lineages. Curr Biol 14:231-5

Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum Genet 60:331-50

Croke DT (2000) Genetic archaeology and the origins of the Irish population. Ir J Med Sci 169:258-61

Cunliffe B (1997) The Ancient Celts. Oxford University Press, Oxford

Cunliffe B (2001) Facing the Ocean: The Atlantic and its People. Oxford University Press, Oxford

Darwin CH (1875) Marriages between first cousins in England and their effects. J. Statist. Soc. 38:153-184

Davies N (2000) The Isles - A History. Papermac, London

Davison A, Birks JD, Brookes RC, Messenger JE, Griffiths HI (2001) Mitochondrial phylogeography and population history of pine martens Martes martes compared with polecats Mustela putorius. Mol Ecol 10:2479-88

Dawson GW (1964) The Frequencies of the ABO and Rh (D) Blood Groups in Ireland from a Sample of 1 in 18 of the Population. Ann Hum Genet 28:49-59

Decorte R (1996) Genetic analysis of single hair shafts by automated sequence analysis of the mitochondrial d-loop region. Adv Forensics Haemogenet 6:17-19

Di Benedetto G, Erguven A, Stenico M, Castri L, Bertorelle G, Togan I, Barbujani G (2001) DNA diversity and population admixture in Anatolia. Am J Phys Anthropol 115:144-56

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci U S A 91:3166-70

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci U S A 88

Dimo-Simonin N, Grange F, Taroni F, Brandt-Casadevall C, Mangin P (2000) Forensic evaluation of mtDNA in a population from south west Switzerland. Int J Legal Med 113:89-97

Dolukhanov PM (1993) Foraging and farming groups in north-eastern and north-western Europe: identity and interaction. In: Chapman D, Dolukhanov P (eds) Cultural transformations and interactions in eastern Europe. Aldershot, Avebury, pp 122-145

Dubut V, Chollet L, Murail P, Cartault F, Beraud-Colomb E, Serre M, Mogentale-Profizi N (2004) mtDNA polymorphisms in five French groups: importance of regional sampling. Eur J Hum Genet 12:293-300

Duffy S (1997) Ireland in the Middle Ages. Gill and Macmillan, Dublin

Duffy S (2000) The Concise History of Ireland. Gill and Macmillan, Dublin

Dumolin-Lapegue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. Genetics 146:1475-1487

Dupanloup I, Bertorelle G (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. Mol Biol Evol 18:672-5

Dupuy BM, Stenersen M, Egeland T, Olaisen B (2004) Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci. Hum Mutat 23:117-24

Dutilleul P (1993) Modifying the t test for assessing correlation between two spatial process. Biometrics 49:305-314

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435-45

Elliott C, Brodwin P (2002) Identity and genetic ancestry tracing. Bmj 325:1469-71

Esmonde-Cleary AS (1989) The Ending of Roman Britain. Batsford, London

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-91

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567-87

Flores C, Maca-Meyer N, Perez JA, Gonzalez AM, Larruga JM, Cabrera VM (2003) A predominant European ancestry of paternal lineages from Canary Islanders. Ann Hum Genet 67:138-52

Forster P, Cali F, Rohl A, Metspalu E, D'Anna R, Mirisola M, De Leo G, Flugy A, Salerno A, Ayala G, Kouvatsi A, Villems R, Romano V (2002) Continental and subcontinental distributions of mtDNA control region types. Int J Legal Med 116:99-108

Forster P, Toth A (2003) Toward a phylogenetic chronology of ancient Gaulish, Celtic, and Indo-European. Proc Natl Acad Sci U S A 100:9079-84

Foster EA, Jobling MA, Taylor PG, Donnelly P, de Knijff P, Mieremet R, Zerjal T, Tyler-Smith C (1998) Jefferson fathered slave's last child. Nature 396:27-8

Francalacci P, Bertranpetit J, Calafell F, Underhill PA (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. Am J Phys Anthropol 100:443-60

