AN EXPLORATION OF IRISH SURNAME HISTORY THROUGH PATRILINEAL GENETICS

by

Stephanie Kay Farmer

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Biological Sciences

Indianapolis, Indiana

December 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Susan Walsh, Chair

Department of Biology

Dr. Kathleen Marrs

Department of Biology

Dr. Lata Balakrishnan

Department of Biology

Approved by:

Dr. James Marrs

Head of the Graduate Program

To my parents, who always pushed me to be the best that I could be.

ACKNOWLEDGMENTS

I would not be where I am today without my amazing parents. To them I owe all my gratitude and all my love. They have sacrificed and provided for me so that I may have the chance at a better future and to pursue my dreams. They have given me more love and support than I thought possible and have truly been a solid foundation in my life.

I want to thank my family, friends and loved ones as well: Tiffany, Ailish, Taylor, Tommy and Jenny. Thank you for putting up with my complaints and always offering your unwavering support as I went through my graduate career. I would not be nearly as sane as I am now without the love you all offered.

If not for Dr. Walsh, I am not sure where I would even be in life right now. I would like to thank you for offering me a spot in your lab all those years ago when I was just an undergraduate student. Being in your lab has been an incredible learning experience, but not just from the knowledge you have passed down. I have learned a lot more about myself, so thank you for continuing to push me throughout the years.

My research would not be where it is without my lab mates, but, more importantly, my time in the Walsh lab would not have been nearly as enjoyable without them. To each and every one of you, thank you for your friendship. I could not have asked for a better team to work with. Thank you, Krystal Breslin and Charanya Muralidharan, for guiding me and shaping my experience in the lab. Thank you, Bailey Wills, Wesli Kay Stubbs and Mirna Ghemrawi for dealing with me and putting a smile on my face on even the hardest days. Thank you, Case Jacobus and Clare Hartman, for being wonderful and helping me with the mundane research, it truly helped tremendously. Thank you to Noah Herrick, Ryan Eller, Racquel Hopkins, Emma Fort, Sarah Alexander, Lydia Hawthorne, Paige Wallace and Morgan Howard for being incredible lab mates and friends.

Thank you to Dr. Lata Balakrishnan and Dr. Kathy Marrs for agreeing to be on my committee and offering your guidance and support.

Finally, thank you to our collaborators, Prof. William Smyth, Prof. Tommie McCarthy, Dr. Manfred Kayser and Mannis van Oven for all the hard work and dedication they have put into this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Farmer, Stephanie Kay. MS
Institution: Purdue University
Degree Received: December 2018
Title: An Exploration of Irish Surname History Through Patrilineal Genetics
Committee Chair: Susan Walsh

Due to Ireland's secluded geographical location, its genetic structure is a popular topic of study. The indigenous inhabitants of Ireland remained undisturbed for a long period time, allowing for a distinct genetic population to be created. This peace was disrupted by conflict with invading forces, such as the Nordic Vikings and Anglo-Norman forces. However, these historical events helped to shape both the culture of Ireland and the ancestry seen in the Irish population today. In Ireland, quite like many countries around the world, the male's surname is passed from father to son, just as the Y-chromosome. The relationship between Irish surnames and their corresponding Y-haplogroups was examined to determine if common and rare Irish surnames can be genetically linked to the historical invasions listed above. The surnames chosen for this study were selected based on their prevalence in Ireland, rare or common, and their proposed historical origin, Irish, Norse or British. To discover any possible patterns in surnames and Y-chromosomal DNA, Y-haplogroups were generated from the DNA of 630 Irish male subjects using an assay specifically developed for the region. The assay contains twenty single-nucleotide polymorphisms (SNPs) that were selected to further resolve the R1b-L21 Y-haplogroup for Irish ancestry, the most prevalent haplogroup in Western Europe, and Ireland in particular. Additional Y-STR data was also generated to examine recent surname history within the collected individuals. Each surname was examined to determine whether one haplogroup occurred more frequently and with this method, distinct patterns in Irish surnames and geographical locations were discovered. In addition to resolving Y-surname history patterns, it is also believed that this assay may be beneficial in determining if an unknown DNA sample is of Western European origin and even in some cases, if a more specific Irish origin can be predicted.

CHAPTER 1. INTRODUCTION

1.1 Purpose

Ireland, on a genetic level, is an interesting population to study due to its relatively isolated, islandic position and unique history, in particular, the male specific genetic makeup of Ireland. Patrilineal genetics have been proposed to tie in with high accord with its complicated history and society. Several previous studies in this area typically focus on two subjects: the history and genetics of Irish surnames and the influence invading forces had on the country's genetics.

The purpose of this study is to analyze the genetic makeup of Ireland using surnames and the Y-chromosome to better understand its history, by means of Y-haplogroup subclades and Y-STRs. Throughout the history of Ireland, conflict with multiple invading groups has shaped the culture and genetics of its people, including the Viking raids that plagued the country for over a century [1]. Though the Norse Vikings had such a large influence on many aspects in Ireland, including their culture, history and genetic profile, much of this impact has remained unstudied. Studies focusing on the impact the Vikings had on Ireland have begun to look at admixture analysis using the Y chromosome [2], though very few have used surnames to analyze and study the impact Vikings had on the Y chromosome specifically in the Irish populations. However, many studies have been done to analyze this impact on England [3, 4], a country that has also had a historical and genetic influence on Ireland through the Anglo-Norman invasions [5].

In both Ireland and Norway, like in many other patrilineal countries around the world, the male's surname is traditionally passed on from father to child, just as the Y-chromosome is passed on from father to son. Descendants of indigenous Irishmen should possess an Irish surname and an Irish specific Y-chromosome, while the Norse descendants should carry a Norse surname and a Y-chromosome distinctive from the Irish descendants, more similar to the modern Norse population. This pattern shall be examined within this collection of Irish individuals.

Additionally, this project aims to discover the geographical distribution of Ireland using the Y-chromosome. While most studies simply claim Ireland is broadly made up of the Y-haplogroup R1b [2], or group individuals based on their Y-STR profile [3, 6-8], individuals can be further classified into subclades. These subclades, that further resolve the Y-haplogroup, can then be linked to historic Irish families to better understand Ireland's past.

1.2 Historical Irish Families

Perhaps the most famous of the Irish dynasties, the Uí Néill kingdom claims to be descendants of the prehistoric Irish King of Tara, Niall Noígíallach, also known as Niall of the Nine Hostages. Niall Noígíallach, who ruled over the northern half of Ireland, is known as a 4th century warlord and regarded as semi-legendary. His descendants spread throughout and conquered Ireland, alternating the status of High King between family septs from the 7th to the 11th century [9].

The Eoganachta dynasty came into power in the 7th century but was founded in the 5th century by Conall Corc, descendant of the semi-legendary king Ailill Olum's first born son, Eogan [10, 11]. The Eoganachta reigned over Munster, though never produced a High King. The rule of the Eoganachta was mostly regarded as peaceful, achieved through wealth and diplomacy rather than war. However, in the late 10th century, they were eventually overthrown by a much smaller, Southern dynasty, the Dál Cais [12].

The Dalcassian dynasty are descendants from Cas, another semi-legendary king from the 4th century. While also residing in Munster, this clan's ancestry is mostly contained to Counties Clare, Tipperary and Limerick [13]. The most famous of the Dalcassians was perhaps Brian Borumha, who overthrew the Eoganachta, battled with the northern kingdoms (under the rule of Uí Néill's descendants) and became a High King of Ireland, the first to end the Uí Néill dynasty's rein [14, 15]. While it was theorized that the founders of the Eoganachta and Dalcassian dynasties were brothers, the current belief is now that the Uí Néill's briefly supported the Dalcassians in an attempt to weaken the Eoganachta [16]. Brian Borumha is also well known for the Battle of Clontarf, in which his armies fought the advancing Irish-Norse alliance forces to stop the increasing spread and reign of Norse Vikings. The Irish-Norse alliance was mainly comprised of the men of Leinster and the Viking Dublin "kingdom." Though Brian Borumha was victorious and helped contain the spread of Norse influence, he, his son and his grandson all perished in battle [17].

1.3 Historical Invasions

The Irish remained relatively undisturbed from outside forces until the first Viking Age in 795 AD, where Vikings began pillaging off the west coasts of Ireland [18]. These raids are characterized by quick "hits," where the Vikings would loot a village and then retreat to their ships. In 902 AD, the Vikings were pushed out of Dublin, where they continued to raid the coasts of Britain and other

areas along the Irish Sea [19]. The Vikings returned shortly after in 914 AD and created settlements, establishing some of the first modern ports cities in Ireland, including Dublin, Wexford, Waterford, Cork and Limerick [20]. It is believed that these towns were ruled by the Norse noblemen, warriors and merchants but were also inhabited by indigenous Irish [21].

The Viking Age ended in the 12th century during a power struggle between an ousted Irish king and those currently in power. The fallen king called upon the help of Norman armies, with the backing of King Henry II of England, to regain his throne. Successful, the fallen king then named one of the Normans an heir to his kingdom. As the Norman armies continued to expand their conquest, King Henry II, afraid of an impending rival, ordered his armies to invade Ireland to take control of the situation, beginning about the start of the British control over Ireland [22].

1.4 Irish Surname History

Ireland has the oldest recorded use of surnames, dating back to the late 9th century. These names were either profession based or patronymic in nature, meaning a surname was taken from a paternal relative's forename. Patronymic names contain either the prefix *Mac* or *O*. The prefix *Mac* stands for "son," meaning the surname was formed from the father's forename. The prefix *O* stands for "descendent of" and is used when the surname was formed from a grandfather or earlier paternal ancestor [23]. Hereditary surnames came into use in the 11th century [24], predominately by upper class families, to show wealth and status. It was not unheard of for some lower-class families to go without a surname until forced in the 16th to 17th century, during which Irish surnames were anglicized to match the Anglo-Norman fashion.

It should be noted that there are instances where an individual may change his name, or a surname may not match a believed ancestry. There are the obvious instances where this may occur, such as nonpaternity, adoption or matrilineal surname transmission. It was also a common practice for a Medieval leader to change their surname to their clan leader's name in an attempt to show allegiance [13]. As early Ireland put such a large emphasis on ancestry, genealogical records were used to validate power or property. These documents were often altered and forged, and as such not all records may truthfully indicated a names origin [9]. There are instances where children born from an Irish man and a Norse woman would typically bear the Norse name [25], which would disrupt a patrilineal naming line. Furthermore, most Irish surnames contain an Irish or Norse origin, but the use of hereditary surnames began after the Viking raids. While many influential families

of Viking descent may have followed early Irish surname practices, it is theorized that some individuals may have followed a trend of selecting a surname based on the Norse language.

1.5 The Y-Chromosome

Genetically, the Y-chromosome is the main determining factor in biological sex, causing a human fetus to develop as a male rather than a female. The Y-chromosome is divided into two sections: the pseudoautosomal region and the non-recombining region. The pseudoautosomal regions are located at the tips of the Y-chromosome and connect with the X-chromosome during miosis. Recombination between the X-and Y-chromosome occurs at these locations [26]. However, the portion of the Y-chromosome excluding the tips is often referred to as the non-recombining, or male specific, region. This area does not undergo recombination, as it has no homologous pair [27]. This leads to a very low mutation rate, meaning that an almost exact copy of a father's Y-chromosome will be directly passed on to his son and any mutation in the father's Y-chromosome will be inherited. Specific mutations, or markers, can then be traced back to recent or ancient ancestors and compared to different mutations other male lines may contain [27]. These mutations can be viewed in the form of Y-STRs (Y-chromosome short tandem repeats) or Y-SNPs (single nucleotide polymorphisms). Simply put, if families are patrilineal, then the Y-chromosome, and its mutations, should be consistent across surnames. As males only have one Y-chromosome, Y-chromosomal markers are mono-allelic. Therefore, Y-markers are considered haplotypes and not genotypes [28].

1.5.1 Y-Chromosome STRs

A Y-STR refers to a string of specific allelic short tandem repeats that differ in size. These sequences are typically 3-6 base pairs in length. The exact number of times a sequence is repeated is dependent on genetic inheritance [29]. A visualization of these repeats can be seen in Figure 1, showing a basic depiction of an STR. With autosomal STRs, there may be a differing number of repeats for each allele, as one allele is inhered from each parent. However, as the Y-chromosome is haploid, Y-markers are mono-allelic and there is typically only one repeated sequence per loci. Multiple loci with a differing number of repeats can be found all along the sequenced region of the Y-chromosome.

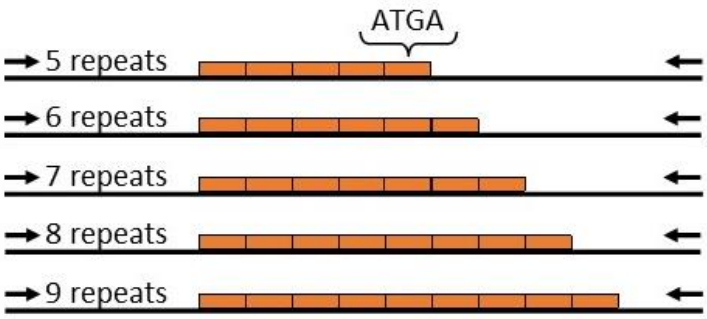


Figure 1: An example of an STR. The base pairs "ATGA" can be repeated a number of different times on one chromosome, and the amount of times it is repeated on each chromosome (one from each parent) becomes the STR genotype. [30] The Y-chromosome is mono-allelic, so only one set of repeats should be seen when viewing Y-STRs.

As a Y-STR profile is unique to a male line [27], it can be used as a reflective tool for identifying paternal relatives in a family. Y-STR profiles are used in a variety of ways in the forensic field: determining if a male was involved in a crime by identifying the presence of a Y chromosome in a DNA sample, finding a genetic paternal figure of a male suspect, separating a male DNA profile from a female DNA profile, separating multiple male suspects in a mixed sample and aiding investigators when searching for unknown, male perpetrators in instances where autosomal DNA profiles are unhelpful [31, 32].

The main application of Y-STRs for this study revolves around familial matching. Y-STRs are predominantly used for paternity testing, demonstrating their ability to accurately depict if two men are father and son, and, to a less accurate degree, if two men are brothers, cousins, grandfather and grandson, etc. [33]. This predication is possible, again, due to the non-recombining portion of the Y-chromosome. As the father passes on his Y-chromosome to his son, there is only one set of repeats able to be inherited. When comparing multiple locations for Y-STRs, a man will share an identical or very similar YSTR profile [34]. However, this makes many Y-STR kits impractical at separating males from the same paternal lineage, as their profiles are likely to read the exact same. Newer Y-STR kits combat this issue with the inclusion of rapidly mutating (RM) markers. RM Y-STR markers have mutation rates above 1×10^{-2} per generation, while most Y-STRs used in forensics have a mutation rate of 1×10^{-3} per generation or lower [35]. A study comparing the Yfiler (Applied Biosystems) Y-STR kit (which contains seventeen non-RM Y-STRs) to thirteen discovered RM Y-STRs found that the RM Y-STRs were able to increase the ability to differentiate between closely and distantly related men, with a 4.4-fold increase of average male relative differentiation [36].

The Promega PowerPlex® Y23 kit is one of many Y-STR kits available for purchase. This Y-STR kit uses seventeen Y-STR loci commonly available in many kits (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and Y-GATA-H4) with six additional, highly discriminating Y-STR loci (DYS481, DYS533, DYS549, DYS570, DYS576, and DYS643) for a total of twenty-three Y-STR loci able for analysis for improved distinguishing ability, illustrated in Figure 2. Of the six additional Y-STR, both DYS570 and DYS576 are RM Y-STRs. The addition of the highly discriminatory and RM Y-STRs was the reason for selecting this kit over others, as they can increase the ability to better differentiate between male relatives. This ability is needed so that direct male relatives are not included during statistical analysis, as this may interfere with results by skewing allele frequency.

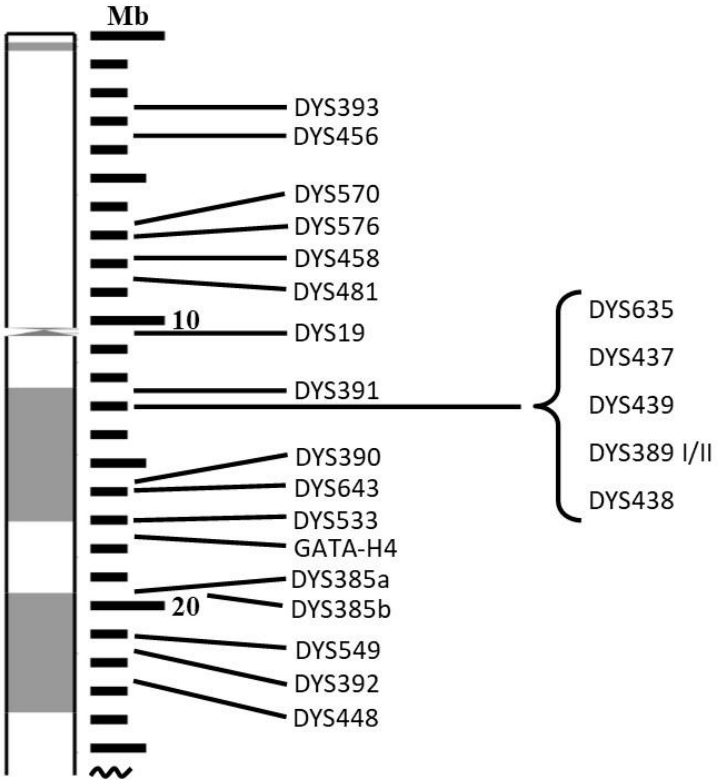


Figure 2: The locations of the 23 loci able to be sequenced in the Promega PowerPlex® Y23 Y-STR kit on the Y-chromosome [37].

1.5.2 Y-Chromosome SNPs

A SNP refers to a single nucleotide located on a strand of DNA that can be variable among a population. This is the most common form of genetic variance, as one change in a base pair can

lead to many of the phenotypic differences seen among the human population, such as pigmentation or disease [38].

The Y-chromosome also contains genetic polymorphisms which are associated with a type nonvisible genotypic difference, called a Y-SNP [39]. Y-SNPs have been shown to be useful in predicting the geographic origin of a male's paternal ancestor [40].While Y-STRs can be used to detect paternal ancestry, this estimation is limited to current or more recent geological origins. Y-SNPs are more applicable as genealogy markers as they have a lower rate of mutation than Y-STRs [41]. While Y-STR's typically have a mutation rate between 1×10^{-2} to 1×10^{-3} [35], Y-SNPs have a much slower rate of mutation of 2×10^{-8} base pairs per generation [42]. SNPs, therefore, are a better marker for analyzing more ancient DNA and family lines, which are represented by different Y-Haplogroups.

1.6 Haplogroups

Haplogroups are determined by a specific set of mutations inherited from a single parent and are defined by SNPs [43]. Y-haplogroup A is believed to be the very first Y-haplogroup and is most prominent in Southern Africa, suggesting that this is where it originated. The migration patterns of Y-haplogroups can then be viewed by frequency in populations and geographical area. The changes in Y-haplogroup frequencies around the world reflect the migration patterns suggested in the "Out of Africa" theory [44]. These two theories support each other, suggesting that as humans traveled out of Africa, the Y-chromosome slowly mutated over time, developing new haplogroups as populations began to diverge [45]. Similar to a phylogenetic tree mapping the evolution of a species as it slowly diverges from a recent common ancestor, the Y-chromosome phylogenetic tree maps the changes in mutations as major Y-haplogroups diverge into new haplogroups, and those major haplogroups branching off into many subclades [7]. The implications of this discovery not only allowed the Y-chromosome to be used for anthropological purposes, but also for more in depth ancestral and genetic studies as well.

1.6.1 Haplogroup R1b

The Y-haplogroups most common in contemporary Europe are E, G, I, J, N and R, with more than 50% of the European population belonging to Y-haplogroup R [46]. Y-haplogroup R can be further divided into subclades R1 and R2, with most Europeans belonging to R1. R1 can then be further

divided into R1a, which is more prevalent in Eastern Europe, and R1b, which is more prevalent in Western Europe. This divergence is believed to have occurred about 25,000 years ago [47]. With R1b being the most frequent haplogroup found in Ireland and the surrounding area [48], this subclade was chosen for its specificity to the Irish population.

1.6.2 Subclades and Further Resolution of R1b

Twenty SNPs linked to subclades within the Y-haplogroup R1b were chosen for this project based on their location in the R1b phylogenetic tree. Most males in Ireland belong to the subclade L21 [49] and so this assay was designed to further resolve the classification of Irish males. Subclades located higher in the tree, indicating more ancient subclades, act as hubs for subclades father down the tree, meaning these subclades represent individuals with a more ancient common ancestor to the Irish population who migrated elsewhere. This includes L11, P312, Z195, L21 and DF63. The chosen, higher resolved subclade, i.e. the mutations that are relatively more recent and specific to current families, help increase the detail and accuracy of ancestral predictions. These subclades include FGC11134, DF21, DF5, Z3017, Z2534, DF49, M222, S588, DF41, L513, L193, Z251, L1335, S764 and Z255. The phylogenetic tree specific to only these subclades can be seen in Appendix A.

Many of these highly resolved subclades are proposed to be linked to countries based on unpublished work in the genealogical community. Z255, Z2534, FGC11134, M222 and S588 are theorized to be specific to Ireland. S764 and L193 are theorized to be specific to Scotland. DF41, DF21, Z3017 and DF5 are theorized to be mainly found in Ireland and Britain. L513 is theorized to be mainly found in Ireland and Scotland. These theorized are based on the collected findings in public databases such as FamilyTreeDNA.