Gamble C, Davies W, Pettitt P, Richards M (2004) Climate change and evolving human diversity in Europe during the last glacial. Philos Trans R Soc Lond B Biol Sci 359:243-53

Genest P (1973) [Hereditary transmission for 300 years of a satellited Y chromosome in a family]. Ann Genet 16:35-8

Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K (1994) Identification of the remains of the Romanov family by DNA analysis. Nat Genet 6:130-5

Giraldo A, Martinez I, Guzman M, Silva E (1981) A family with a satellited Yq chromosome. Hum Genet 57:99-100

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

Gonzalez AM, Brehm A, Perez JA, Maca-Meyer N, Flores C, Cabrera VM (2003) Mitochondrial DNA affinities at the Atlantic fringe of Europe. Am J Phys Anthropol 120:391-404

Gouin N, Grandjean F, Pain S, Souty-Grosset C, Reynolds J (2003) Origin and colonization history of the white-clawed crayfish, Austropotamobius pallipes, in Ireland. Heredity 91:70-7

Grenham J (1999) Tracing your Irish Ancestors. Gill and Macmillan, Dublin

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

Hamblin MT, Di Rienzo A (2000) Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. Am J Hum Genet 66:1669-79

Haywood J (1995) The Penguin Historical Atlas of the Vikings. Penguin, London

Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, Ward R, Sykes B (2001) mtDna and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. Am J Hum Genet 68:723-37

Helgason A, Hrafnkelsson B, Gulcher JR, Ward R, Stefansson K (2003) A population wide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. Am J Hum Genet 72:1370-88

Helgason A, Sigurethardottir S, Gulcher JR, Ward R, Stefansson K (2000a) mtDNA and the origin of the Icelanders: deciphering signals of recent population history. Am J Hum Genet 66:999-1016

Helgason A, Sigurethardottir S, Nicholson J, Sykes B, Hill EW, Bradley DG, Bosnes V, Gulcher JR, Ward R, Stefansson K (2000b) Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. Am J Hum Genet 67:697-717

Hill EW (1999) A Molecular Genetic Investigation of Irish Origins. Ph.D thesis, University of Dublin

Hill EW, Jobling MA, Bradley DG (2000) Y-chromosome variation and Irish origins. Nature 404:351-2

Hofmann S, Jaksch M, Bezold R, Mertens S, Aholt S, Paprotta A, Gerbitz KD (1997) Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. Hum Mol Genet 6:1835-46

Hofreiter M, Serre D, Poinar HN, Kuch M, Paabo S (2001) Ancient DNA. Nat Rev Genet 2:353-9

Housley RA, Gamble CS, M. S, P P (1997) Radiocarbon evidence for the late glacial human recolonisation of northern Europe. Proc Prehist Soc 63:25-44

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

Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708-13

Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nat Genet 12:417-20

James S (1999) The Atlantic Celts. British Museum Press, London

Jobling MA (2001) In the name of the father: surnames and genetics. Trends Genet 17:353-7

Jobling MA, Tyler-Smith C (2000) New uses for new haplotypes the human Y chromosome, disease and selection. Trends Genet 16:356-62

Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598-612

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc Natl Acad Sci U S A 75:2868-72

Kittles RA, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, Goldman D, Long JC (1999) Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. Am J Phys Anthropol 108:381-99

Kouvatsi A, Karaiskou N, Apostolidis A, Kirmizidis G (2001) Mitochondrial DNA sequence variation in Greeks. Hum Biol 73:855-69

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Paabo S (1997) Neandertal DNA sequences and the origin of modern humans. Cell 90:19-30

Lahermo P, Laitinen V, Sistonen P, Beres J, Karcagi V, Savontaus ML (2000) MtDNA polymorphism in the Hungarians: comparison to three other Finno-Ugric-speaking populations. Hereditas 132:35-42

Larruga JM, Diez F, Pinto FM, Flores C, Gonzalez AM (2001) Mitochondrial DNA characterisation of European isolates: the Maragatos from Spain. Eur J Hum Genet 9:708-16