1.7 Subclades, Families and Locations

Influential and ancient Irish families and geographical locations in Ireland have previously been linked to specific R1b subclades [9, 13]. These families can be grouped into three categories: Irish Type I, Irish Type II and Irish Type III. Irish Type I represents North West Ireland and is associated with the *Uí Néill* kingdom, representing descendants from fifth-century king, Niall of the Nine Hostages. Irish Type I is believed to be linked to the subclade M222 [9, 50]. Irish Type II represents Southern Ireland and is theorized to be linked to the Eoganachta kingdom from Munster. Due to

the unpublished, but largely supported, work done by Dr. Ken Nordtvedt, the Irish Type II group has been linked to the subclade CTS4466 (CTS4466 is not included in the R1b-L11 Multiplex assay, but CTS4466 is a newer mutation that directly branches off FGC11134 and will be represented as such). Irish Type III represents South West Ireland and is associated with the Gaelic Dalcassian families, who represent descendant families of Cas, another ancient Irish king [13]. Irish Type III has been linked to the subclade L226 [50] (L226 is not included in the R1b-L11 Multiplex assay, but L226 is a newer mutation that directly branches off Z2534 and will be represented as such). These Irish Types were originally discovered using Y-STR profiles of males with surnames believed to be descendants of the historical kings mentioned above. Later, these Y-STR profiles were able to be linked with a Y-SNP. Y-SNPs are more useful, as the Y-STR profiles need to include possible 1-step mutations. The Y-STRs will only continue to mutate as time continues, diverging from the proposed Y-STR profile type. The Y-SNPs will mutate at a much slower rate and, if a new mutation does occur, the ancestral subclade belonging to the Irish Type can still be easily determined.

Complications with assessing surnames using the Y-chromosome include mutations on the Y-chromosome, multiple founders attributing to a single surname, nonpaternity, adoption, matrilineal surname transmissions and genetic drift [6]. Even still, a general trend of a common subclade should be observed within surnames.

1.8 What is a Viking?

Viking is a term typically used for raiders hailing from Scandinavia, which is present day Norway, Denmark and Sweden. They began their voyages at the end of the 8th century and frequented areas in Britain and Ireland, particularly along the coasts of the Irish Sea. These raids started as quick plunders where Vikings would gather riches to return home. Beginning in the 9th century, they began to conquer kingdoms and create settlements [51]. Many countries were affected by the Vikings, but studies mainly focus on this effect in the British and Irish populations [2, 3].

The most common Y-haplogroups in Norway are I1, R1a, R1b and N, respectively. Y-haplogroup I1 is believed to have originated in Denmark, while Y-haplogroup N is common among the Finnish [52]. The modern-day presence of R1b in Norway is perhaps explained by the shared history between Ireland, England and the Vikings. Individuals of Norse descent are most likely to belong

to either the I1 or R1a haplogroup and studies use this belief when examining the presence of Vikings in their population [2, 3].

One such study [3] examined the genetic signature of Vikings in areas of northwest England, an area where Viking activity has been historically documented. Two distinct groups of unrelated males were collected: modern (males with two generations of residency in the area) and medieval (males with known ancestry in the region and a surname historically known to be present in the area during medieval times). The study then compared the Y-haplogroup frequencies of the two groups with Norway. It was determined that based upon the higher frequency of both Y-haplogroups I and R1a, the medieval group is more genetically similar to the genetic structure of Norway. The study also used Y-STR allele frequencies to compare the two groups to Norway and found that the Y-STR allele frequencies of the medieval group were more genetically similar to that of the current Norse population. This led to the conclusion that there was Viking presence in northwest England as historically suggested and that sampling historical surnames provides a more accurate depiction of past genetic events.

A similar study was performed on the Irish population [2]. One group was comprised of males with believed Irish surnames and the other of males with surnames believed to have Norse roots. Y-STRs were used to place samples into Y-haplogroups and the haplogroup frequencies of males were used to determine the amount of genetic influence Vikings had on Ireland. It was determined there was not a significant difference between the Irish group and the Norse surname group when comparing them to the Y-haplogroup frequencies from the Norway data. The findings reported that, when comparing the Y-STR alleles, the Norse surname group was not more genetically similar to Norway and was more similar to the Ireland group. This led to the conclusion that while there may have been a small presence of Vikings in Ireland, the presence may not be as large as historically indicated. However, this study contains a relatively small sample size, using only 47 men from 26 different surnames. This is not an accurate depiction of the Irish population, which may skew the results. Also, the study used Y-STRs to compare the different populations as opposed to Y-SNPs, which are more indicative of ancestral lineages.

Both studies show a common practice for identifying Viking presence in non-Norse populations. The presence of Y-haplogroups I and R1a heavily imply the presence of Viking DNA [2, 3]. A comparison of the Y-STR profiles of males from the population being studied to the Y-STR profile of modern males from Norway can also show how closely related the two groups may be.

1.9 Norse Admixture in Ireland

A large scale study was recently performed on Ireland using autosomal DNA to examine the country's genetic fine scale structure [53]. The study was able to find seven clusters of predominantly Irish membership, each distinct from one another, and three clusters of shared Irish-British ancestry. The discovered distribution of Irish clusters follows the geographic borders of Ireland's provinces, showing that individuals in Ireland do not typically move from the area they were born. Another focus of this study was on the present admixture within Ireland. An upper estimate of 20% Norwegian ancestry was found within the Irish population. This ancestry mainly comes from individuals originating from counties on the northern or western coasts of Norway where Norse Viking activity is believed to have originated from. Further analysis of these results suggests this admixture event can be dated around the time of historically documented Viking activity in Ireland. The conclusion of this study was heavily in support of the significant presence of Viking in Ireland, agreeing with historical documentation.

1.10 Unknown Person Identification Case

R1b-L21 is the most common subclade in Ireland [49] and the highly resolved subclades (FGC11134, DF21, DF5, Z3017, Z2534, DF49, M222, S588, DF41, L513, L193, Z251, L1335, S764 and Z255) chosen for the R1b-L11 Multiplex Assay are believed to be specific to certain countries (Ireland, Britain or Scotland) by the genealogical community. This assay therefore has the potential to propose an Irish ancestry for a male sample. Depending on the subclade the sample is found to be, an ancestral location within Ireland, such as the province, or potential surname may be proposed as well.

A male body was found in a grave with a headstone only marked as JB-55 in Griswold, Connecticut, with JB assuming to be the initials of the male and 55 his age. The male is believed to have died between 1810-1830. The examining lab theorized this male to contain Irish ancestry. To test this theory, two extracts of DNA from the body (JB-55-1 and JB-55-2) and two reagent blanks (JB-55-RB1 and JB-55-RB2) were sent to be tested with the R1b-L11 Multiplex assay to determine the likelihood of Irish ancestry. A potential surname was also hoped to be discovered. The generated Y-STR profile was given as well to help with analysis. This analysis will be discussed further in the Results and Discussion section.

CHAPTER 2. MATERIAL AND METHODS

2.1 Sample Collection and Preparation

IRB approval was received for this study. A total of 630 DNA samples from male volunteers making up 65 different surnames were collected. The surnames of interest, included in Appendix B, were chosen for their proposed ancestry (Irish, Norse or English/Scottish) and significant history, based upon MacLysaght's book, *The Surnames of Ireland* [54]. Common names with confirmed origin history as either Irish (McCarthy and O'Sullivan) or British (Johnston and Taylor) were chosen as controls for this study. Related surnames with multiple spellings were grouped together for analysis. Individuals containing surnames of interest were selected from a telephone book and contacted through postal letters detailing the study. Those choosing to participate mailed back a questionnaire detailing paternal history locations, consent forms and two cytology brushes for testing. Two separate cohorts were collected. Cohort 1 included surnames considered to be less frequent, or rare, in Ireland, representing surnames with a proposed Norse surname. Cohort 2 included surnames found more commonly in modern day Ireland, representing surnames with a proposed Irish surname. The number of families contributing to the samples was determined by comparing Y-STR profiles of individuals with the same surname. Surnames were ranked by frequency (65 being the least frequent, 1 being the most frequent) using information from The National Archives of Ireland Census of 1911, the most current full censuses open to the public.

2.2 Extraction and Quantification

For Cohort 1, DNA was extracted by a collaborating lab via a standard phenol/chloroform protocol with proteinase-K digestion.

DNA from the swabs that were not extracted from the collaborating lab (Cohort 2) were extracted via an inorganic salting out method, as followed: the swab was injected into a 2ml Eppendorf tube containing 1300µl lysis buffer (DNA Free), 15µl Proteinase K (New England Biolabs, Ipswich, MA), and 130µl of 10%SDS (Dot Scientific Inc., Burton,MI). These tubes are then incubated at 36ºC overnight. Following this, the swab was removed and 200µl of 0.5M NaCl (Dot Scientific Inc.) was added and the tube was immediately inverted several times and incubated at room temperature for 10 minutes. The samples were then centrifuged for 10 minutes at 10ºC at 16200rcf.

1ml of supernatant was then transferred to a new tube with 1ml of 100% Isopropanol (Fisher Scientific International Inc. Hampton, NH) being added, slowly inverting several times followed by another 10 minute incubation period at room temperature. The tubes were then spun in the centrifuge for 15 minutes at 10ºC at 16200rcf. The supernatant was then removed from the tubes without disturbing the DNA pellet and discarded. 500ml of 70% Ethanol (Decon Laboratories Inc., King of Prussia, PA) was added to the pellet to wash any remaining contamination and the tube was spun down for another 5 minutes 10ºC at 16200rcf. The ethanol was discarded, and the pellets were then washed a second time following the same specifications. Following the second ethanol wash, after the ethanol was discarded, the pellets were left to dry in the hood. After any remaining ethanol has dried off, 30µl of Molecular Grade Water (Dot Scientific Inc.) was added to resuspend the samples. Samples were stored at -20 ºC for further use.

The DNA of the extracted samples was quantified using the Invitrogen™ Qubit™ Fluorometer by the manufacturer's guidelines for the DNA High Sensitivity Assay kit (Fisher Scientific International Inc.).

2.3 Genotyping

2.3.1 Collaborating Lab

Samples from Cohort 1 (*n*=275) were genotyped by a collaborating lab, with a focus on investigating Y-haplogroup R1b. While some subclades underneath L21 were included in the preliminary analysis, M222 (*n*=64), L513 (*n*=513), L144 (*n*=1), L226 (*n*=1) and P314 (*n*=1), 135 samples were genotyped as L21. It was decided that a new assay needed to be developed to delve further into the subclade L21 to achieve a higher resolution and better investigate the genetic population structure of Ireland. Samples that fell outside of haplogroup R1b were analyzed with an existing global Y-SNP multiplex tool developed previously by collaborators, van Oven et al. [55].

2.3.2 R1b-L11 Multiplex Assay

For the successful genotyping of these DNA samples, a PCR SNaPShot Multiplex assay was designed. A set of 20 different Y-SNPs that are believed to correspond with certain R1b subclades were designed for a custom genealogical assay: L11 rs9786076, P312 rs34276300, Z195 rs568477247, L21 rs11799226, DF63 rs758915822, Z255 rs770144323, Z2534 rs771470152,

FGC11134 rs762444220, Z251 rs746402982, DF49 rs769171919, M222 rs20321, S588 rs953965077, L1335 rs778756430, S764 rs960989927, DF41 rs372620228, L513 rs577488529, L193 rs777935843, DF21 rs138322855, Z3017 rs148603310, DF5 rs865977022. Primer3Plus, a web-based designed tool, was used to design both the flanking and single base extension (SBE) primers for the 20 Y-SNPs. The information on primers used in the PCR1 reaction is located in Table 1 and the information on primers used in the SBE is located in Table 2. The optimal direction for the SBE primer was then selected using information from NCBI, as only one direction is needed for extension. All SBE primers were able to be design using the forward direction. The default parameters of the program were used for this design [56]. Fragment sizes were limited to 150bp or less to cater to degraded samples. In order to appropriately separate the SBE products, poly-T tails were added to the 5' end of the SBE primer. After successful primer design, the Autodimer software program was utilized to eensure that no primer-primer interactions took place [57]. All primer sequences and concentrations are provided in Appendix C.

The Y-SNP DF13 rs373989227 was originally included in the above multiplex assay. However, due to its location on the Y-chromosome, the Y-SNP could not amplify in a sufficient manner to be consistently viewed on the CE. As most major subclade clusters that fall beneath DF13 are included in the R1b-L11 Multiplex Assay, the assay was not significantly impacted by removing DF13 from the selected Y-SNPs.

Table 1: Primer information for the PCR1 reaction in the R1b-L11 Multiplex Assay

Subclade	PCR Primers	Bases	Sequence	Conc.	Product Size
L21	rs11799226_F	21	CCCTCCTCAGCAACAGTAAAA	0.5 µm	104bp
	rs11799226_R	20	GGAAGCATTCAGAAGCAGGT	0.5 µm	
M222	rs20321_F	21	CATGATCAAGCAGTGGTGCTA	0.5 µm	57bp
	rs20321_R	22	TTTCCTGAGCAAGAAGTATGGA	0.5 µm	
DF49	rs769171919_F	20	ACAGCATGGGGGTAACTGAC	1.0 µm	122bp
	rs769171919_R	25	TTCAAAGAGATAGTTCCTAGGTCCA	1.0 µm	
Z2534	rs771470152_F	20	CATTGCTTCATGGACCCTTC	0.5 µm	96bp
	rs771470152_R	20	ACATGTACCCCGGAACTTCA	0.5 µm	
L513	rs577488529_F	22	ACTGATTCCACATGATGGTGAG	0.5 µm	168bp
	rs577488529_R	22	AAAGCAAACTGTGTTCAAATGC	0.5 µm	
L193	rs777935843_F	22	AGAATTTAAAACAGGCCAGATG	1.0 µm	218bp
	rs777935843_R	20	GGAGTGCAGTGGCACAATTT	1.0 µm	
Z251	rs746402982_F	20	GTTCCAGAAAGGTGGAGCAG	0.5 µm	108bp
	rs746402982_R	22	AAAAGAACAAGATGTTCCCCTT	0.5 µm	
DF41	rs372620228_F	20	AACCCATTGGAGAATGGAAA	1.0 µm	154bp
	rs372620228_R	20	CTCAAACGCAAAGCAATGAA	1.0 µm	
P312	rs34276300_F	21	CTTTCTCAACCCACTGTCTGC	0.5 µm	153bp
	rs34276300_R	20	GGTGGAGTTGGGGCTAAAGT	0.5 µm	
L1335	rs778756430_F	20	CCCCTCCTCTAGGAGACCAA	0.5 µm	93bp
	rs778756430_R	20	TATGGCCTCCCTTTCACTTG	0.5 µm	
Z255	rs770144323_F	22	AAGCAAAACAAGACAGACCAGA	0.5 µm	119bp
	rs770144323_R	21	TCTGGTTCTCCAGCTTCCAGA	0.5 µm	
L11	rs9786076_F	22	TTACCTGTGGGCATTTGTAAGA	0.5 µm	136bp
	rs9786076_R	22	AGCACCCACATCCTGTTTTAGT	0.5 µm	
DF63	rs758915822_F	21	TTGCCAAGATTGTTCTGTTCC	0.5 µm	109bp
	rs758915822_R	20	GGTTAGGATTGTCTTGGCCA	0.5 µm	
DF21	rs138322855_F	24	CATTAACAAATTCCCAATTTCAGG	0.5 µm	123bp
	rs138322855_R	24	GTATTTGGTGCTGTGAATTTCTGA	0.5 µm	
FGC11134	rs762444220_F	22	GCTCAAGTCCATCATCACAGAG	0.5 µm	106bp
	rs762444220_R	22	TTGCAAACTGTCTGAGAAATCC	0.5 µm	
Z195	rs568477247_F	20	AGGCTCCAACCACCAAAAAT	0.5 µm	89bp
	rs568477247_R	20	CCTGATTTCAACCCAGCCTA	0.5 µm	
S764	rs960989927_F	20	TGGGGAAATAATTGGCTGTG	0.5 µm	93bp
	rs960989927_R	20	GCAGGGCTCTAGCCATACAG	0.5 µm	
DF5	rs865977022_F	20	CTGATCAAAATGCCTGCAAA	0.5 µm	101bp
	rs865977022_R	27	CTGACTACAAAATAGATATCCTTCACA	0.5 µm	
S588	rs953965077_F	23	TGGCCAATTCTTAATCTTCAGTT	0.5 µm	99bp
	rs953965077_R	23	TGACCACAAAAGTAGGAAGAAAA	0.5 µm	
Z3017	rs148603310_F	21	GATCCCATTTGTCAATTCTGG	0.5 µm	152bp
	rs148603310_R	20	ATACTGGACCGCTTCCTTCC	0.5 µm	

Table 2: Primer information for the SBE reaction in the R1b-L11 Multiplex Assay

SBE Primers	Model Input	Bases	Sequence	Conc.
rs11799226_F	G/C	23	TTTTTCCCAATTTTATTGCGCTG	1.30 µm
rs20321_F	G/A	29	TTTTTTTTCATTCCTAATCCCTCATCCGA	0.04 µm
rs769171919_F	C/T	32	TTTTTTTTTTTTTGTCTCCATCCAAATCTCAT	2.30 µm
rs771470152_F	A/G	30	TTTTTTTTTGCTTAGAGTCAAAGGGATGGG	0.30 µm
rs577488529_F	G/A	37	TTTTTTTTTTTTTTTCACATTGCATTTATTGTGCACC	0.50 µm
rs777935843_F	C/T	40	TTTTTTTTTTCCAGCTAATTTTTTGTATTTTTAGTAGAGA	2.00 µm
rs746402982_R	C/T	43	TTTTTTTTTTTTTTTTTTTTTCCCTAAACTATGTTTGATACTC	0.50 µm
rs372620228_F	A/G	45	TTTTTTTTTTTTTTTTTTTTTTTAACTGGTTTAATATCATGCACAA	2.50 µm
rs34276300_F	C/A	48	TTTTTTTTTTTTTTTTTTTTTTTTTCTGCTAATGTATCTGCTGCACTG	1.00 µm
rs778756430_F	T/C	51	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGTCAAGGCTTTTTTGATTTC	0.60 µm
rs770144323_F	A/G	54	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTGGGGAACCCTAATATAGGTATGT	0.20 µm
rs9786076_F	A/G	60	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAACAGACAGAACCAAAAGTTCTTC	0.80 µm
rs758915822_F	T/C	60	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCAAGATTGTTCTGTTCCAT	0.30 µm
rs138322855_F	G/A	63	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTGAAATTGTGGAATACTGTAGGAG	0.70 µm
rs762444220_F	A/T	68	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGATAAACCTGCTATTCAGTTTCCTGT	0.04 µm
rs568477247_F	C/T	74	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAGTAGCTTCAAAATTAAATAAATCCATGT	0.30 µm
rs960989927_F	G/A	78	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCAATTTCCTTATATTGACTCAC	0.80 µm
rs865977022_F	A/G	78	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGTGAAAATTGAGAGATAAATATAAAGC	0.80 µm
rs953965077_R	C/T	79	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATGGAAAAGAACTGTATTACGTAAGTT	2.00 µm
rs148603310_F	C/T	82	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGTATTTTAGTCGTGAACTCTTTGC	0.80 µm

PCR amplification was performed for all 20 Y-SNPs containing 1µl genomic DNA extract (at 1ng), 1X PCR buffer (Applied Biosystems), 2.5 mM MgCl2 (Applied Biosystems), 220 µM of each dNTP (Roche, Mannheim, Germany) 1.75 U AmpliTaq Gold DNA polymerase (Applied Biosystems). A Mastercycler Nexus SX1 thermocycler (Eppendorf) using the following parameters was used for amplification: 95° C for 10 minutes, 33 cycles of 95°C for 30 seconds and 62°C for 30 seconds, 62°C for 5 minutes. The amplified products from the first PCR were then purified with ExoProStar-S (GE Healthcare Europe GmbH, Eindhoven, The Netherlands) and incubated at 37° C for 45 minutes and 80° C for 15 minutes. The SBE reaction utilizes 2µl of previously purified PCR product to simultaneously genotype all 20 Y-SNPs of interest in the presence of 1µl of SNaPshot Ready Reaction Mix (Applied Biosystems). This reaction took place at the following parameters: 96° C for 2minutes; 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds, 60° C for 30 seconds. 1µl Shrimp Alkaline Phosphatase (SAP) (USB Corporation) was used to purify the products of the SBE reaction. Incubation was done at 37°C for 45 minutes and

inactivation was done at 80°C for 15 minutes. The purified products from the SBE reaction were then analyzed on the ABI 3500 Genetic Analyzer (Applied Biosystems) with POP-7 polymer on a 36 cm capillary array under an injection voltage of 2.5kV for 10 seconds and with a running time of 500 seconds at 60° C. Gene Mapper v3.7 software program (Applied Biosystems) was used for the analysis of the samples.

2.3.3 Promega PowerPlex® Y23 YSTR

All samples were genotyped for 23 Y-chromosomal STR loci (DYS576, DYS389I/II, DYS448, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438 (penta), DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS385a/b, DYS456 and GATA-H4) using the Promega PowerPlex® Y23 kit according to the manufacturer's instructions. The product was run on the ABI 3500 Genetic Analyzer (Applied Biosystems) with POP-7 polymer and visualized with Gene Mapper v3.7 software program (Applied Biosystems) for the analysis of the samples.

2.4 Genetic Distance using Y-STR population data

Y-haplogroup data was taken from the 1000 Genomes Project database [58] to compare the collected Irish Y-haplogroup and subclade data with populations throughout the world.

To compare the Y-STR profiles of the collected Irish samples to other populations, samples were placed into three groups based on suspected surname origin: Irish Suspected Surnames (ISS), Irish Norse Suspected Surnames (INSS) and Irish English Suspected Surnames (IESS). These groups represent Irish individuals whose ancestors are believed to be indigenous to Ireland, Irish individuals whose ancestors are believed to be Norwegian and Irish individuals whose ancestors are believed to be English or Scottish, respectively. With surnames separated based upon proposed surname origin, population similarities can be visualized to see which population each group is genetically closest too based upon available Y-STR data from other studies. Y-STR profiles and frequency data from the United Kingdom and Ireland were taken from Aliferi *et. al* [59], while global Y-STR data was taken from Purps *et. al* [60]. The populations taken from the global dataset were: Finland, Sweden, Denmark, Germany, the Netherlands and Wales. These populations, including the United Kingdom and Ireland, were all selected for their proximity to each other. As there is no current study published containing an extensive dataset of Norse individuals with a

comparable number of Y-STR loci, data was taken from the Norway DNA Project found at FamilyTreeDNA [61]. By comparing these datasets with the ISS, INSS and IESS samples, it is possible to observe if the ISS samples are more genetically similar to modern day individuals from Ireland, if the INSS samples are more genetically similar to modern day individuals from Norway and if the IESS samples are more genetically similar to modern day individuals from the United Kingdom. The relationship of the ISS, INSS and IESS groups were compared to the aforementioned population groups using genetic distance, or R_{ST}. An R_{ST} value is an analog for a fixation index, or an F_{ST} value. F_{ST} measures the genetic distance between subpopulations using allele frequencies. R_{ST} is a form of F_{ST} that is typically preferred in ancestry studies, as it more suited for markers with a stepwise mutation model and is not affected by within-population variation [62]. Statistical analysis on all populations was done using the Arlequin package [63]. This assessment was visualized using a Multidimensional Scaling (MDS) graph. The MDS was plotted using the Excel add-on XLSTAT [64].

2.5 Additional Analyses

Allele frequency data from all collected samples was analyzed using the online tool STRAF [65]. STRAF was also used to create a Principal Component Analysis (PCA) plot from the Y-STR data of the collected Irish population based upon Y-haplogroups and subclades.

NEVGEN is one of many online tools available to provide haplogroup predictions based on a Y-STR profile. A fitness score is used to estimate the accuracy of the most probable Y-haplogroup [66]. Samples from Cohort 2 that were not able to be classified by the R1b-L11 Multiplex Assay were placed into a Y-haplogroup based upon this method.

CHAPTER 3. RESULTS AND DISCUSSION

3.1 Y-Haplogroup R1b Subclade Genotyping

The majority of Irish samples (88.3%) fall into the Y-haplogroup R1b, with different percentages occurring when separating the samples into categories based on their proposed surname origin. In surnames with a proposed Irish surname, 91.5% of samples fall into R1b. Of the proposed Norse origin surnames, 86.6% of samples belong to R1b and 77.3% of samples from the proposed English surnames belong R1b. Irish individuals with a proposed English surname are less likely to belong to R1b than either of those with the proposed Irish or Norse origin surnames. However, the proposed English surname samples belonging to R1b typically fall into subclades higher in the phylogenetic tree (L11, P312, Z195 and L21) or subclades that are theorized to be British or Scottish in origin by the genealogical community (L513, DF5, and DF41)

The subclade results from this study agree with previous reports of the subclade R1b-M222 being the most common subclade in Ireland [49, 50], with a large proportion of samples (19.0%) falling into this group. In the surnames with a proposed Irish surname, 25.3% of samples are resolved to M222. Of the proposed Norse origin surnames, 14.6% of samples belong to R1b-M222 and 4.6% of samples from the proposed English surnames.

Subclade R1b-S588 is a further subclade of M222, meaning it is a relatively newer mutation within the M222 subclade and is therefore highly specific to the Irish population. However, only 1.4% of those with a proposed Irish surname belong to S588 while 10.1% of those with a proposed Norse surname belong to S588 (with no samples in the proposed English surname group belonging to S588). This could be indicative of an Irish specific clan taking on a surname with Norse root. This theory will be discussed further in another section.

All subclades included in the R1b-L11 Multiplex Assay were found in the Irish population with the exclusion of DF63 and L193. Proposed theories in the genealogical community suggest that L193 is specific to Scotland, which may be the reason it is not seen in the Irish population. DF63 is theorized to be found in Western Europe, but this subclade is either very rare or non-existent in the Irish population. Subclades DF5, DF41, Z251, L1335 and S764 are all found at very low frequencies in the collected Irish data.

Other Y-haplogroups found in this study were I, E, J, G and T. As mentioned, Y-haplogroup I is frequently found in Norway and other Scandinavian countries [52]. Y-haplogroup E is most commonly found in the northern coast of Africa in countries such as Algeria and Egypt., but is also frequently seen all throughout Europe [67]. Y-haplogroup J is considered a Middle Eastern haplogroup. Its presence in Europe is described by the westward migration of people from this area [68]. Y-haplogroup G is similar, as it is most frequent in the Middle East but also Caucasus[69]. Again, Y-haplogroup T is frequent in the Middle East and has migrated westward into Europe [70]. Y-haplogroups J, G and T are believed to be linked with the spread of agriculture into Europe.

A breakdown of subclades by population for the samples that could be successfully genotyped is shown in Table 3. Individuals in Cohort 2 who could not be placed into an R1b subclade with the R1b-L11 Multiplex assay were categorized into a most probable Y-haplogroup based on their Y-STR profile using NEVGEN [66]. R1b represents individuals who were grouped into R1b based on their Y-STR profile but could not be typed using the R1b-L11 Multiplex Assay, meaning this subclade falls above R1b-L11, which is a more ancient subclade believed to have originated in the Bronze Age (4.8-5.5 kya) [71]. This is more common in males with a proposed English origin surname.

Table 3: Distribution of Y-haplogroups and subclades of the Irish population based on proposed surname origin

Irish Haplogroups			Norse Haplogroups			English Haplogroups			Total Haplogroups		
SNP	#	%	SNP	#	%	SNP	#	%	SNP	#	%
L11	12	4.10	L11	18	6.72	L11	4	9.09	L11	34	5.62
P312	12	4.10	P312	10	3.73	P312	7	15.91	P312	29	4.79
Z195	1	0.34	Z195	8	2.99	Z195	1	2.27	Z195	10	1.65
L21	22	7.51	L21	41	15.30	L21	6	13.64	L21	69	11.40
DF63	0	0.00	DF63	0	0.00	DF63	0	0.00	DF63	0	0.00
FGC11134	24	8.19	FGC11134	30	11.19	FGC11134	0	0.00	FGC11134	54	8.93
DF21	20	6.83	DF21	13	4.85	DF21	0	0.00	DF21	33	5.45
DF5	1	0.34	DF5	3	1.12	DF5	1	2.27	DF5	5	0.83
Z3017	21	7.17	Z3017	2	0.75	Z3017	0	0.00	Z3017	23	3.80
Z2534	23	7.85	Z2534	4	1.49	Z2534	0	0.00	Z2534	27	4.46
DF49	19	6.48	DF49	0	0.00	DF49	1	2.27	DF49	20	3.31
M222	74	25.26	M222	39	14.55	M222	2	4.55	M222	115	19.01
S588	4	1.37	S588	27	10.07	S588	0	0.00	S588	31	5.12
DF41	1	0.34	DF41	2	0.75	DF41	3	6.82	DF41	6	0.99
L513	19	6.48	L513	13	4.85	L513	6	13.64	L513	38	6.28
L193	0	0.00	L193	0	0.00	L193	0	0.00	L193	0	0.00
Z251	1	0.34	Z251	3	1.12	Z251	0	0.00	Z251	4	0.66
L1335	1	0.34	L1335	0	0.00	L1335	0	0.00	L1335	1	0.17
S764	1	0.34	S764	2	0.75	S764	0	0.00	S764	3	0.50
Z255	11	3.75	Z255	17	6.34	Z255	1	2.27	Z255	29	4.79
R1b	1	0.34	R1b	0	0.00	R1b	2	4.55	R1b	3	0.50
R1a	1	0.34	R1a	0	0.00	R1a	3	6.82	R1a	4	0.66
I	18	6.14	I	21	7.84	I	6	13.64	I	45	7.44
E	3	1.02	E	14	5.22	E	0	0.00	E	17	2.81
J	2	0.68	J	1	0.37	J	0	0.00	J	3	0.50
T	0	0.00	T	0	0.00	T	1	2.27	T	1	0.17
G	1	0.34	G	0	0.00	G	0	0.00	G	1	0.17

3.2 1000 Genomes Comparison Data

It should be noted that the 1000 Genomes data [58] does not contain a specific Irish population in its Y-chromosomal dataset. The closest populations that an Irish individual may be categorized under are the CEU (Utah Residents with Northern and Western European Ancestry) and GBR (British in England and Scotland) populations. Table 4 contains the lists the population included in the 1000 Genomes Project's Y-chromosome data [58] and the sample sizes for each population.

Table 4: The populations and their codes included in the 1000 Genomes Project's Y-chromosome data [58] and the sample sizes for each population

Code	Population	Super Population	Pop. Size
ACB	African Caribbeans in Barbados	African	47
ASW	Americans of African Ancestry in SW USA	African	26
BEB	Bengali from Bangladesh	South Asian	42
CDX	Chinese Dai in Xishuangbanna, China	East Asian	50
CEU	Utah Residents (CEPH) with Northern and Western European Ancestry	European	49
CHB	Han Chinese in Beijing, China	East Asian	46
CHS	Southern Han Chinese	East Asian	52
CLM	Colombians from Medellin, Colombia	Ad Mixed American	43
ESN	Esan in Nigeria	African	53
FIN	Finnish in Finland	European	38
GBR	British in England and Scotland	European	46
GIH	Gujarati Indian from Houston, Texas	South Asian	58
GWD	Gambian in Western Divisions in the Gambia	African	55
IBS	Iberian Population in Spain	European	54
ITU	Indian Telugu from the UK	South Asian	60
JPT	Japanese in Tokyo, Japan	East Asian	56
KHV	Kinh in Ho Chi Minh City, Vietnam	East Asian	46
LWK	Luhya in Webuye, Kenya	African	45
MSL	Mende in Sierra Leone	African	42
MXL	Mexican Ancestry from Los Angeles USA	Ad Mixed American	33
PEL	Peruvians from Lima, Peru	Ad Mixed American	41
PJL	Punjabi from Lahore, Pakistan	South Asian	48
PUR	Puerto Ricans from Puerto Rico	Ad Mixed American	54
STU	Sri Lankan Tamil from the UK	South Asian	55
TSI	Toscani in Italia	European	53
YRI	Yoruba in Ibadan, Nigeria	African	52

The 1000 Genomes dataset was used to compare the frequency of R1b in the rest of the world's population to that of its frequency in Ireland (the samples collected in this study) to reinforce the idea that R1b is more prevalent in the Irish population, and that the subclades chosen for the R1b-L11 Multiplex Assay are more specific to Ireland. The Y-haplogroup R1b occurs in only 17.4% of the global samples in the 1000 Genomes data, including both European and non-European populations, compared to 88.3% of samples in the Irish population (IP) data collected in this study (including both Cohort 1 and Cohort 2) belonging to R1b. A visual representation of the Y-haplogroup frequencies can be seen in Figure 3, while the IP data is used to represent Ireland.

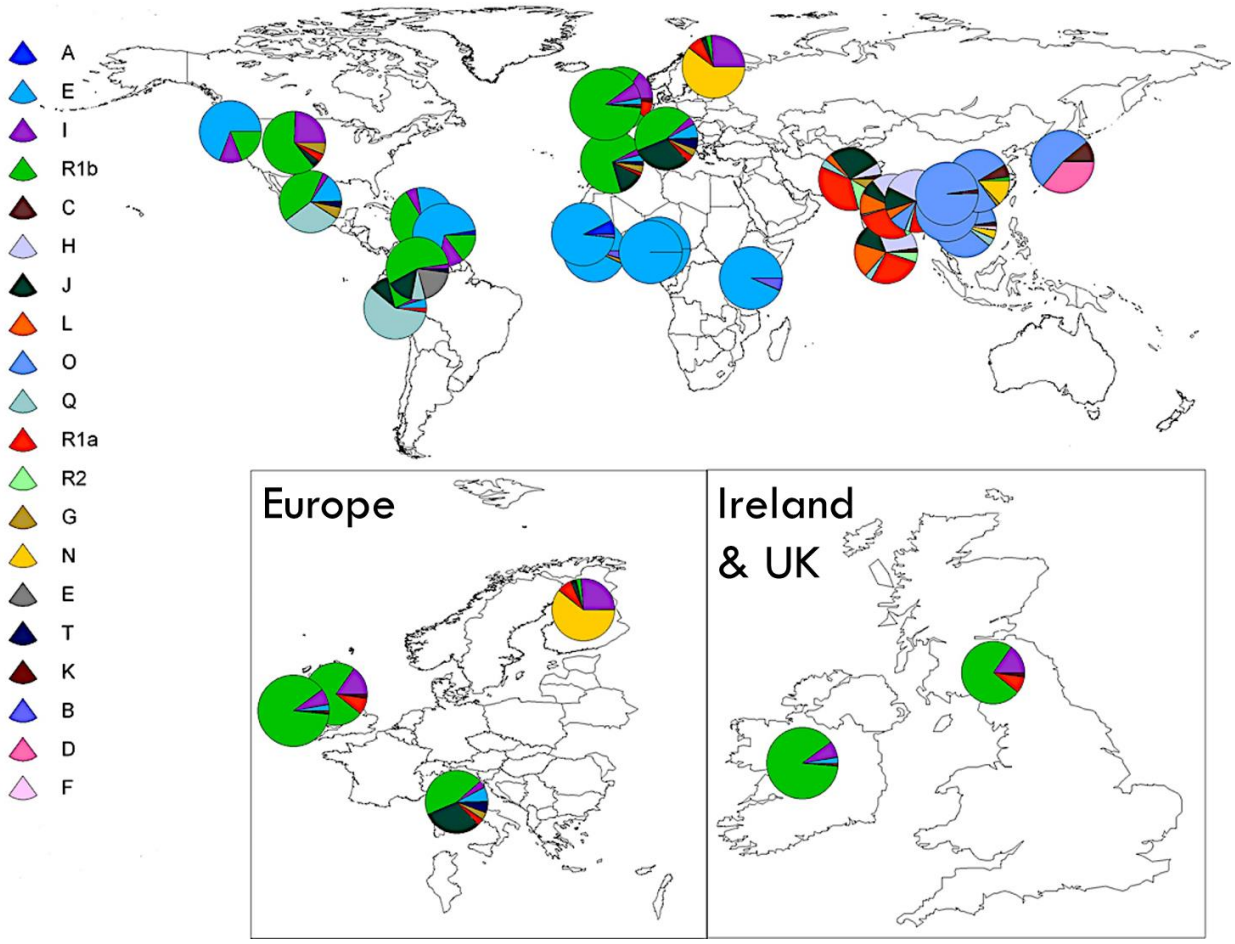


Figure 3: A global distribution of Y-haplogroups by country using data from the 1000 Genomes Project. The distribution of Y-haplogroup frequencies for Ireland was taken from the IP data.

The populations containing the Y-haplogroup R1b or a subclade of R1b are ACB (African Caribbeans in Barbados), ASW (Americans of African Ancestry in SW USA), CEU (Utah Residents with Northern and Western European Ancestry), CLM (Colombians from Medellin, Colombia), FIN (Finnish in Finland), GBR (British in England and Scotland), IBS (Iberian Population in Spain), MXL (Mexican Ancestry from Los Angeles USA), PEL (Peruvians from Lima, Peru), PUR (Puerto Ricans from Puerto Rico), and TSI (Toscani in Italia). These populations are all either European or have a historical connection with Europe. Excluding CEU and GBR, R1b occurs, on average, in 36.1% of the samples in these populations. However, the majority of the samples (89.3%) that are in the Haplogroup R1b do not fall into the subclade R1b-L21 or in the subclades below R1b-L21 in the phylogenetic tree. A visual representation of the R1b subclade frequencies can be seen in Figure 4, while the IP data is used to represent Ireland.

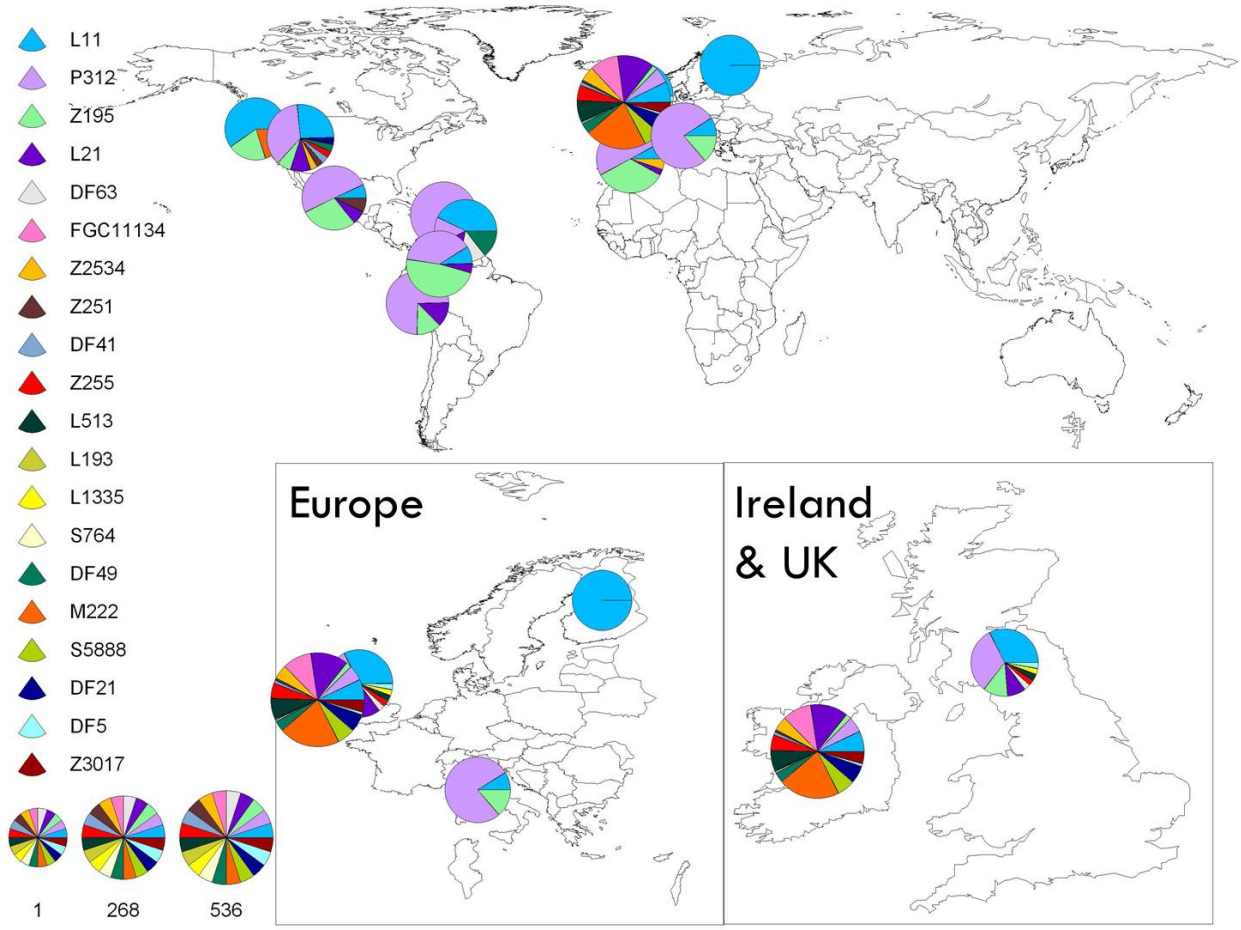


Figure 4: A global distribution of Y-haplogroup R1b subclades by country using data from the 1000 Genomes Project. The distribution of Y-haplogroup R1b subclade frequencies for Ireland was taken from the IP data.

Of the samples in the GBR population, 73.9% belong to R1b, with the rest of the population falling into Y-haplogroups I1, I2, J2 and R1a. Many of the samples (76.5%) that are in the R1b are above R1b-L21. However, the subclades that are seen at or below R1b-L21 (L21, DF63, Z255, L513, L1335 and DF5) are theorized to be linked to British and Scottish populations, apart from the single sample that fell into the subclade Z255.

Of the samples in the CEU population, 61.2% belong to R1b, with the rest of the population falling into Y-haplogroups G2, I1, I2, J2 and R1a. While CEU is the closest population in the 1000 Genomes dataset to the Irish population, it is important to note that it includes multiple countries falling in Northern and Western Ancestry. Though the majority of the samples (70%) belonging to R1b are above R1b-L21, which is less than in the GBR population, the subclades seen below R1b-L21 are believed to be linked with ancestral Irish heritage (Z2534, Z251, DF41, Z255, DF49, and DF21).

As stated, of the samples in IP data, 88.3% belong to R1b. The rest of the population falls into Y-haplogroups I, E, R1a, G, J and T. Only a small percentage of the samples (12.6%) are above L21, while only 3 samples belonging to R1b could not be classified with the R1b-L11 Multiplex Assay (meaning they are located in R1b but above R1b-L11). All Y-SNPs used in the R1b-L11 Multiplex Design were seen in the population apart from DF63 and L193, which are not specifically linked to the Irish population and were chosen as controls to distinguish other origins that may be present within the Irish population.

This comparison demonstrates that the Y-haplogroup R1b, and more specifically the subclade R1b-L21, is highly specific to the Irish population, with higher frequencies than seen in other populations within the 1000 Genomes dataset. This comparison also suggests that the R1b-L11 Multiplex Design assay created could help distinguish if a male individual may be of Western Europe, even Irish origin, which is further illustrated in the queried analysis of JB-55.