Long JC (1991) The genetic structure of admixed populations. Genetics 127:417-28

Loogvali EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, et al. (2004) Disuniting Uniformity: A Pied Cladistic Canvas of mtDNA Haplogroup H in Eurasia. Mol Biol Evol 21:2012-2021

Lutz S, Weisser HJ, Heizmann J, Pollak S (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. Int J Legal Med 111:67-77

Mac Niocaill G (1972) Ireland before the Vikings. Gill and Macmillian, Dublin

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B, Torroni A (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64:232-49

MacShamhraín A (2002) The Vikings - An Illustrated History. Wolfhound Press, Dublin

Malyarchuk BA, Derenko MV (2001) Mitochondrial DNA variability in Russians and Ukrainians: implication to the origin of the Eastern Slavs. Ann Hum Genet 65:63-78

Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Wozniak M, Miscicka-Sliwka D (2002) Mitochondrial DNA variability in Poles and Russians. Ann Hum Genet 66:261-83

Manni F, Toupance B, Sabbagh A, Heyer E (2004) New method for surname studies of ancient patrilineal population structures, and possible application to improvement of Y-chromosome sampling. Am J Phys Anthropol 126:214-218

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220

Mascheretti S, Rogatcheva MB, Gunduz I, Fredga K, Searle JB (2003) How did pygmy shrews colonize Ireland? Clues from a phylogenetic analysis of mitochondrial cytochrome b sequences. Proc R Soc Lond B Biol Sci 270:1593-9

Mathews A (1968) Origin of the surname McGuinness with a short history of the sept. A. Mathews, Dublin

McLysaght E (1982) More Irish Families. Irish Academic Press, Dublin

McLysaght E (1985a) Irish Families: Their Names, Arms and Origins. Irish Academic Press, Dublin

McLysaght E (1985b) The Surname of Ireland. Irish Academic Press, Dublin

Megaw JVS, Megaw MR (1996) Ancient Celts and modern ethnicity. Antiquity LXX:175-181

Megaw JVS, Megaw MR (1998) The mechanism of (Celtic) dreams ? : a partial repsonse to our critics. Antiquity LXXII:432-435

Meyer S, Weiss G, von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152:1103-10

Mogentale-Profizi N, Chollet L, Stevanovitch A, Dubut V, Poggi C, Pradie MP, Spadoni JL, Gilles A, Beraud-Colomb E (2001) Mitochondrial DNA sequence diversity in two groups of Italian Veneto speakers from Veneto. Ann Hum Genet 65:153-66

Moore LT (2004) Y Chromosomes and Irish Phylogeogrpahy. Ph.D thesis, University of Dublin

Nicholls KW (2003) Gaelic and Gaelicized Ireland in the Middle Ages. Lilliput, Dublin

Ó Corraín D (1972) Ireland before the Normans. Gill and Macmillan, Dublin

Ó Cróinín D (1995) Early Medieval Ireland 400-1200. Longman, Essex

Ó Cuív B (1988) Personal Names as an indicator of relations between native Irish and settlers in the Viking period. In: Bradley J (ed) Settlement and Society in Medieval Ireland. Boethius Press, Kilkenny

Ó Donnabhaín B (2000) An Appalling Vista? The Celts and the archaeology of later prehistoric Ireland. In: Desmond A (ed) New Agendas in Irish Prehisory. Wordwell, Dublin, pp 189-196

O'Donnell KA, O'Neill C, Tighe O, Bertorelle G, Naughten E, Mayne PD, Croke DT (2002) The mutation spectrum of hyperphenylalaninaemia in the Republic of Ireland: the population history of the Irish revisited. Eur J Hum Genet 10:530-8

Ó Muraíle N (Mac Fhirbhisigh D) (2003) The Great Book of Irish Genealogies. De Búrca, Dublin

Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet 29:20-1

Opdal SH, Rognum TO, Vege A, Stave AK, Dupuy BM, Egeland T (1998) Increased number of substitutions in the D-loop of mitochondrial DNA in the sudden infant death syndrome. Acta Paediatr 87:1039-44