3.3 Y-Haplogroup I

As discussed earlier, the most frequent Y-haplogroups in Norway are I1, R1a, R1b and N, respectively, with Y-Haplogroup I1 specifically pertaining to Viking ancestry [52, 72]. Y-haplogroups I and R1a are believed to be representative of native Norwegians, while R1b may have two origins based on its subclade. One origin is believed to be from Germanic settlers moving into Norway from what is considered present day Germany [73], with these individuals typically belonging to the subclade R1b-U106 [74], which is a subclade under R1b-L11. The second origin is believed to be representative of Norway's history with Ireland and those who were either captured or relocated willingly.

The likelihood of a surname containing the Y-haplogroup I in the Irish population is not significantly increased when separating Irish surnames by proposed Norse origin and proposed Irish origin. Y-haplogroup I appears in 6.1% of the samples in the proposed Irish surname group, while appearing in 7.8% of the samples in the proposed Norse surname group. The presence of the Y-haplogroup I heavily suggests the presence of Vikings in Ireland. However, with a similar frequency of Y-haplogroup I in both groups, this suggests some proposed Irish origin surnames may actually be Norse, with the reverse being true for the proposed Norse origin surnames. Upon further examination of the Y-haplogroups, an agreement or disagreement of the proposed origin

of these surnames may be reached, which has been done in this study. A sample belonging to the proposed English surname group is more likely to belong to Y-haplogroup I, with 16.1% of samples falling into this group. The increase in the number of individuals in this group belonging to Y-haplogroup I may have occurred during the Norman invasion, where Vikings settled along the coasts of northwest England [75]. The Y-haplogroup would have then been brought back the Ireland during the Anglo-Norman invasion. This data is found in Table 3 above.

3.4 Allele Frequencies

All 630 male individuals were run with the Promega® Powerplex Y23 kit to type 23 different Y-STR markers. Only unrelated members were used for genetic analysis to prevent any bias in results. The online tool STRAF was used to calculate allele frequencies, which can be found in Appendix D. Genetic parameters of the Irish population can be seen in Table 5. The locus with the highest discriminatory power (PD) is DYS385, with the next highest three loci being DYS456, DYS576 and DYS481, with probabilities of 0.8556, 0.7231, 0.7206 and 0.6912, respectively. The loci with the least discriminatory power DYS393, with a probability of 0.1814. The same pattern is shown regarding genetic diversity (GD), meaning that DYS385 is the most variable Y-STR in the Irish population with a GD of 0.8189. DYS393 is the least variable loci with a GD of 0.1818. The Polymorphic Information Content (PIC) value measures how informative a genetic marker is for linkage studies [76]. Again, the same pattern can be observed, with DYS385 having the highest PIC value, meaning this locus contains the most allelic variation with a PIC value of 0.8020. DYS393 contains the least allelic variation with a PIC value of 0.1756. The match probability (PM), or the probability of a match occurring between two unrelated individuals [65], follows the opposite trend. DYS385 has the lowest PM value at 0.1444, meaning two unrelated males are unlikely to have the same allele at this locus. DYS393 has the highest PM at 0.8186, meaning two unrelated males are more likely to share the same allele at this locus. These results are comparable to a previous study [59] on the Irish population.

Table 5: Genetic parameters of the Irish population using 23 Y-STR loci generated by STRAF [65].

Locus	N	GD	PIC	PM	PD
DYS385	513	0.8189	0.802	0.1444	0.8556
DYS456	471	0.7246	0.6712	0.2769	0.7231
DYS576	502	0.7221	0.6703	0.2794	0.7206
DYS481	496	0.6926	0.6533	0.3088	0.6912
DYS549	488	0.6624	0.5984	0.339	0.661
DYS570	503	0.647	0.6055	0.3542	0.6458
DYS439	470	0.6442	0.587	0.3572	0.6428
DYS458	493	0.6278	0.582	0.3735	0.6265
DYS390	484	0.6229	0.5621	0.3784	0.6216
DYS533	475	0.6125	0.5349	0.3888	0.6112
DYS389II	470	0.5924	0.5425	0.4089	0.5911
DYS448	482	0.5881	0.5167	0.4132	0.5868
DYS392	466	0.568	0.5133	0.4332	0.5668
DYS391	497	0.512	0.4095	0.4891	0.5109
DYS635	493	0.4703	0.4396	0.5306	0.4694
YGATAH4	459	0.3925	0.3529	0.6083	0.3917
DYS389I	494	0.3716	0.3368	0.6292	0.3708
DYS437	453	0.3423	0.3116	0.6585	0.3415
DYS643	456	0.3413	0.3254	0.6594	0.3406
DYS19	463	0.2978	0.2717	0.7029	0.2971
DYS438	479	0.2933	0.2693	0.7073	0.2927
DYS393	504	0.1818	0.1756	0.8186	0.1814

3.5 Multidimensional Scaling Plot

The Arlequin package [63] was used to calculate the genetic distance between differing populations (R_{ST}). For this comparison, the collected Irish data was separated into three groups: Irish Suspected Surnames (ISS), which included Irish individuals with surnames of proposed Irish origin, Irish Norse Suspected Surnames (INSS) which included Irish individuals with surnames of proposed Norse origin and Irish English Suspected Surnames, which included Irish individuals with surnames of proposed English or Scottish origin. These three groups were then compared to nine other populations: Ireland, the United Kingdom, Norway, Wales, the Netherlands, Sweden, Finland, Denmark and Germany. Ireland, Norway and the UK were chosen to discover if the ISS group was more genetically similar to Ireland, if the INSS groups was more genetically similar to

Norway and if the IESS group was more genetically similar to the UK, based on the Y-STR profiles. Wales was chosen based on its geographic proximity to Ireland. Denmark, Finland and Sweden were chosen based on their geographic proximity to Norway. Germany and the Netherlands were chosen based on their geographical location between the United Kingdom and Norway. The resulting R_{ST} values can be seen in Table 6. Using XLSTAT [64], a Multidimensional Scaling (MDS) plot was generated using the R_{ST} values between the ISS, INSS and IESS groups and other populations with a stress value of 0.160. A stress value is used to determine how well the MDS graph is actually representing the data, meaning how well the genetic distances across populations is mapped [77], much like a goodness-of-fit measurement. A stress value of 0.160 lies between a fair (0.1) and poor (0.2) score. This may be caused by an imbalance in sample sizes, as the sample sizes in the collected Irish data are limited to participants and quality of generated Y-STR profiles, and the comparison data is limited to information available (ISS *n*=241, INSS *n*=168, IESS *n*=14, Ireland *n*=701, UK *n*=1057, Norway *n*=1121, Wales *n*=118, Netherlands *n*=2085, Sweden *n*=261, Finland *n*=416, Denmark *n*=185, Germany *n*=1718). Another cause may be due to the comparison being comprised of populations with very similar Y-STR data.

Table 6: Population Structure: R_{ST} calculations between ISS and INSS with nine populations

	Germany	Denmark	Finland	Sweden	Netherlds.	Wales	Norway	UK	Ireland	IESS	INSS	ISS
ISS	0.113	0.170	0.290	0.207	0.080	0.028	0.203	0.049	0.003	0.037	0.001	0.000
INSS	0.109	0.161	0.287	0.201	0.080	0.038	0.201	0.049	0.004	0.032	0.000	
IESS	0.033	0.053	0.201	0.085	0.017	0.025	0.107	0.007	0.013	0.000		
Ireland	0.099	0.158	0.281	0.193	0.067	0.015	0.187	0.033	0.000			
UK	0.033	0.061	0.194	0.084	0.013	0.028	0.096	0.000				
Norway	0.033	0.021	0.172	0.023	0.051	0.167	0.000					
Wales	0.090	0.126	0.229	0.163	0.057	0.000						
Netherlands	0.020	0.020	0.164	0.036	0.000							
Sweden	0.023	0.004	0.130	0.000								
Finland	0.158	0.160	0.000									
Denmark	0.012	0.000										
Germany	0.000											

Shown in Figure 5 below, the IESS group is genetically closer to the United Kingdom group than the Ireland group based on modern Y-STR profiles. IESS falls between the Netherlands and the United Kingdom, suggesting individuals in the IESS are more related to populations east of Ireland, closer toward central Europe. The ISS and INSS groups are not significantly different to each other

and overlap. Neither of the two groups are significantly different to Ireland. As such, Irish males with surnames of proposed Norse origin do not appear to be more genetically similar to present day Norse males than Irish males with surnames of proposed Irish origin, based on the 23 Y-STR loci used in this study. However, prominent Norwegian ancestry has been shown in the Irish population through admixture analysis [53]. As Y-STR variations are more reflective of the last 1,000 years, Y-SNPs provide a better assessment of older mutations and specific clades within Y-haplogroup R's evolutionary history.

Though Y-STRs are more useful in forensics and modern population genetics, they are not as ancient or established as Y-SNPs [42]. As shown in the subclade data in Table 3, many individuals belong to Y-haplogroups I and R1a, highly suggesting the presence of Vikings in Ireland. Vikings are representative of ancient Norse DNA. Y-STR profiles are representative of modern-day DNA, and, as such, are not as well suited for this comparison. The main applications of Y-STRs in this study was to gain a sense of the variation that exists in the population of modern-day paternal Ireland and to determine if the collected data contained close relatives to ensure the Y-haplogroups seen in the collected surnames were not being overestimated.

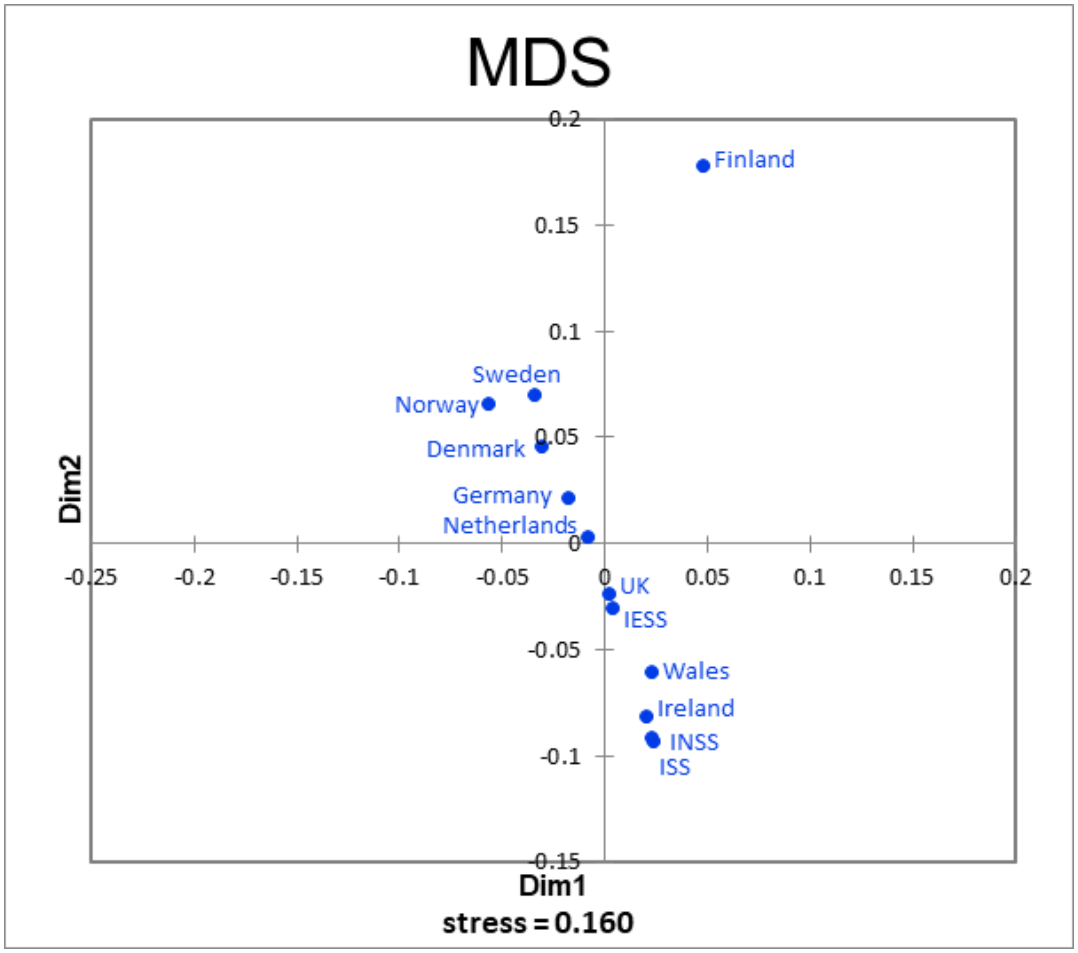


Figure 5: Multidimensional Scaling (MDS) plot comparing ISS and INSS to nine other populations. The ISS group represents Irish males in the collected data set with surnames of proposed Irish origin. The INSS group represents Irish males in the collected data set with surnames of proposed Norse origin. The IESS group represents Irish males in the collected data set with surnames of proposed English origin.

3.6 Principal Component Analysis

A Principal Component Analysis (PCA) is used to visualize variations in a dataset and to see patterns that may be hard to observe. The PCA creates a new axis to maximize variation in the data, thus creating the principal components for the X and Y axis [78].

A PCA plot was generated with the online tool STRAF [65], using the Y-STR profiles from the collected Irish set. However, PCA plots are best viewed using a three-dimensional plot to best view variation. As such, two separate two-dimensional graphs were created, which can be seen in Figures 6 and 7.

Viewing the PCA plot using the first and second principal components in Figure 6, three distinct groups can be seen. The first group is comprised of individuals belonging to the Y-haplogroup

R1b subclades, excluding R1b-M222 and its subclade R1b-S588. Individuals belonging to these two subclades are located farther down on the plot in a second group, appearing to be separated from the rest of the Irish population, including the direct ancestor of M222, R1b-DF49. The third group is comprised of individuals belonging to the Y-haplogroups found in Irish population other the R1b (I, R1a, E, J, T and G). Graphing with these two principal components, individuals belonging to R1b contain a Y-STR profile distinct from individuals in other Y-haplogroups, showing a separate ancestry. However, individuals belonging to M222 and S588 also share a Y-STR profile distinct from the rest of the Irish R1b population. This shows that these individuals may be a part of a distinct and prominent ancestry, supporting the belief of M222 belonging to the Uí Néill kingdom [9, 50], a highly influential subset of individuals in Irish history.

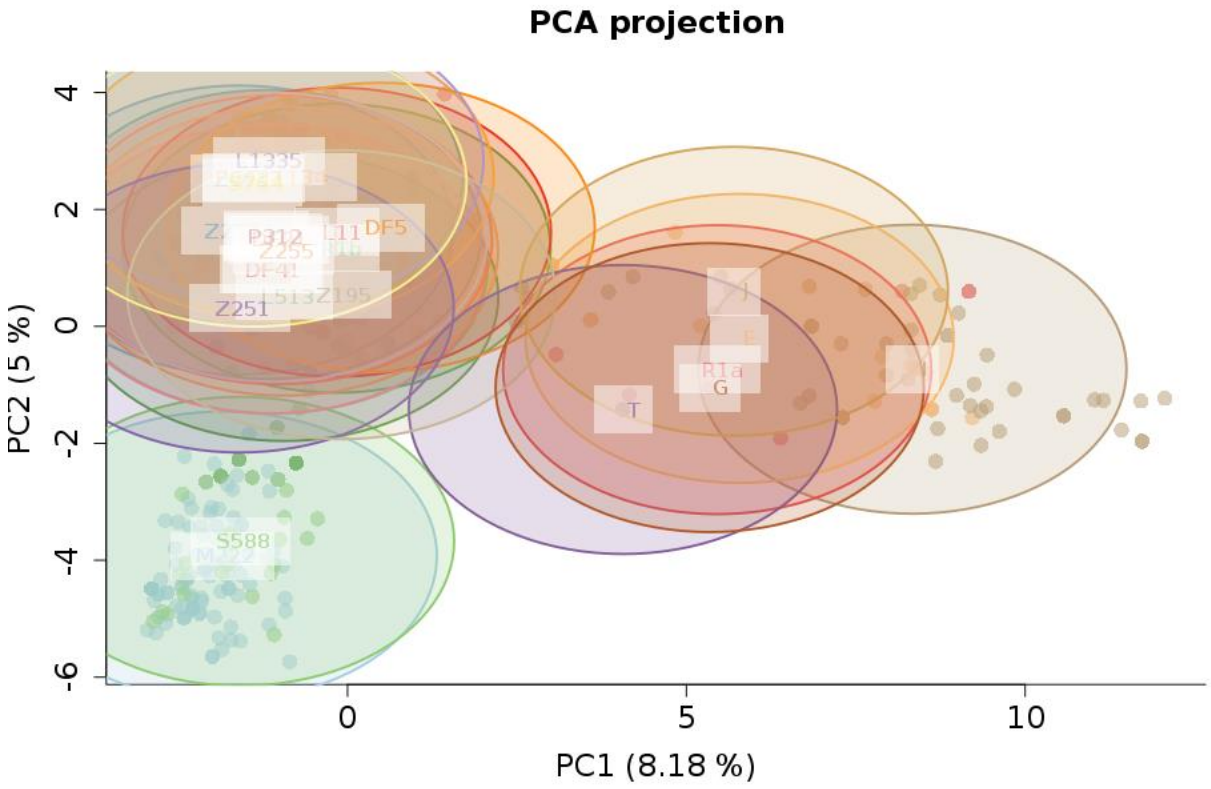


Figure 6: A PCA plot of the collected Irish data separated by subclades and Y-haplogroups using the principal components 1 and 2.

Viewing the PCA plot using the first and third principal components in Figure 7, six distinct groups can be seen. The first is comprised of all samples belonging to R1b, including M222 and S588. The other five are comprised of the Y-haplogroups found in Irish population other the R1b (I, R1a, E, J, T and G), which are separating from each other. Again, graphing with these two principal

components, individuals belonging to R1b contain a Y-STR profile separate to individuals in other Y-haplogroups, showing a separate ancestry. Individuals in the other five Y-haplogroups are also separated from each other, showing their own ancestry/lineage.

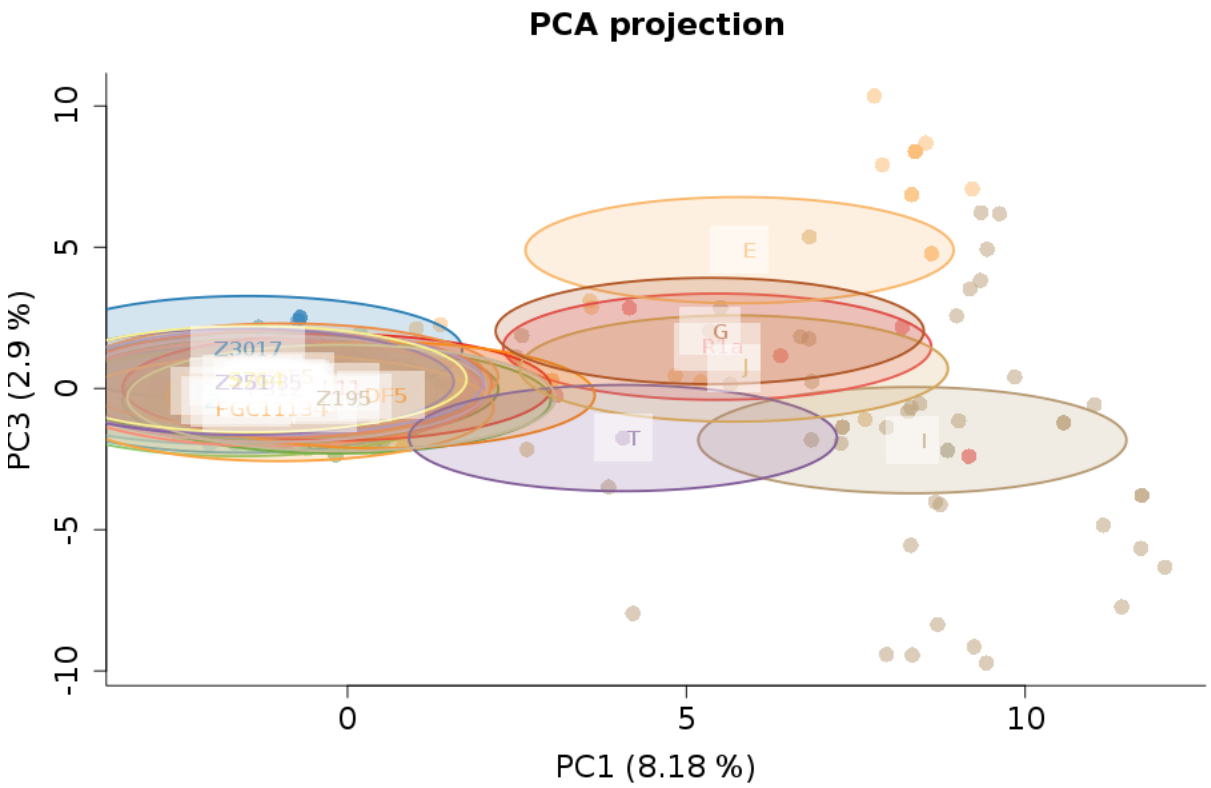


Figure 7: A PCA plot of the collected Irish data separated by subclades and Y-haplogroups using the principal components 1 and 3.

3.7 Surname Frequencies and Subclades

In other countries, such as Great Britain and Spain, surname frequency is related to the number of founders and Y-haplogroups found within that surname [6, 79]. Unrelated male individuals with surnames that are less common in the population are more likely to belong to the same broad haplogroup as an individual with the same surname, demonstrating that these unrelated males contain a common ancient ancestor (the founder of the surname). These surnames are also more likely to only contain one founder. The reverse is true for individuals with more common surnames, which may contain multiple haplogroups and founders. However, this pattern cannot be applied to the Irish population, where most male individuals belong to Y-haplogroup R1b regardless of surname frequency. This is observed when viewing the collected IP data and is also true with

regards to subclades. There is no correlation between high-frequency surnames sharing a subclade. While multiple surnames have their own unique patterns of subclade occurrence, this cannot be linked to surname frequency. Due to Ireland's historical and geographical seclusion to outside countries and extensive history regarding surnames, this is not surprising. Some popular names may have had multiple founders, particularly the profession-based names (such as Butler or Coppinger), but this is not the extent of issues causing this occurrence. This issue may also be due to anglicization of many Irish surnames by the British. As mentioned earlier, other issues can be linked to name changes, nonpaternity, adoption, matrilineal surname transmission and record falsification. As such, the history of each name may need to be consulted as well, studying both the historical origin of the surname with its frequency in the population.

3.8 Province Distributions and Subclade Patterns

Individuals who provided the province of origin for either their paternal grandfather or themselves were placed in one of four groups: Ulster (n=99), Connacht (n=101), Munster (n=151) or Leinster (n=146). A clear pattern was observed when comparing subclade frequency with geographical location in Ireland, shown in Figure 8. Individuals with R1b-P312 and R1b-Z195 haplogroups were found in Connacht, Leinster and Munster but not Ulster. These subclades are more ancient and less specific to the Irish population. Individuals belonging to these established subclades that were not included in the R1b-L11 Multiplex Assay, may instead originate from other European populations. Therefore, males who travelled into Ireland from another country may have been more likely to settle into these provinces as opposed to Ulster, which was a province anciently controlled by the Uí Néill's kingdom for several hundreds of years and could have prevented integration.

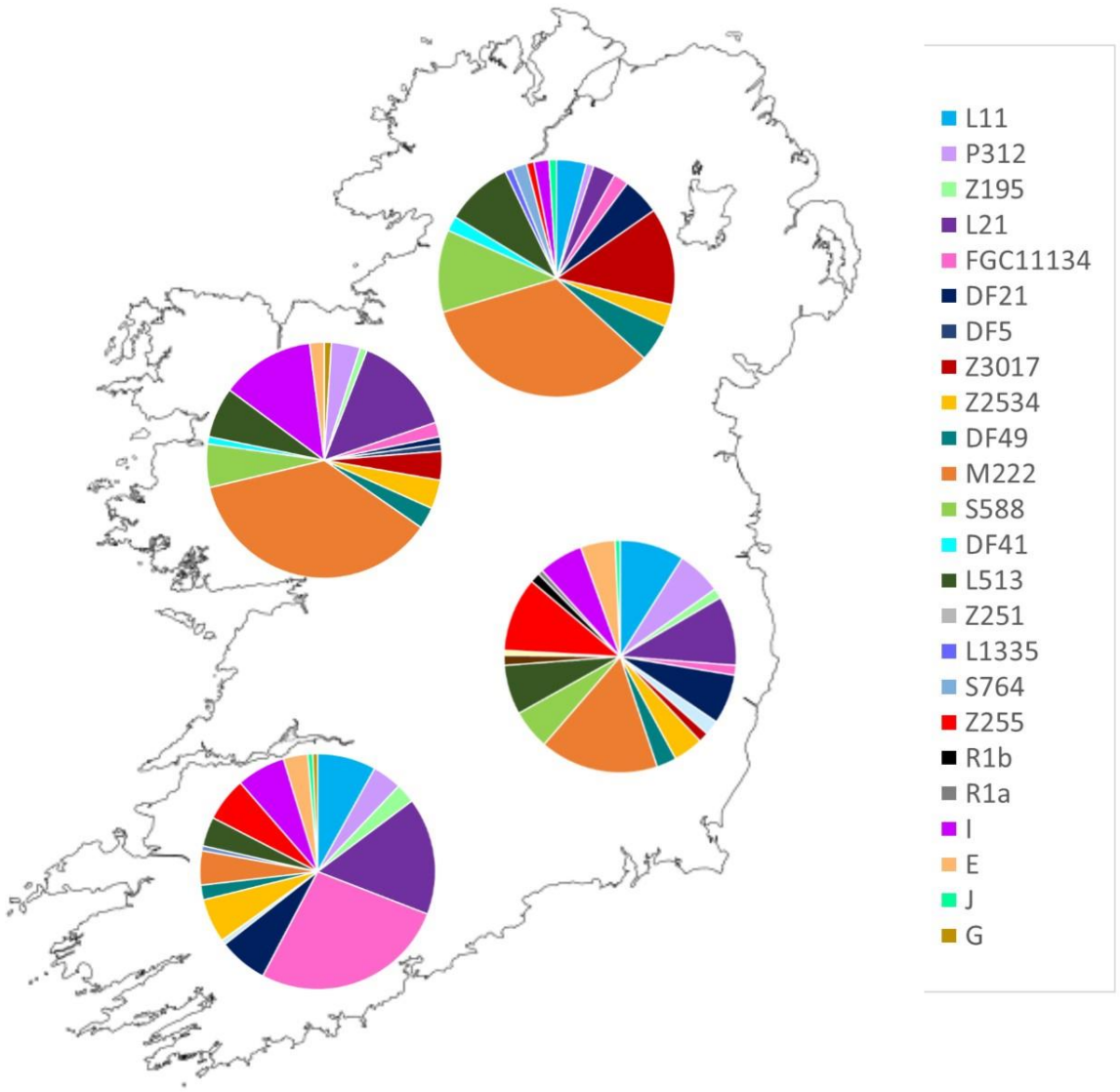


Figure 8: The distribution of Y-haplogroups and subclades of the Irish population by province.

Subclade M222, which has been linked to the North West Irish Type I group [50], is the most common haplotype in both Ulster (33.33%) and Connacht (36.63%), reinforcing the finding of M222 belonging to North West Ireland. However, it is also found at a high frequency in Leinster (16.44%), though not as significantly as it is seen in the previous two provinces. While M222 is seen in Munster, it is not at a significant frequency (4.64%). This pattern helps support the idea that the subclade M222 originated in North West Ireland [9], and was carried Southeast, presumably by the *Uí Néill* family, before being stopped as they travelled south, likely by the

Eoganachta kingdom. Subclade S588 is a mutation downstream of M222 and follows a similar pattern to its ancestral subclade: most common in Ulster (11.11), with decreasing frequency in Connacht (5.94%) and Leinster (5.48%) but is then non-existent in Munster.

Subclade FGC11134, which has been linked to the South Irish Type II group by Dr. Ken Nordtvedt, is the most common subclade found in Munster (26.49%) and is only found at low frequencies in Ulster (2.02%), Connacht (1.98%) and Leinster (1.37%). This lends support to the hypothesis that the subclade FGC11134 originated in Southern Ireland but did not travel outside of Munster. It also agrees with historical reports of the Eoganachta, who preferred to rule by means of wealth and federalizing strategies rather than warfare [12].

Subclade Z2534, which has been linked to the South West Ireland Type III [50], does not show a distinct pattern regarding geographical location. While Z2534 appears to be seen at a slightly higher frequency in Munster (5.96%), this is not significantly higher than in the other three provinces (4.11% in Leinster, 3.96% in Connacht and 3.03% in Ulster). However, the Dalcassian tribe, which is associated with the Irish Type III group [50], was smaller than the Eoganachta and the Uí Néill's [12], suggesting the genetic signature of the Dalcassians may not be as well spread as the larger dynasties. This clan also fought throughout Ireland: in the north of Ireland against the Uí Néill kingdoms and in Leinster against the Norse-Irish alliance [15]. Further research should be done to test the true location of Irish Type III, while historical documentation may be referenced to understand the proper link of Z2534 to the Dalcassian families.

Subclade Z255, which is often referred to as the Irish Sea Haplotype, is only seen in Leinster and Munster, though at a higher frequency in Leinster (10.27% in Leinster, 5.96% in Munster and 1.01% in Ulster). The Irish Sea Haplotype is named as such because it is found most commonly along the coasts touching the Irish Sea in Ireland, Britain, Norway, Sweden and France, which has been researched by Joe Flood, mathematician and administrator of the Cornwall Y-DNA Geographic Project, in unpublished work. While Z255 has not been linked to a specific Irish family, this subclade may be indicative to a prominent Leinster kingdom. However, this subclade is also commonly found when viewing samples with a proposed Norse origin surname (see Table 1). The first Viking towns were founded along the eastern and southern half of Ireland, such as Dublin and Wexford [80]. This pattern may also show a possible link between Norse invaders and the Irish population.

Subclade Z3017 is seen at a higher frequency in Ulster (13.13%) and a lower frequency in Connacht (3.96%) and Leinster (1.37%) while being absent in Munster. Z3017 has been proposed to have a link with the British population, which would explain why this subclade is more commonly found in Ulster, as part of Ulster now belongs to Northern Ireland, which is and was heavily influenced by the British population.

Y-haplogroup I is commonly linked to the Norse population and to the Vikings [72]. Y-haplogroup I is seen at its highest frequency in Connacht (12.87%), an area where the some of the first Viking raids took place [81]. However, with the first major towns in Ireland, which are found in Munster and Leinster, Y-haplogroup I should be more common in these provinces but is only found in 6.62% and 6.16% of the samples respectively. This is still a large indicator of the presence of Vikings, though not in the areas expected. As stated earlier, it is believed that these towns were ruled by the Norse noblemen but mainly inhabited by indigenous Irish [21]. This haplogroup is very rare in Ulster (2.02%) and is only seen in one of the samples. Subclade R1b-L11 is also found at high frequency in Munster (7.95%) and Leinster (8.90%). Subclade R1b-U106 is below L11 [74], but, because it is not included in the R1b-L11 assay, will be genotyped as L11. U106 is theorized to be representative of Vikings with a Germanic ancestry, meaning their ancestors travelled though present-day Germany to reach Norway and Denmark. The high presence of L11 in both Munster and Leinster support this theory and suggest that the standards for determining the presence of Vikings in Ireland may need to be broadened to include specific subclades within R1b.

The Irish DNA Atlas reported a general trend of gene flow across Ireland with three areas of low migration [53]. The first reported barrier is in the West of Ireland, including the coast of Connacht, where the first Viking raids occurred. This is perhaps where the large proportion of Y-haplogroup I comes from, originating in Connacht but not traveling far from the coasts. The second barrier refers to the relatively low genetic migration near the Leinster Munster border. This is seen in the high proportion of subclade FGC1134, which is almost exclusive to Munster, and the low proportion of subclade M222 in Munster. When viewing Leinster, it is the most diverse of the four provinces, and contains the highest proportion of R1b-L21 subclades and non-R Y-haplogroups. The distributions of these subclades and Y-haplogroups is also more even than the other three provinces, with no one subclade dominating the others. This observation may in part be due to Leinster containing the main port into Ireland, and therefore being the easiest province to travel to

and settle in. The third barrier is within Ulster, including Scotland. This is perhaps seen in the large proportion of subclade Z3017 in Ulster.

Y-SNPs are ancient genealogical markers and represent ancient male lines. Provincial information shown in Figure 8 is representative of where the grandfather of the male giving the sample was born (if this was not provided, the father's birth place would be used followed by the male himself). Often, the male giving the sample was born in the same province as his father and grandfather. Even using more modern DNA, historical patterns are still seen in Ireland's present-day genetic makeup. A low rate of immigration between the provinces is seen in the collected data, which is in agreement with the DNA Atlas. The DNA Atlas also found a link between Ulster and Connacht, which can be expected due to the patrilineal and military ties between the two provinces [53]. This similarity between the two provinces is noticeable when viewing the subclades. Ulster and Connacht contain the highest proportions of subclades M222 (33.33% and 36.63%), S588 (11.11% and 5.94%) and Z3017 (13.13% and 3.96%) and given these regions patrilineal history, this seems in agreement with the proposed 'Irish Types' in the region.

3.9 Proposed Irish Origin Surnames

The following proposed Irish surnames are found to be of Irish origin due to particular subclade frequencies: Boland, Curley, Duggan, Gilmore, Kennedy, Loughlin, Loughman, McArdle, McCabe, McGettigan, McGinty, McKeever, McMahon, McNea, O'Donovan, O'Mahony, Reilly and Sullivan.

The surname Butler is discussed in further detail below. Butler is a profession-based surname where both the profession and the surname linked to it came from England during the Anglo-Norman Invasion. Having multiple founders from multiple origins contributing to the surname, this name cannot be placed into a specific origin and is representative of Ireland after the admixture occurring from both the Viking and Anglo-Norman invasions.

As discussed below in more detail, the surname Hanly is of proposed Irish origin but the high proportion of the Y-haplogroup I present does not support this belief and therefore a re-evaluation of the surname origin should be performed, perhaps examining additional individuals to provide more information on its potential for being Norse.

The surnames Byrne, Curran, McCarthy, McManus, McSweeny, O'Loughlin and O'Toole are proposed Irish origin, but contain one or more additional Y-Haplogroups. Byrne, McCarthy and O'Toole are discussed in further detail below.

The surname Curran (*n*=17) is equally widespread throughout all four provinces in Ireland [23], and seven different R1b subclades are found within the samples (L11, P312, L21, FGC11134, M222, Z251, and Z255) with two samples belonging to Y-haplogroup I. This suggests multiple founders, and so the presence of the Y-haplogroup I cannot be used to infer Norse origin.

The surname McManus (*n*=9) is believed to be of two separate origins: one from a king of Ireland, the other originating in Ulster [23]. The surname McSweeny (*n*=6) is historically linked to the Uí Néill family [23]. Due to the history of these names, the presence of the one sample belonging to Y-haplogroup I in both surnames may be explained by adoption or nonpaternity.

The O'Loughlin's (*n*=7) are suggested to be of to be a part of the Dalcassian sept [54]. However, there is no presence of the subclade Z2534. The subclades seen are L21 and M222 (two samples each) and DF21 (one sample). The Y-haplogroups I and G each contain one sample. While it is questionable that O'Loughlin is a sept belonging to the Dalcassians, another possible origin cannot be determined due to the presence of multiple subclades and Y-haplogroups at low frequencies.

A conclusion regarding its origin could not be reached for Carr or Kilpatrick. The surname Carr (*n*=4) contains one sample belonging to Y-haplogroup I with another sample that could not be haplotyped (this sample could not be classified with the R1b-L11 Multiplex assay and the Y-STR profile contained drop outs, preventing further analysis). Due to small sample size, a proper analysis of the surname could not be reached. Both samples in Kilpatrick belong to Y-haplogroup E. These two males are closely related, differing by only one repeat at DYS385, and do not properly represent this surname group alone.

The surname Harold-Barry is uncommon from the others, as it is a hyphenated name. While the surname Harold is believed to be of Norse origin, the surname Barry is thought to be Irish [54]. As Barry is the second name, this should be the inherited name from the father. All three samples belong to R1b-L11, suggesting that this name is not of Irish origin. However, this is a small sample size and may not properly represent this name. The distribution of all Y-haplogroups in these surnames is located in Apendix E.

3.10 Proposed Norse Origin Surnames

The following proposed Norse surnames that are found to be in agreement with being of Norse origin due to the frequency of a Y-haplogroup I or R1a are Coll, Dolphin, Doyle, Gorey, Harold, Reynolds, Cotter (for reasons discussed in detail below), and Higgins. The surnames Dolphin, Doyle, Gorey, Harrold, Cotter and Higgins are discussed in further detail in another section.

Coll (*n*=5) is typically believed to be of Norse origin [54]. Two of the samples belong to Y-haplogroup I (40%). Though this is a small sample size, the large percentage of Y-haplogroup I being found is suggestive of Norse origin.

The surname Reynolds (*n*=15) is from the old Norse name Rognvald [82]. Two samples belong to Y-haplogroup I while 60% of the samples are categorized as L21. A high proportion of surnames believed to be of Norse origin from the collected data are categorized as R1b-L21 (15.3%), as shown in Table 3. This suggests there may be another subclade underneath L21 not included in the R1b-L11 Multiplex Assay specific to the Norse population. Therefore, considering the history of the surname with the genetics, Reynolds likely is of Norse origin.

The following proposed Norse surnames do not lend support to being of Norse origin due to the complete absence of a Y-haplogroup I or R1a: Kettle, McAuliffe, McCauley, McGorry, McLoughlin, McSorley, O'Rourke, Somers, Toner and Turley.

The surnames Arthur, Grimes, O'Hagan, Coppinger and Murphy present a mixture of Y-haplogroups and subclades that suggest multiple founders of multiple origins. Coppinger and Murphy are discussed in more detail below. The surname Arthur (*n*=17) contains one Y-haplogroup I sample, but also contains twelve samples from closely related males belonging to Y-haplogroup E. As most of the samples from Y-haplogroup E are believed to be from one family, this is not a good representation of the Y-haplogroups in the surname.

The surname Grimes (*n*=10) is thought to be of Norse origin, though there is not much history on the name [54]. There is one sample belonging to Y-haplogroup I, but also to Y-haplogroup J. The other four subclades present are L11 (two samples), S588 (four samples) and DF21 and DF5 (one sample each). Subclades DF21 and DF5 have been theorized to be specific to the British population and L11 is a lower resolved subclade. The presence of these subclades with Y-haplogroup I suggests that this name is of Norse/English origin and perhaps came to Ireland during the Anglo-Norman Invasion.

This is similar to surname O'Hagan (*n*=9), which has conflicting report of origin. The surname is documented as Irish [54], but Norsemen have been documented as carrying this surname [83]. The largest subclade in the surname is S588 (four samples) followed by DF21 (two samples) and then subclades L11 and M222 (one sample each). One sample belongs to the Y-haplogroup I. Again, it is suggested that this name is of Norse/English origin and perhaps came to Ireland during the Anglo-Norman Invasion. The high proportion of S588 in Grimes and O'Hagan will be discussed further detail below.

The surnames Broderick, Hammond, McDowell and Costin are undetermined. Broderick (*n*=25) contains nineteen samples belonging to L21, which has a high-resolution in the R1b tree. This suggests that the samples in this subclade may belong to a further subclade of several European origins (Norse, English, etc.), therefore a conclusion cannot be reached. The same is true of the surname Hammond, where all samples belong to lower resolved subclades (P312, L21 and Z195). Neither of these names contain the Y-haplogroup I. The best approach for these surnames would be to genotype outside the L21 design that has been performed to gain information on which direction in the tree these individuals could have developed from, as opposed to subclades below R1b-L21 which have already been analysed.

The surname McDowell (*n*=10) comes from the name Mac Dubhghaill, meaning son of Dubhghall. Dubhghall can be translated to "dark stranger" [84], a name often used to describe Norsemen [23]. Though one sample in the surname group did not belong to R1b-L11, it could not be further classified (this sample could not be genotyped with the R1b-L11 Multiplex assay and the Y-STR profile was inconclusive, preventing further analysis). A conclusion cannot be reached for this name unless more samples are collected, and further analysis can be done.

Only two samples were collected for the surname Costin. While both samples belong to L11, the small sample size does not allow the collected data to be of good representation of the surname. The distribution of Y-haplogroups in these surnames is located in Apendix E.

3.11 Proposed English Origin Surnames

The following proposed English surnames are found to be of English origin due to higher frequencies of lower resolved subclades (L11, P312, Z195 and L21) or subclades proposed to be of British or Scottish origin (L513, DF41 and DF5) Goodman, Holden and Lamb. and Taylor.

The surnames Johnston and Thompson are found to have a mixture of Norse influence, which is discussed in more detail below.

The surname Taylor contains only one sample. It was genotyped as subclade P312, but due to poor sample size this is not a good representation of the surname and no conclusion could be reached.

The surname Guy contains only one sample. As this sample was of poor quality, no information could be gathered on this surname.

3.12 In depth look at surnames with potentially significant subclade patterns

Many surnames showed a distinct pattern of subclades within R1b or Y-haplogroups outside of R1b. A full list of all the surnames and their Y-haplogroup and subclade distributions can be seen in Appendix E. A select number of surnames with significant patterns will be discussed below, helping to support the hypothesis that Y-SNPs can be used to correctly trace ancestry in Irish Surnames with a higher and more accurate resolution that Y-STRs.

3.12.1 O'Reilly

O'Reilly is one of the most frequent surnames in Ireland, and until recently did not use the prefix *O*. This sept is named as a descendant of Raghallach, originally named Ó Raghallach, and originated in Ulster [85]. The surname O'Reilly has been linked to the Uí Néill kingdom by Brian McEvoy at Trinity College in Dublin, which is often cited but has not been published [86].

There are 24 Irish samples belonging to the O'Reilly group. Of these 24, 70.83% of samples contain the subclade M222, which is linked to the Uí Néill kingdom [50]. It should be noted that five individuals who belong to M222 have identical Y-STR profiles (three of these individuals being named Reilly and the other two O'Reilly). The other subclades seen are L11, P312 and Z3017, each subclade containing two samples. The two individuals belonging to Z3017 share an identical Y-STR profile. One individual in this surname group belongs to the Y-haplogroup E, which may be due to adoption or nonpaternity. This information can be seen in Figure 9.

Due to the high presence of M222 present in the O'Reilly surname, this data supports the claim that O'Reilly belongs to the Uí Néill kingdom.