Orekhov V, Poltoraus A, Zhivotovsky LA, Spitsyn V, Ivanov P, Yankovsky N (1999) Mitochondrial DNA sequence diversity in Russians. FEBS Lett 445:197-201

Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, Goodwin W (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature 404:490-3

Pääbo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M (2004) Genetic Analyses From Ancient DNA. Annu Rev Genet 38:645-679

Parson W, Parsons TJ, Scheithauer R, Holland MM (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. Int J Legal Med 111:124-32

Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Borresen-Dale AL, Underhill PA (2002) Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. Eur J Hum Genet 10:521-9

Pereira L, Prata MJ, Amorim A (2000) Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. Ann Hum Genet 64:491-506

Perez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martinez-Arias R, Clarimon J, Fiori G, Luiselli D, Facchini F, Pettener D, Bertranpetit J (1999) Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. Am J Hum Genet 65:208-19

Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P (1999) Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. Int J Legal Med 112:291-8

Piercy R, Sullivan KM, Benson N, Gill P (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int J Legal Med 106:85-90

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-59

Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, Paabo S (1994) Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. Biol Chem Hoppe Seyler 375:837-40

Raftery B (1994) Pagan Celtic Ireland: the enigma of the Irish Iron Age. Thames and Hudson, London

Relethford JH, Crawford MH (1995) Anthropometric variation and the population history of Ireland. Am J Phys Anthropol 96:25-38

Renfrew C (1987) Archaeology and Language. Jonathan Cape, London

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

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 M, Macaulay V, Sykes B, Pettitt P, Hedges R, Forster P, Bandelt HJ (1997) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 61:247-251

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, Sykes B (1998) Reply to Barbujani et al. Am J Hum Genet 62:491-492

Richards MB, Macaulay VA, Bandelt HJ, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62:241-60

Roberts DF, Hiorns RW (1965) Methods of analysis of the genetic composition of a hybrid population. Human Biology 37:38-43

Rohl A, Brinkmann B, Forster L, Forster P (2001) An annotated mtDNA database. Int J Legal Med 115:29-39

Rolf B, Meyer E, Brinkmann B, de Knijff P (1998) Polymorphism at the tetranucleotide repeat locus DYS389 in 10 populations reveals strong geographic clustering. Eur J Hum Genet 6:583-8

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 MS (2001) PASSAGE. Pattern Analysis, Spatial Statistics, and Geographic Exegesis. Version 1.1.1.3, Department of Biology, Arizona State University, Arizona

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

Rousselet F, Mangin P (1998) Mitochondrial DNA polymorphisms: a study of 50 French Caucasian individuals and application to forensic casework. Int J Legal Med 111:292-298

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928-33

Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjaerg L, Gedde-Dahl T, Issel-Tarver L, DiRienzo A, Paabo S (1995) Genes and languages in Europe: an analysis of mitochondrial lineages. Genome Res 5:42-52

Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Paabo S (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci U S A 93:12035-9

Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo A (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. Eur J Hum Genet 6:365-75

Schaffner SF (2004) The X chromosome in population genetics. Nat Rev Genet 5:43-51

Schmitz RW, Serre D, Bonani G, Feine S, Hillgruber F, Krainitzki H, Paabo S, Smith FH (2002) The Neandertal type site revisited: interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. Proc Natl Acad Sci U S A 99:13342-7

Schneider S, Roessli D, L. E (2000) ARLEQUIN ver. 2000: software for population genetic analysis, Genetics and Biometric Laboratory, University of Geneva, Switzerland

Seddon JM, Santucci F, Reeve NJ, Hewitt GM (2001) DNA footprints of European hedgehogs, Erinaceus europaeus and E. concolor: Pleistocene refugia, postglacial expansion and colonization routes. Mol Ecol 10:2187-98
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

Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Mennecier P, Hofreiter M, Possnert G, Paabo S (2004) No evidence of Neandertal mtDNA contribution to early modern humans. PLoS Biol 2:313-17

Shoemaker JS, Painter IS, Weir BS (1999) Bayesian statistics in genetics: a guide for the uninitiated. Trends Genet 15:354-8

Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. Mol Biol Evol 12:914-20

Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics 134:983-93

Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G (2000a) Geographic patterns of mtDNA diversity in Europe. Am J Hum Genet 66:262-78

Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G (2000b) Reconstruction of prehistory on the basis of genetic data. Am J Hum Genet 66:1177-9

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-62

Smouse PE, Long JC, Sokal RR (1986) Multiple Regression and correlation extension of the Mantel test of matrix correspondence. Systematic Zoology 35:627-632

Sokal RR, Harding RM, Oden NL (1989) Spatial patterns of human gene frequencies in Europe. Am J Phys Anthropol 80:267-94

Stoneking M (1998) Women on the move. Nat Genet 20:219-20

Stumpf MP, Goldstein DB (2001) Genealogical and evolutionary inference with the human Y chromosome. Science 291:1738-42

Sunderland E, Tills D, Bouloux C, Doyle J (1973) Genetic Studies in Ireland. In: Roberts DF, Sunderland E (eds) Genetic Variation in Ireland. Taylor and Francis, London, pp 141-160

Sykes B, Irven C (2000) Surnames and the Y chromosome. Am J Hum Genet 66:1417-9

Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2:930-42

Tagliabracci A, Turchi C, Buscemi L, Sassaroli C (2001) Polymorphism of the mitochondrial DNA control region in Italians. Int J Legal Med 114:224-8

Thomas MG, Bradman N, Flinn HM (1999) High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome. Hum Genet 105:577-81

Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB (1998) Origins of Old Testament priests. Nature 394:138-40

Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc Natl Acad Sci U S A 97:7360-5

Tills D (1977) Red cell and serum proteins and enzymes of the Irish. Ann Hum Biol 4:35-42

Tills D, Teesdale P, Mourant AE (1977) Blood groups of the Irish. Ann Hum Biol 4:23-4

Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus ML, Bonne-Tamir B, Scozzari R (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62:1137-52

Torroni 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

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835-50

Torroni A, Richards M, Macaulay V, Forster P, Villems R, Norby S, Savontaus ML, Huoponen K, Scozzari R, Bandelt HJ (2000) mtDNA haplogroups and frequency patterns in Europe. Am J Hum Genet 66:1173-7

Trask RL (1996) Historical Linguistics. Arnold, London

Tremblay M, Vezina H (2000) New estimates of intergenerational time intervals for the calculation of age and origins of mutations. Am J Hum Genet 66:651-8

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al. (2001) The sequence of the human genome. Science 291:1304-51

Wang J (2003) Maximum-likelihood estimation of admixture proportions from genetic data. Genetics 164:747-65

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

Wilder JA, 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 36:1122-1125

Wilson D (1999) Dark and Light: The story of the Guinness Family. Orion, London

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

Wooding S, Ostler C, Prasad BV, Watkins WS, Sung S, Bamshad M, Jorde LB (2004) Directional migration in the Hindu castes: inferences from mitochondrial, autosomal and Y-chromosomal data. Hum Genet 115:221-9

Woulfe P (1923) Sloinnte Gaedeal is Gall. Gill and Son, Dublin

Y Chromosome Consortium (2002) A nomencalture system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339-348

Young FW (1987) Multidimensional Scaling - history, theory, and applications. Lawrence Erlbaum Associates, Hilsdale, New Jersey

Zerjal T, Beckman L, Beckman G, Mikelsaar AV, Krumina A, Kucinskas V, Hurles ME, Tyler-Smith C (2001) Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. Mol Biol Evol 18:1077-87

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, Rosenberg NA, Feldman MW (2003) Features of evolution and expansion of modern humans, inferred from genome wide microsatellite markers. Am J Hum Genet 72:1171-86

Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L (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

Zvelebil M (2000) The Social Context of the Agricultural Transition in Europe. In: Renfrew C, Boyle K (eds) Archaeogenetics: DNA and the population prehistory of Europe. McDonald Institute for Archaeological Research, Cambridge, pp 57-79