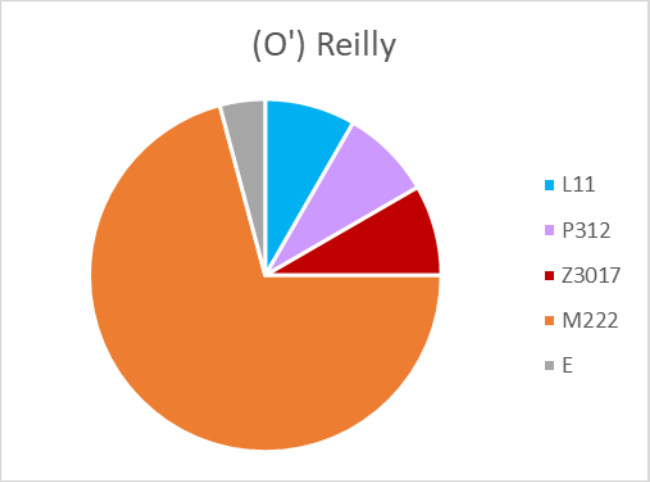


Figure 9: Y-haplogroup and R1b subclade frequency in O'Reilly (*n*=24)

3.12.2 O'Donovan and O'Sullivan

The O'Donovan family is said to have the most authentic information regarding their family history [23]. They have a pedigree of the eldest branch dating back to the Gaelic times that has been verified by the Genealogical Office and can be linked to the Eoganachta kingdom [87]. O'Sullivan is one of the most common surnames in Ireland and is the most common surname in Munster. The O'Sullivan family have a well-known origin as one of the leading families of the Eoganachta kingdom [23, 87]. All samples for O'Donovan come from men descendant from Munster, and all samples in O'Sullivan who marked their provincial information are from men descendant from Munster, showing that individuals from this prominent kingdom are still highly prevalent in their ancestral regions.

Both surnames contain a high percentage of subclade FGC11134 (O'Donovan with 66.66% and O'Sullivan with 64.29%), which is linked to the Eoganachta, based on Dr. Ken Nordtvedt's work. There are no related individuals in the O'Donovan group; however, the sample size is small (*n*=6). Two of the nine males in O'Sullivan belonging to FGC11134 have an identical Y-STR profile. Only two other subclades are present in O'Donovan (one sample each for L21 and M222). The four other subclades present in O'Sullivan are L11, Z2534, L513 (each containing one sample) and L21 (containing two samples). This information can be seen in Figures 10 and 11.

Due to the high presence of FGC11134 in the O'Donovan and O'Sullivan surnames, this data supports the historic claim that they belong to the Eoganachta kingdom. As there are verified

records reporting that the O'Donovan's and O'Sullivan's are descendants of Eoganachta, these results also demonstrate the ability of Y-SNPS to correctly verify the ancestry of individuals based on their surnames, and that subclade FGC1134 is in fact representative of the Eoganachta kingdom and the province of Munster.

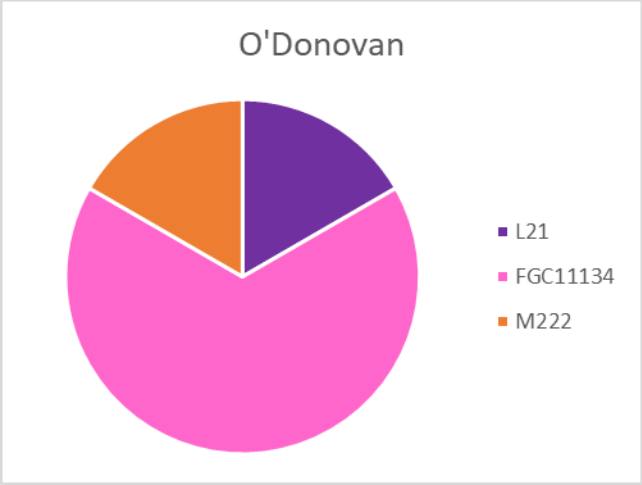


Figure 10: R1b subclade frequency in O'Donovan (*n*=6)

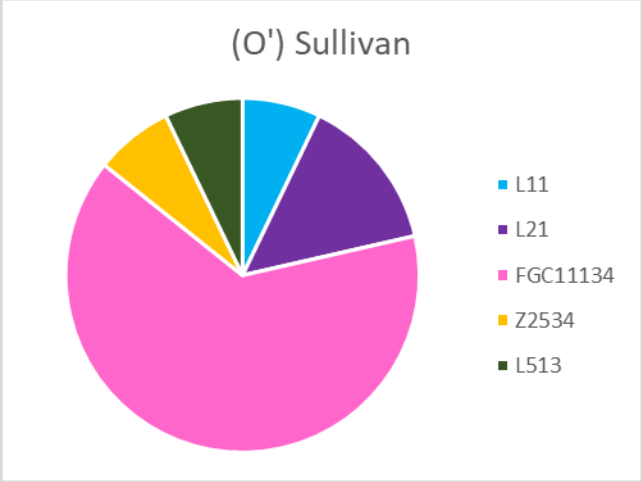


Figure 11: R1b subclade frequency in O'Sullivan (*n*=14)

3.12.3 Boland and Kennedy

The surname Boland, anglicized from the surname Ó Beollain, belongs to at least two distinct septs: the Ui Fiachrach from Connacht and the Dalcassians of Thomond, who are descendants of Mahon, brother of Brian Boru. Kennedy is another common name in Ireland. The ancestor of the

Kennedy's was Cinnéide, nephew of Brian Boru. However, the Kennedy's of Ulster are believed to be of Scottish origin as Kennedy is also a Scottish name [23].

While the most frequent subclade in both Boland and Kennedy is Z2534, which is believed to be representative of the Dalcassians [50], this subclade is not found at as high a frequency as demonstrated with M222 and FGC11134 (37.50% for Boland and 28.57% for Kennedy). Boland contains four other subclades (one sample each for DF21, M222 and L513 and two sample for Z255). Kennedy contains five other subclades (one sample in L21, two samples each in L11, L513 and Z255 and three samples in P312). Two individuals in subclades L513, Z2534 and Z255 each contain an identical Y-STR profile. These results can be seen in Figures 12 and 13.

The results for Boland are consistent with the belief of multiple founders in the surname. Individuals in Z2534 would belong to the Dalcassian kingdom. The individual belonging to M222 is likely to be representative to the Uí Fiachrach sept, which originates from Connacht and has close ties with the Uí Néill's. The individuals that belong to Z255 may be representative of a third sept originating in Leinster. The results for Kennedy seem to be consistent with the belief that this name belongs to both a Dalcassian sept and Scotland. While Z2534 would be representative of the Dalcassians, the lower resolved subclades (L11, P312 and L21) and L513 may be representative of a Scottish heritage.

As mentioned above, the Dál Cais were a small kingdom in comparison to the Eoganachta and Uí Néill dynasties [12]. This alone would be reason to suggest a smaller frequency of Z2534 seen in the Dalcassians than M222 in the Uí Néill's [50] and FGC11134 in the Eoganachta, linked by the work of Dr. Ken Nordtvedt. On top of a smaller kingdom size, Brian Borumha's direct family line was cut after the death of both his son and grandson in the Battle of Clontarf [17]. Finally, Uí Néill was a 4th century warlord [9], while Brian Borumha was not born until sometime in the 10th century [88]. The subclade M222 is there for older than Z2534 and would have had roughly six centuries to spread throughout Ireland during war and conquest before Brian Borumha was able to spread the Dalcassian line. Considering these facts, it is acceptable for surnames descended from the Dalcassians to contain a lower frequency of Z2534 than would be expected for M222 and FGC11134.

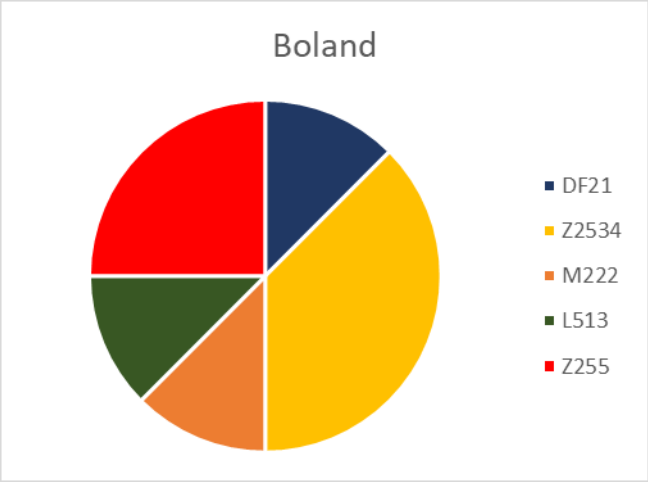


Figure 12: R1b subclade frequency in Boland (*n*=8)

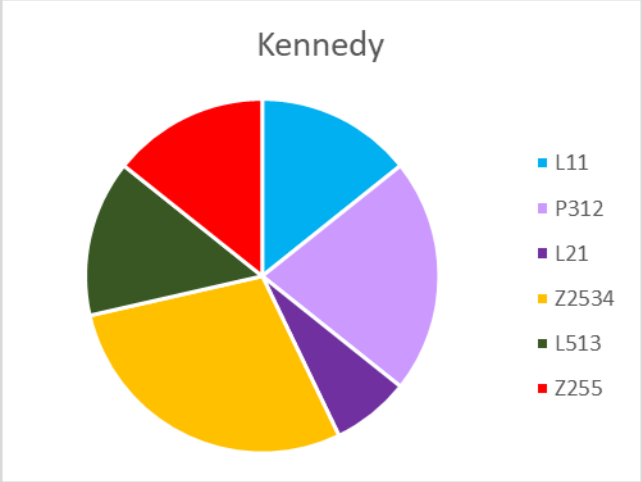


Figure 13: R1b subclade frequency in Kennedy (*n*=14)

3.12.4 Lamb

The surname Lamb is believed to be of two origins: an English name from English descent and an anglicized version of the surname O'Loan [54]. The most frequent subclade in Lamb is DF41 (37.50%), followed by L11 (25.00%). The three other subclades, Z195, L21 and DF5, contain one sample each. These results can be seen in Figure 14. While none of these subclades fall outside of R1b, they are either lower resolved subclades (L11, Z195 and L21) or subclades that are theorized to be linked to the British population (DF41 and DF5). These results suggest that individuals with

the surname Lamb are of English descent and provide a good understanding of how an English surname will appear in comparison to an Irish surname regarding Y-SNPs and subclade data.

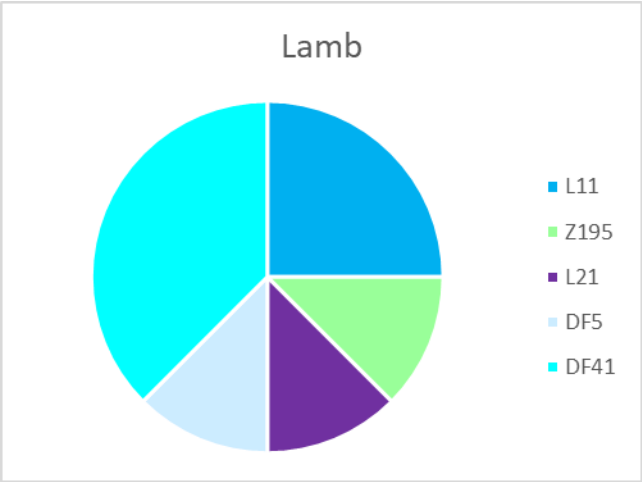


Figure 14: R1b subclade frequency in Lamb (*n*=8)

3.12.5 Dolphin

The surname Dolphin is believed to be of Norse origin and thought to have arrived in Connacht from England during the Norman invasion [54]. Though the sample size is small for this surname (*n*=4), Dolphin is an exemplary surname to demonstrate Norse origin. All individuals with this surname contain distinct Y-STR profiles, so none of the males are from the same direct family. Y-haplogroup I is the most frequent haplogroup present (75%), with the fourth sample belonging to subclade M222. These results can be seen in Figure 15. All samples in Dolphin who marked their provincial information are descend from Connacht. The individual belonging to M222 may represent a male who changed his surname to reflect allegiance, fathered a son with a Norse female or an instance of nonpaternity or adoption. These results heavily favor the belief that the surname Dolphin is Norse in origin and was introduced to Ireland in Connacht. These results also demonstrate that Y-SNPs are an accurate way to determine surname origin and that it is best to view surnames individually based on history rather than as a group of proposed origin.

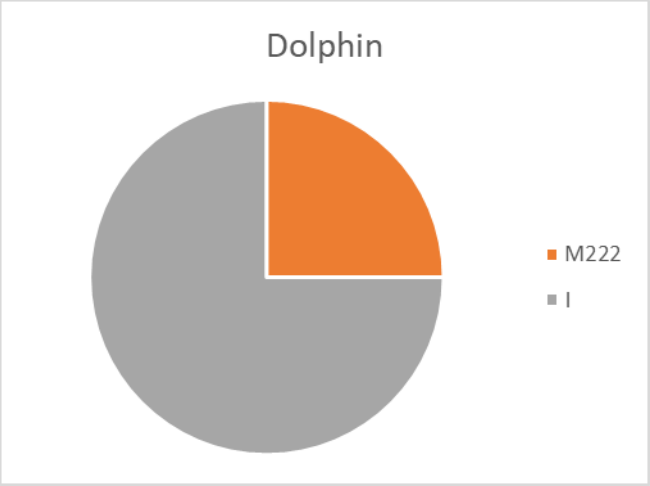


Figure 15: Y-haplogroup and R1b subclade frequency in Dolphin (*n*=4)

3.12.6 Coppinger

The surname is proposed to be of Norse origin, arriving in Co. Cork (Munster) around or before the 14th century [54], meaning the name is believed to be of Viking decent, which may have been introduced to England during the Norman conquest and then brought to Ireland. Y-haplogroup I is tied as the most frequent group present in Coppinger with subclade R1b-Z195 (35.71%). The second most frequent group is Y-haplogroup E (two samples) followed by FCG11134 and P312 (each with one sample). These results can be seen in Figure 16. Interestingly, the individuals belonging to Y-haplogroups I and E are descendant from Connacht, while the samples from individuals belonging to subclade Z195 are descendant from Munster. These results show two distinct patterns and families. Perhaps the individuals who settled in Co. Cork are of English descent and those who settled in Connacht are of true Norse origin, separate from those who settled in Co. Cork.

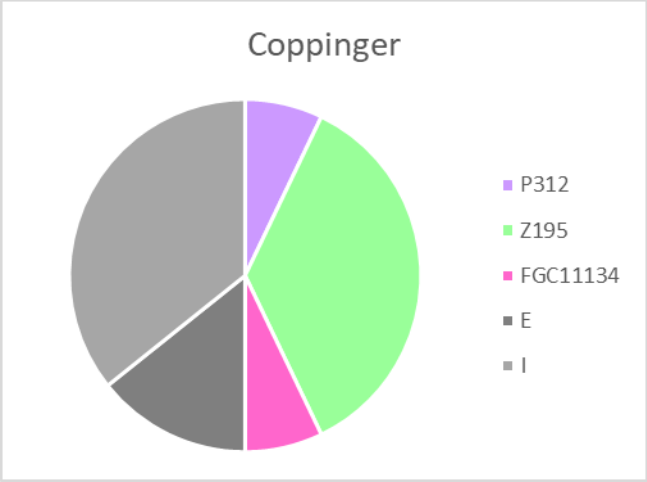


Figure 16: Y-haplogroup and R1b subclade frequency in Coppinger (*n*=14)

3.12.7 Harrold and Cotter

The surname Harold is theorized to be of Norse origin and established before the Anglo-Norman invasion. The surname Cotter comes from Mac Oitir (meaning Son of Oitir), an old Gaelic-Irish name formed from the Norse personal name Oitir [54]. Though Cotter is from a Norse personal name, much of its history is attributed to Irish families. However, these two names share a very common pattern.

The most frequent subclade in Harold is FGC11134 (62.50%) followed by the Y-Haplotype I (25.00%), with one sample belonging to L11. Two individuals belonging to FGC11134 share an identical Y-STR profile. The most frequent subclade in Cotter is L11 (55.55%) followed by the only other subclade, FGC11134 (44.45%). These results can be seen in Figures 17 and 18.

While Cotter does not contain any samples with the Y-haplogroup I (which may be due to a small sample size, *n*=9), it shows a very similar distribution to Harrold. Both surnames contain subclade L11. As mentioned above, subclade L11 may be representative of subclade R1b-U106 [74], which is shown to be a subclade for Vikings with a Germanic origin. This possibility is supported by the presence of Y-haplogroup I in Harold. All individuals who marked their provincial information are descendant from Munster, which is explanatory for the high volume of FGC11134 present. These results suggest the two surnames may come from Viking settlements in Munster, which were then inhabited by indigenous Irish.

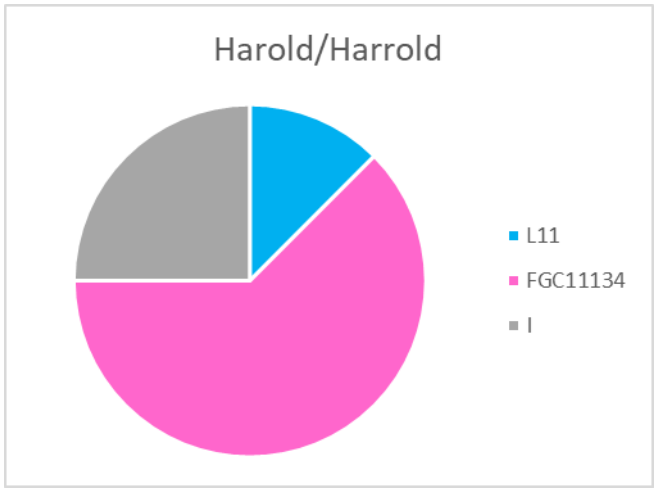


Figure 17: Y-haplogroup and R1b subclade frequency in Harold (*n*=8)

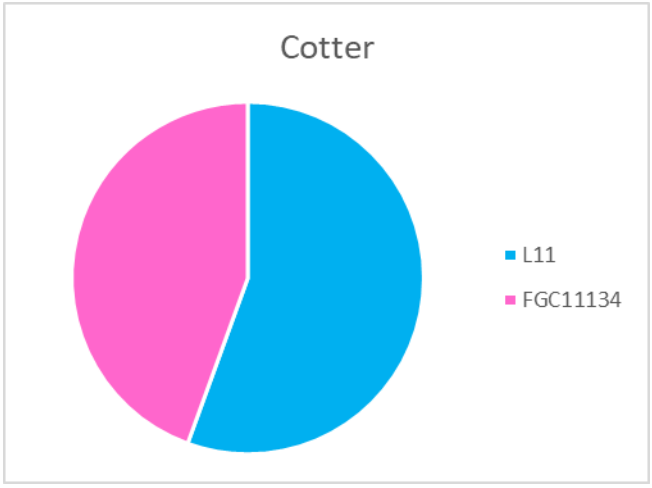


Figure 18: R1b subclade frequency in Cotter (*n*=9)

3.12.8 Doyle and Murphy

Doyle is a common Irish surname, frequently found in Leinster, which can be translated to mean dark foreigner, a name typically used to refer to Norsemen or Scandinavians. This surname is not found in great Gaelic genealogies, which lends to the belief that this surname was dependent upon a Norseman who settled in Ireland pre-Norman times. This belief is also supported by the presence of this surname being more frequent around areas near the sea coast [23]. Murphy is the most common surname in Ireland and is believed to have originated independently in multiple parts of Ireland, meaning it contains multiple founders, but the most powerful sept belonged to Leinster.

Historians believe there is a Norse influence on Murphy, due to its original name, Ó Murchadha, breaking down into the words “sea” and “battle/warrior” [23].

The most common subclade in both Doyle and Murphy is Z255 (60.00% and 26.32%). The second most common subclade for Doyle is DF41 (with two samples), with subclades P312, S288, L513 and Y-haplogroup I each containing one sample. Two individuals belonging to Z255 share an identical Y-STR profile. The second most common subclade for Murphy is L21 (21.05%), followed by M222 (15.79%), with subclades L11, FGC11134, DF21, Z3017, S588, Z251 and Y-haplogroup I each containing one sample. These results can be seen in Figures 19 and 20.

While Doyle appears to have multiple founders, there is a high proportion of individuals with Z255. All individuals who marked their provincial information and are classified as Z255 descended from Leinster. Some subclades seen in Doyle contain a lower resolution (P312) or are linked to British and Scottish proposed subclades (DF41 and L513), whose presence can be easily explained to this surname being frequent along the Irish Sea coast. The presence of the Y-haplogroup I supports the theory of Norse origin; however, with such a low frequency in combination with the high proportion of Z255, this may be another example of a case in which a Norseman ruled an area inhabited by indigenous Irish. Murphy appears to have multiple founders as theorized, with multiple major family lines being seen (both M222 and FGC11134). However, the pattern of subclade 255 with the presence of a Y-haplogroup I individual mirrors the patterns seen in Doyle.

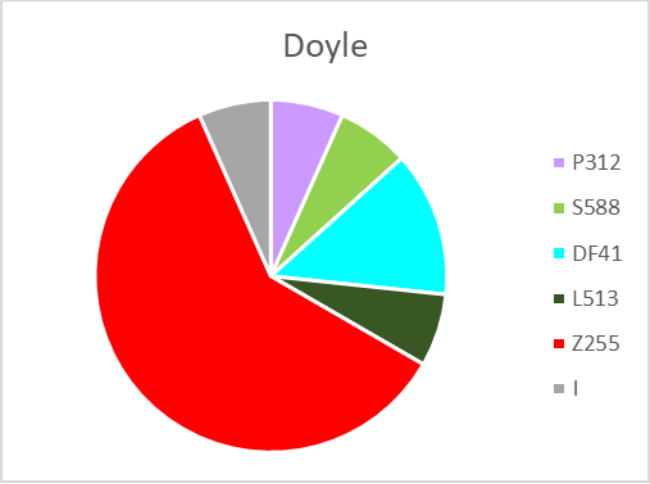


Figure 19: Y-haplogroup and R1b subclade frequency in Doyle (*n*=15)

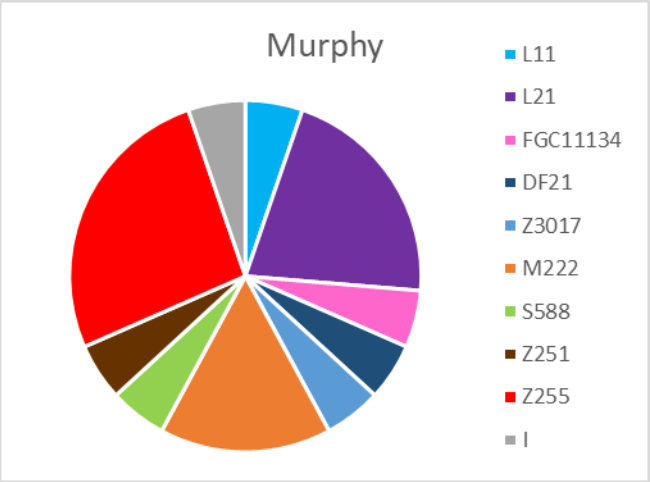


Figure 20: Y-haplogroup and R1b subclade frequency in Murphy (*n*=19)

3.12.9 Byrne and O'Toole

The Byrne and O'Toole families share a history. O'Byrne is from Ó Broin, which means descendant of Bran, King of Leinster (d. 1052) [23]. The O'Toole's, from Ó Tuathail (translated as "people mighty') were also a great Leinster sept [54]. Both families were pushed from their territory, which is now modern-day Co. Kildare, during the Anglo-Norman invasion. They settled south of Wicklow around the year 1200, where they put up a largely successful resistance to English aggression until the end of the 17th century, when Ireland relented to the British control [23].

The subclade frequency between these surnames is almost identical, which is to be expected when considering their shared history. The most frequent subclade in both Byrne (*n*=11) and O'Toole (*n*=12) is M222 at 36.4% and 33.3%, respectively. Following this, the subclades L21, DF21, Z2534 and Z255 at comparable frequencies in the two surnames. While one male in O'Toole is classified as Y-haplogroup I, in Byrne, the Y-haplogroups I and J contain one sample each. One sample in O'Toole could not be classified (this sample could not be genotyped with the R1b-L11 Multiplex assay and the Y-STR profile was inconclusive, preventing further analysis). The individuals in O'Toole belonging to M222 share an identical Y-STR profile. These results can be seen in Figures 21 and 22.

The subclades Z255 and DF21 are expected to be seen due to the location of the surnames and perhaps represent the original founders. Z255 has been referred to as the Irish Sea Haplotype by Joe Flood in unpublished work, and the genealogical community believed DF21 to be specific to the Irish and British populations. The coast on Leinster lies along the Irish Sea accounting for the subclade Z255. The location may also account for the presence of DF21, as the Irish Sea lies between Ireland and England. The presence of major historical familied can be seen in M222 (linked to the Uí Néill's [9, 50]) and Z2534 (linked to the Dalcassians [50]). As the surname Byrne is of King Bran of Leinster, the additional Y-haplogroups present do not seem to be representative of a Norse origin. It is likely that the additional Y-haplogroups arose from non-paternity or a Viking joining the ranks of the family.

The surname O'Toole is not linked to a particular king or historical Irish figure. With the surname only meaning "descendants of the mighty people," the presence of the Y-haplogroup I cannot be easily determined. Perhaps this family is a faction of the Byrne's, further explaining the similar frequency patterns. The surname may even contain a Norse origin, where the "people mighty" refers to the Vikings, and over the years the presence of the Y-haplogroup diminished over time while mixing with families such as the Byrne's.

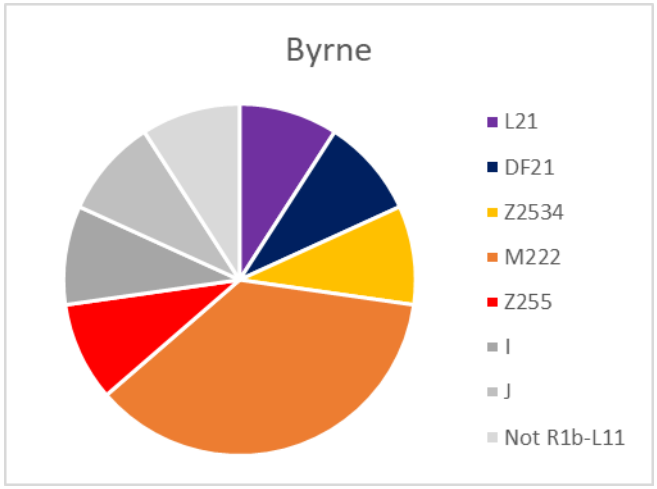


Figure 21: Y-haplogroup and R1b subclade frequency in Byrne (*n*=11)

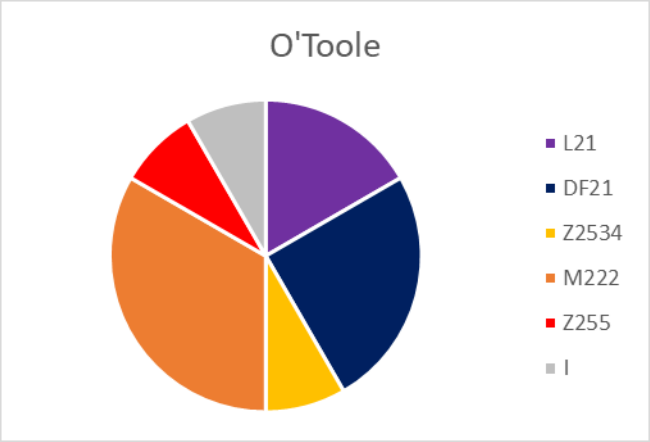


Figure 22: Y-haplogroup and R1b subclade frequency in O'Toole (*n*=12)

3.12.10 McCarthy

McCarthy is one of the most common surnames in Ireland and is the chief family of the Eoganachta [23]. The surname McCarthy is derived from Cárthach, lord of the Eóghannacht. As the Eoganachta kingdom has been linked with FGC11134 by Dr. Ken Nordtvedt, this was expected to be the most prominent subclade seen within the surname. However, of the samples collected for McCarthy (*n*=21) both FGC11134 and DF21 were tied for the most frequent subclades at 23.8% (five samples each). Following this is Z2534 at 19%. Two samples are Z255 and another two belong to the Y-haplogroup I. Subclades P312 and L21 each contain one sample, and one sample was classified as R1b (this sample was not able to be classified using the R1b-L11 Multiplex Assay, but the generated Y-STR profile was predicted as R1B by NEVGEN [66], meaning the sample belongs to a more ancestral subclade). Two individuals belonging to FGC11134 have an identical Y-STR profile. These results can be seen in Figure 23.

The subclade FGC11134 originates from the Eoganachta, but the distributions of the DF21, Z2534 and Z255 subclades with the Y-haplogroup I is similar to the pattern seen in the surnames Byrne and O'Toole. This pattern, along with the lower than expected frequency of FGC11134, suggests multiple founders for the surname McCarthy. Perhaps another sept originated within Leister in a similar location with the Byrne's and O'Toole's.

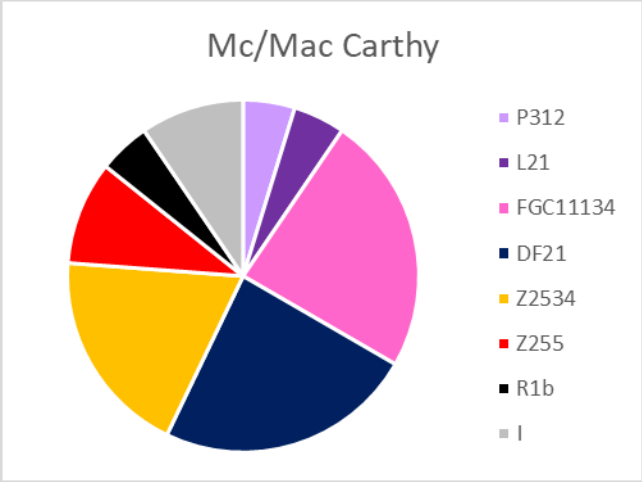


Figure 23: Y-haplogroup and R1b subclade frequency in McCarthy (*n*=21)

3.12.11 Gorey and Higgins

The surname Gorey is proposed to be of Norse origin and thought to be derived from the Norse personal name, Gofraidh [54]. The surname Higgins was anglicized from the name Ó hUiginn. The name hUiginn is believed to be a nickname meaning ‘viking’ or ‘sea-rover,’ allowing the original surname to be translated as “descendant of a viking” [23].

These two surnames show a similar pattern. The most frequent subclade in Gorey is FGC11134 (41.67%), followed by M222 (25.00%), then L21 (with two samples) and DF21 and Y-haplogroup I (with one sample each). The most frequent subclade in Higgins is M222 (61.90%), followed by DF21 (14.29%), then FGC11134 (with two samples) and DF5, Z2534 and Y-haplogroup I (with one sample each). Two individuals belonging to M222 share identical Y-STR profiles. These results can be seen in Figures 24 and 25.

These results contradict the belief that these surnames were founded based on Norse origin, as Gorey appears to have multiple founders from two major families, based upon the presence of M222 and FGC11134. Higgins appears to have multiple founders from three major families, based upon the appearance of M222, FGC11134 and Z2534. However, each surname contains one sample belonging to Y-haplotype I. This may again be indicative of a Norseman ruling an area inhabited by indigenous Irish, but this does not represent the only origin found in these names.

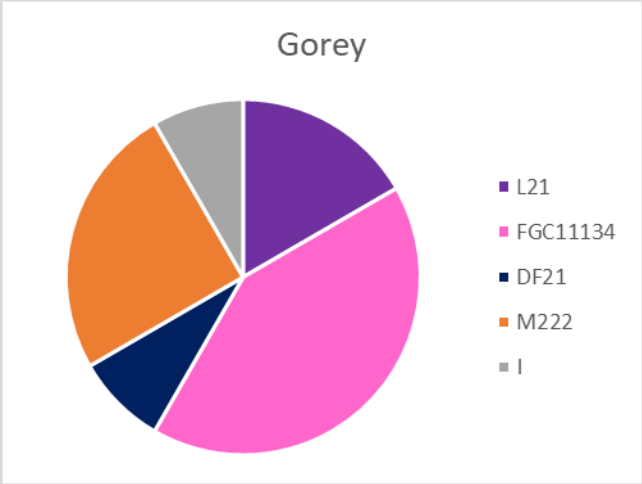


Figure 24: Y-haplogroup and R1b subclade frequency in Gorey (*n*=12)

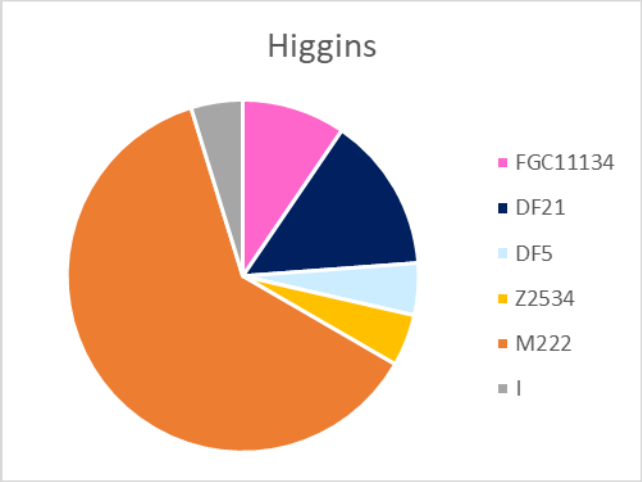


Figure 25: Y-haplogroup and R1b subclade frequency in Higgins (*n*=21)

3.12.12 McCauley

The surname McCauley is believed to have two possible origins. The first, a descendent of Amhlaidh, which is a Norse personal name. The second, an anglicized version of MacAuley, meaning “son of Auley” [54], who is a proposed descendent of Uí Néill. The most frequent subclade in McCauley is M222 (57.14%), followed by S588 (28.57%), with one sample belonging to L513. These results can be seen in Figure 26. As M222 a marker for the Uí Néill kingdom (with S588 being a subclade underneath M222), these results support the theory that the surname McCauly is derived from the name MacAuley and belongs to the Uí Néill kingdom.

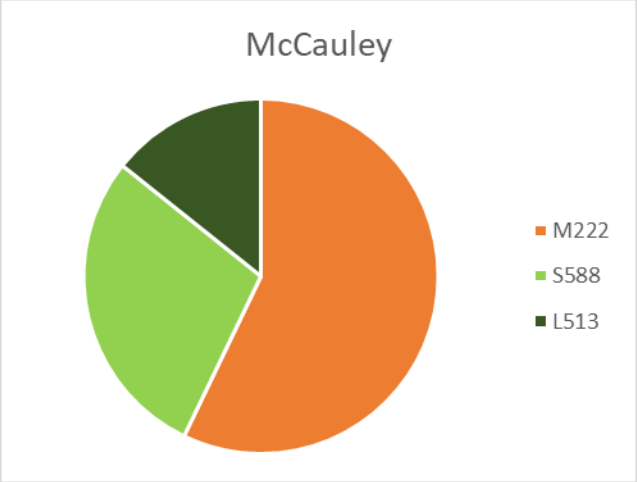


Figure 26: R1b subclade frequency in McCauley (*n*=7)

3.12.13 Butler

Butler is a relatively common name found all across Ireland. As there is no reliable pedigree for this name, its origin is not very well known. The surname and the profession linked to (butler or servant) it came to Ireland during the Anglo-Norman Invasion. Multiple individuals with different origins would have adopted this surname throughout Ireland, and so many different septs of Butler exist and are recorded [23]. There is a high frequency of the Y-haplogroup I in Butler (35.3%) and lower resolved clusters (P312, Z195 and L21 with a combined frequency of 28.6%). Subclades DF21 and M222 each contain two samples. Subclades DF5, DF49 and L513 each contain one sample. These results can be seen in Figure 27. Butler is shown to have a wide range of Y-haplogroups and subclades suggesting multiple founders of Irish, Norse and English origin. Consequently, Butler cannot be claimed as solely having an Irish or English origin.

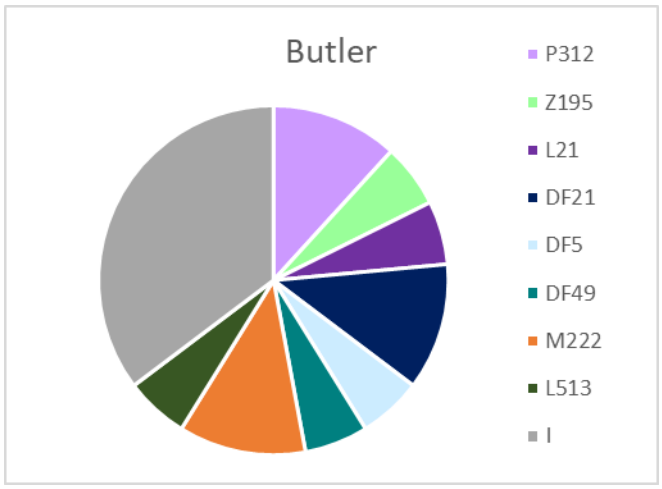


Figure 27: Y-haplogroup and R1b subclade frequency in Butler (*n*=17)

3.12.14 Hanly

Hanly is a proposed Irish surname from the anglicized form of Ó hAinle, a name believed to come from the Gaelic word áluinn, meaning beautiful [23]. There is a high proportion of samples belonging to Y-haplogroup I (25%), suggesting a high amount of Norse influence on this name with proposed Irish ancestry. Interestingly, though the name Hanly is not linked to the Dalcassian family, it shows a similar pattern to the surnames Boland and Kennedy (a large amount of Z2534 in combination with L513). The subclade with the highest frequency is L513 (41.7%), and in third is Z2534 (16.7%), behind Y-haplogroup I. These results can be seen in Figure 28. Hanly appears to have multiple founders, perhaps of both Norse and Irish origin.

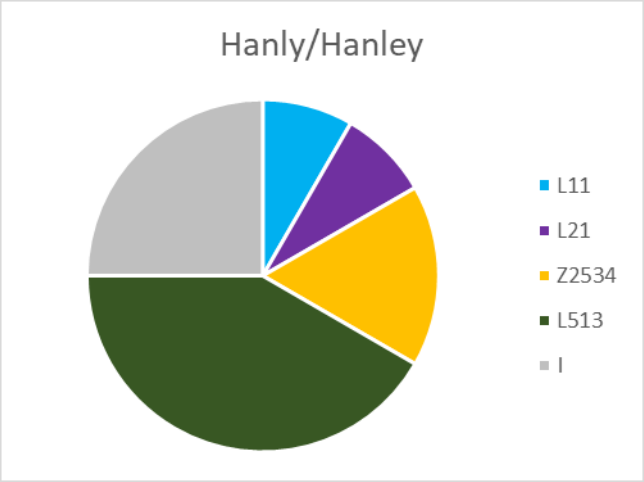


Figure 28: Y-haplogroup and R1b subclade frequency in Hanly (*n*=12)

3.12.15 Johnston and Thompson

Johnston and Thompson are bother considered to be of Scottish origin [54]. Both surnames contain a large proportion of Y-haplogroups suggestive of a Norse influence. A high frequency of samples in Johnston belong to Y-haplogroup I (41.7%) and one sample belongs to R1a. The second most frequent subclade is P312 with three sample, and subclades L11, DF49 and M222 each contain one sample. These results can be seen in Figure 29.

While the most frequent subclade in Thompson is L21 (35.7%), one sample belongs to Y-haplogroup I and two samples belong to R1a. Subclade P312 contains two samples and subclades M222 and Z255 each have one sample. One sample also belongs to Y-haplogroup T and one sample could not be classified (this sample could not be genotyped with the R1b-L11 Multiplex assay and the Y-STR profile was inconclusive, preventing further analysis). These results can be seen in Figure 30.

While the subclades L11, P312, L21 and DF41 are suggestive of an English (Scottish) origin, the high proportion of Y-haplogroups I and R1a propose a heavy presence of Viking influence in these surnames as well.

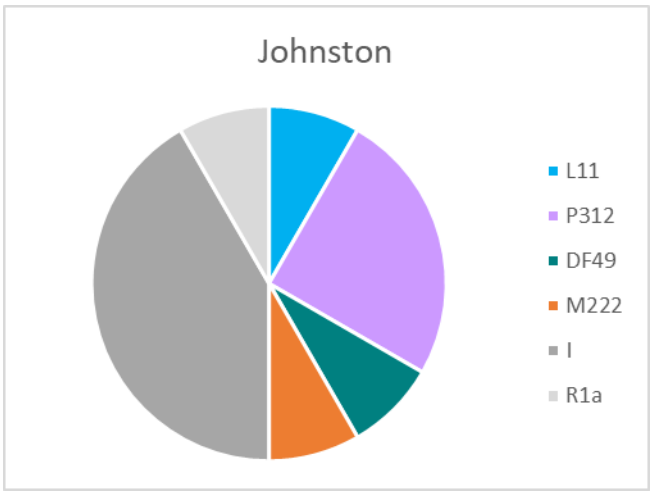


Figure 29: Y-haplogroup and R1b subclade frequency in Johnston (*n*=12)

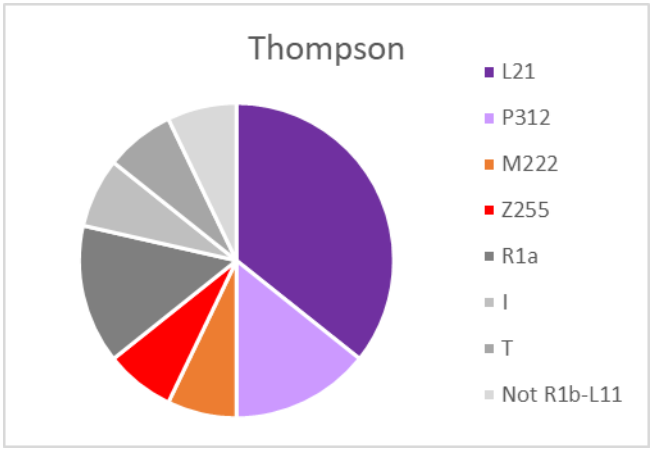


Figure 30: Y-haplogroup and R1b subclade frequency in Thompson (*n*=14)

3.12.16 Kettle, McSorley and Toner

The surnames Kettle, McSorley and Toner all come from Norse personal names. Kettle was originally Mac Coitil, McSorley was MacSomhairle and Toner was Ó Tomhrair [54]. All samples in both Kettle and McSorley belong to S588. The most frequent subclade in Toner is S588 (80.00%), with one sample belonging to FGC11134. These results can be seen in Figures 31-33. Though these surnames are comprised of a small sample sizes (Kettle *n*=3, McSorley *n*=4, Toner *n*=5), S588 is highly specific to the Irish population, as it is a subclade underneath M222, the most prominent subclade in Ireland. Therefore, these surnames cannot be representatives of surnames with a Norse origin.

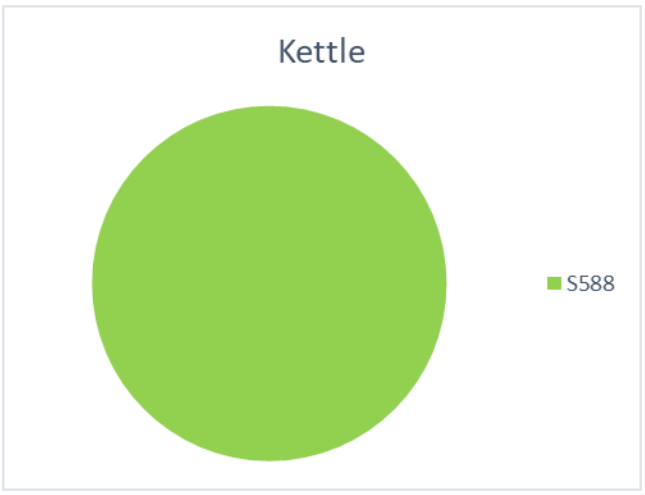


Figure 31: R1b subclade frequency in Kettle (*n*=3)

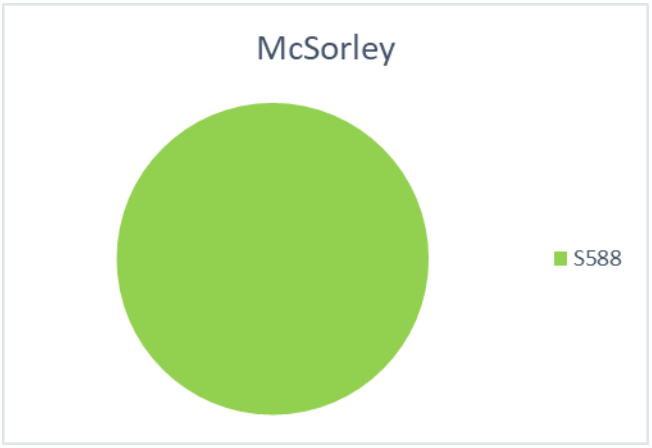


Figure 32: R1b subclade frequency in McSorley (*n*=4)

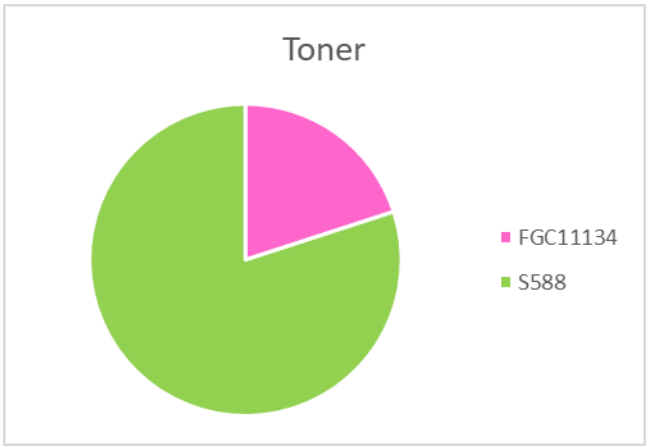


Figure 33: R1b subclade frequency in Toner (*n*=5)

3.13 The Case of Subclade R1b-S588

Many surnames with a proposed Norse origin contain samples belonging to the subclade R1b-S588. Though S588 is a highly Irish specific subclade, it is only found in 4 out of the 293 genotyped samples of proposed Irish origin (1.4%) while being present in 27 out of the 268 haplotypes samples of proposed Norse origin (10.1%). Only two surnames belonging to the proposed Irish origin group contain S588 (Carr and Gilmore) while ten surnames belonging to the proposed Norse origin group contain S588 (Doyle, Grimes, Hagan, Kettle, McCauley, McLoughlin, McSorley, Murphy, Somers and Toner). This high proportion of proposed Norse origin surnames containing this high specific Irish subclade is not expected and has not been noted in previous studies. Perhaps this is one of many reasons why Y-STR profiles are not an efficient way of measuring Norse origin in Irish surnames, and occurrences such as this go unchecked. Interestingly, though Carr and Gilmore are considered to be Irish surnames, arguments could be made to connect these surnames to Norse origins. The surname Gilmore could be contracted from the surname Gill, which is Old Norse for a small ravine [89]. Carr may be linked to the Old Norse word for marsh (*kjarr*) [90]. If these surnames were link to a Norse origin, this would mean that only surnames with a proposed Norse origin contained the Irish specific subclade S588.

This odd case could be caused by a newly formed sept of Irish individuals choosing to take surnames with Norse roots. S588 is a subclade underneath M222, but this does not mean that members of S588 were closely related to the influential Uí Néill kingdom. Due to the slow mutation rate of SNPs, it is likely that the rise of S588 did not occur until a member of the M222 subclade greatly distanced himself from the Uí Néill's. Members of S588 may have been without a prominent clan, and more likely to change their surnames to be able to join a new clan lead by a Norseman. They may also have been a group of lower-class citizens who were unlikely to have a surname until forced to adopt on after the Anglo-Norman invasion. Having a Norse surname was almost seen as a trend at this time, and these names may have been chosen at random. The relationship between these surnames and S588 should be studied further to better understand the genetics of Ireland.

3.14 Unknown Person Identification Case

As a part of principle, the case of JB-55 is an example of identifying an unknown person using Y-DNA in combination with the methods described above. A male body, identified as JB-55 by his gravestone, was found in Jewett City, Connecticut, believing to have died between 1810-1830. Two extracts of DNA from the body (JB-55-1 and JB-55-2) and two reagent blanks (JB-55-RB1 and JB-55-RB2) were sent to be tested with the R1b-L11 Multiplex assay to determine the likelihood of Irish ancestry and a potential surname.

Samples JB-55-1, JB-55-RB1, JB-55-2 and JB-55-RB2 were ran through the R1b-L11 Multiplex assay along with two additional negatives to be sure no contamination was present. Both JB-55-1 and JB-55-2 were genotyped as R1b-P312. JB-55-RB1, JB-55-RB2 and neither of the additional blanks produce any results. The subclade R1b-P312 is prevalent among Western and Northern Europe [91] and cannot be specified to a specific country origin. While P312 is found at low levels in our Irish population dataset, it is more prevalent in individuals with a proposed British surname (proposed Irish at 4.52%, proposed Norse at 3.73% and proposed British at 16.13%). A man of Western European descent may share one of these more ancient subclades with the Irish population (L11, P312, Z195 and L21), and so it is theorized that JB-55 may be of English descent.

The provided Y-STR profile was entered into NEVGEN [66] to see if a more specific Y-haplogroup subclade could be discovered. The most probable prediction given was R1b-A6487 (46.9%), which is a further subclade of R1b-DF21. As R1b-DF21 is included in the R1b-L11 Multiplex Assay and the sample did not present this mutation, this prediction was disregarded. The next highest prediction was R1b-A6292 (15.9%), a subclade under R1b-L238 which is underneath R1b-P312. As this is congruent with the findings, this result was further investigated.

The provided Y-STR profile was used to search the FamilyTreeDNA database to see if any individuals in the database present a matching profile. No individuals grouped into the subclade R1b-L238 presented a similar Y-STR profile.

The search was then broadened to individuals grouped within the subclade confirmed in the assay, R1b-P312, which still narrowed the criteria down from Y-haplogroup R1b. Using this group, two individuals (who shared a matching Y-STR profile) were found to be 1-step-neighbors with JB-55, meaning they shared an almost identical Y-STR profile except at one loci (Y-GATA-H4), where they differed by only 1 repeat (JB-55 has 12 repeats at this location, where the individuals in the database have 11). These individuals are labeled under the surname Barber, which is

indicative of the JB initials found on the gravestone. In a study analyzing the mutation rates for the Promega PowerPlex® Y23 Y-STR loci kit in Serbian father-son pairs [92], a one-step mutation at Y-GATA-H4 was observed once out of 269 father-son pairs. Using these findings, the mutation rate of Y-GATA-H4 was calculated as having a median mutation rate of 0.62×10^{-2} using the Bayesian approach.

The paternity dispute of President Thomas Jefferson can be used as an example to demonstrate the identification of male relatives using Y-STRs [93]. Living male relatives of Thomas Jefferson were compared to the living decedent of Eston Hemings Jefferson, who was the son of Sally Hemings (the African American slave of Thomas Jefferson). The male relatives of Thomas Jefferson and the decedent of Eston Hemings Jefferson were found to have matching Y-STR profiles, except for one repeat difference at a single Y-STR loci, which was explained as a mutation. Eston Hemmings Jefferson was then determined to be fathered by Thomas Jefferson, or his brother. It is therefore plausible that JB-55 is a close relative to the Barber family.

After discovering these two individuals, a search was done on samples containing the surname Barber. Six individuals with this surname contained identical profiles to the two samples mentioned above. Given the information on these samples, the Barber family appears to be from England and mainly inhabited the New England coast. This information was given to the investigating laboratory, and further assessment of Barber as a potential identity will be conducted by them.

JB-55 was buried alongside a female adult named “IB-48” and child “NB-13” in a position that suggested these individuals were a nuclear family. Upon further investigation into historical records, a researcher discovered a potential match to the JB-55 sample. In the Hale Index of Connecticut Burials there is mention of a John Barber whose son, Nathan Barber, died in Griswold, CT in 1826. There is a small discrepancy as Nathan Barber is written to have died at age 12, though the grave indicates an age of 13. Further research about John and Nathan Barber is currently continuing in an attempt to better understand the connection between the JB-55 sample and these individuals.

CHAPTER 4. CONCLUSION

Ireland is a unique and interesting country, not only because of its geographical location and history but now because of its genetics. A complicated history of warriors and high kings placed a large scale of importance on genealogies and family history at an early age. Detailed records of families and their descendants were kept, as well as documentation on events such as the Viking raids and the Anglo-Norman Invasion. The history of Ireland can be studied using the Y-chromosome, as it is passed on from father to son, relatively unchanged, just as surnames are in this patrilineal society.

Ireland is not the only country influenced by the Vikings. Countries along the coasts of the Irish Sea fell victim to raids as well. An extensive study on the Viking influence in the English populations focus on the use of Y-STR profiles and compare groups believed to be influence by the Vikings to the Y-STR profiles of the modern day Norse [3]. While this method was considered successful for the English population, England does not contain a surname history as old or extensive as Ireland [23]. Another study attempted to examine the influence of Vikings on Ireland using a similar approach and concluded that while there is a small amount influence found, the presence of Vikings was much smaller than initially thought. This however was shown to be false when an admixture analysis on Ireland found there to be an upper estimate of 20% Norwegian ancestry within the Irish population [53].

Y-STRs are beneficial in forensic work, particularly in testing for paternity or identifying paternal relatives in a family. Compared to Y-SNPs, Y-STRs have a faster mutation rate and are better suited for studies regarding a more recent history. They are less practical in ancient genealogical studies such as comparing modern day males to events that occurred in the 8^{th} century. The main application of Y-STRs in this study was to check for related males containing the same surname to ensure that the results were not being skewed or overstated. An MDS plot was also made to show that, using Y-STRs, it is difficult to accurately determine the genetic origin of a surname. Using this method, there was shown to be no difference between the proposed Irish surname group and the proposed Norse surname group, though the proposed English surname group was distinct from the rest of the Irish population. However, when viewing Y-SNPs and subclades the

differences between the two groups is more obvious, especially when considering haplogroup frequencies within that list of individuals with the same surname.

Overall, analyzing surname origin groups by Y-haplogroup and subclade frequency provides a better understanding of their genetic differences in a way that can be visualized. Surnames of proposed Norse and English origin are more likely to belong to lower resolved subclades (L11, P312, Z195 and L21) that are less specific to the Irish population. P312 is more specific to the proposed English origin surnames, but the other three subclades are seen at an equal frequency between the two. Subclades below L21 are more likely to be of Irish origin. This was also observed when comparing the collected Irish data to the 1000 Genomes data.

Historical Irish families and kingdoms can be traced using such population specific SNPs. The data from this study supports the finding that M222 is linked to the Uí Néill kingdom [50], demonstrated by its presence in surnames historically linked to the Uí Néill kingdom. The unpublished work by Dr. Ken Nordtvedt, suggesting that FGC11134 is linked to the Eoganachta kingdom is also supported by the findings in the surnames O'Donovan and O'Sullivan. These surnames contain detailed, historical records linking them to the Eoganachta kingdom and both contain a high frequency of FGC11134. The Dalcassian tribe has been linked to Z2534 [50], though this link is not seen to be as strong as M222 is for the Uí Néill kingdom or FGC11134 is for the Eoganachta kingdom. While surnames such as Boland and Kennedy, both linked to the Dalcassians, contain a large proportion of Z2534, they also contain other subclades, such as L513 and Z255. The Dalcassian dynasty was smaller than both the Uí Néill and Eoganachta dynasties. They might have needed to recruit men into their army and required them to change their surname to show allegiance.

When viewing surnames independently, rather than as groups of proposed origins, it is easier to examine the genetic origin of each individual surname. The surname history in Ireland is complex. Adoption and nonpaternity are not the only ways a surname would not match its proposed origin. Maternal surnames may have been passed on if the mother was Norse and the father was Irish [25]. Historical surname documents were often altered, and as such not all records may truthfully indicated a names origin [9]. When those who had not already taken a surname were forced to after the Anglo-Norman invasion, some may have chosen a Norse surname despite being of Irish descent. Irish individuals living under Norse rule may have been required to change their surname to show their loyalty. All these instances are an example of why the Irish population cannot be

easily and cleanly grouped into proposed surname origins and why Y-SNPs should be observed in addition to history. This research has demonstrated when an Irish name is truly of Irish descent and if a Norse surname accurately potentially contains Norse roots. It can also provide information on a surname going against its historically proposed origin.

An example of this is the surname McCauley. McCauley is believed to be of Norse origin, but a large proportion of samples belong to the subclade M222, suggesting an Irish origin. Another example pertains to the surnames Kettle, McSorley and Toner. All three of these surnames are proposed to be of Norse origin, yet most samples from these surnames belong to subclade S588, a distinctly Irish subclade. The relationship between some Norse surnames and S588 might not have been discovered if this study simply relied on Y-STRs.

In addition, a low rate of immigration between the provinces was also seen using Y-haplogroups and subclades, as modern DNA samples still reflect the locations of historical kingdoms. The distribution of subclades in these provinces supports the links found between Irish type clusters and Y-SNPs. A high proportion of M222 and S588 was found in Ulster and Connacht. M222 has been linked to the North West Irish Type I group [50]. Subclade FGC11134, which has been linked to the South Irish Type II group by Dr. Ken Nordtvedt and is the most common subclade found in Munster with very low frequencies elsewhere. Subclade Z2534, which has been linked to the South West Ireland [50], does not show a distinct pattern regarding geographical location. Whilst this subclade seems to show a pattern linking it to the Dalcassian clan, further analysis should be done to test the true location of Irish Type III, while examining historical documentation may help to understand the proper link between Z2534 and the Dalcassian families.

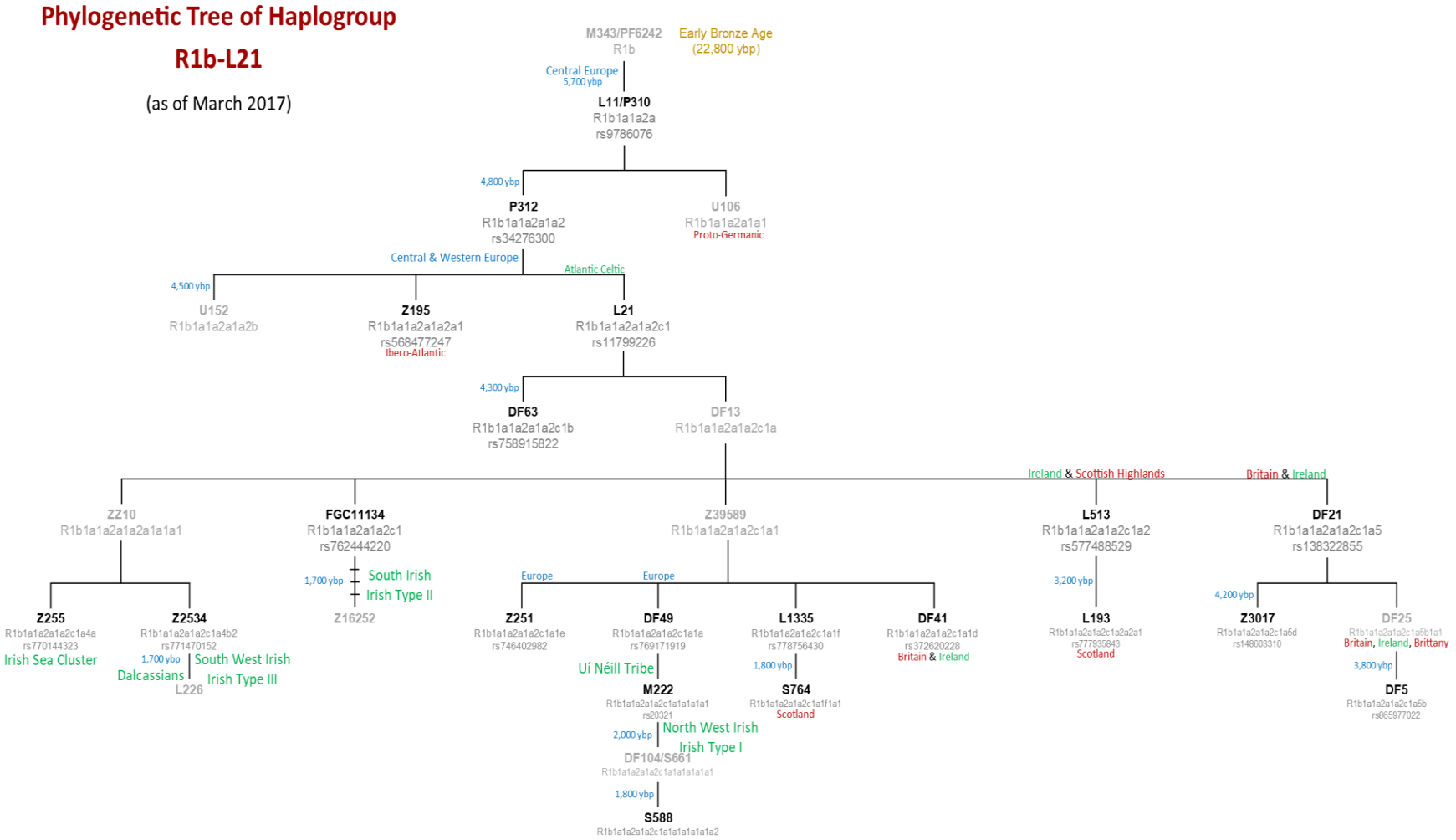
A better understanding of the relationship between Ireland and the Vikings, with regards surnames has been provided with this research. Several surnames were linked with strong support to their true Norse origin. The distribution of the Y-haplogroup I, R1a and other subclades that may have a link to the Viking population (such as R1b-U106 under L11 and other subclades under L21) is also visible. The largest proportion of Y-haplogroup I is in Connacht, which is where the first Viking raids took place. This is followed by both Munster and Leinster, where the Vikings made settlements that became major Irish cities. These provinces also contain the highest proportions of L11 (in Munster and Leinster only) and L21. R1a is only present in Leinster. These results show a high impact of Viking settlement in these areas.

The subclades chosen for the R1b-L11 Multiplex Assay have been shown to be highly effective in determining Irish ancestry, both in comparing the collected Irish data to the 1000 genomes data, and in comparing the Irish collected data to itself in the proposed surname origin groups. The subclades were first able to determine if a sample is of western European ancestry with the low-resolved subclades (L11, P312, Z195 and L21). The comparison with the 1000 Genomes data showed that populations outside of Western Europe were highly unlikely to belong to one of the chosen subclades included in the assay, or even to simply R1b. The 1000 Genomes data also demonstrated that males from the Great British population were more likely to belong to the lower-resolved subclades or subclades theorized to belong to English and Scottish populations (DF63, L513 and DF5). The R1b-L11 Multiplex Assay could therefore be used to help infer if an undetermined male sample is potentially of Irish origin.

This application was demonstrated with the analysis of the JB-55 unknown persons sample. Using only the Y-STR profile an incorrect subclade was predicted. The predicted subclade was R1b-DF21, which is included in the R1b-L11 Multiplex Assay and is believed to be representative of both Irish and British populations. However, when the sample was ran using the R1b-L11 Multiplex Assay, JB-55 did not contain the mutation for DF21 and could only be classified with the more ancient subclade, P312. As P312 is more frequent in surnames with a proposed English origin in the collected data, this sample was reported back as improbable for Irish ancestry. Overall, this research could be particularly helpful in the examination of potentially ancient Irish remains to help piece together the story of the past, in examining potential provincial locations to possible surname links.

APPENDIX A

Phylogenetic tree of Y-haplogroup R1b, focusing on R1b-L21. Subclades in black represent the subclades used in the R1b-L11 Multiplex Assay. Subclades in gray were included to show both structure and major subclades non-specific to the Irish population.



APPENDIX B

A list of surnames collected during sampling.

Surname	Origin	Spelling Variants Included in Samples	# of Samples	# of Families	Frequency
Arthur	Norse	N/A	17	8	49
Boland	Irish	N/A	8	7	34
Broderick	Norse	N/A	26	9	37
Butler	Irish	N/A	17	17	20
Byrne	Irish	N/A	11	10	3
Carr	Irish	N/A	4	4	30
Coll	Norse	N/A	5	5	46
Coppinger	Norse	N/A	15	7	59
Costin	Norse	N/A	2	2	64
Cotter	Norse	N/A	10	7	32
Curley	Irish	N/A	4	4	39
Curran	Irish	N/A	18	14	18
Dolphin	Norse	N/A	4	4	61
Doyle	Norse	N/A	15	10	6
Duggan	Irish	N/A	10	10	24
Fitzpatrick	Irish	N/A	7	7	14
Gilmore	Irish	N/A	16	15	35
Goodman	English	N/A	7	3	57
Gorey	Norse	Gorry	12	6	54
Grimes	Norse	N/A	10	9	29
Guy	English	N/A	1	1	55
Hagan	Norse	O'Hagan	9	7	27
Hamond	Norse	Hammond	10	6	52
Hanly	Irish	Hanley	13	9	31
Harold	Norse	Harrold	8	6	53
Harold-Barry	Norse-Irish	Harrold-Barry	3	2	65
Higgins	Norse	N/A	21	16	17
Holden	English	N/A	4	3	48
Johnston	English	N/A	12	9	9
Kennedy	Irish	N/A	15	9	7
Kettle	Norse	N/A	4	2	60
Kilpatrick	Irish	N/A	2	1	47
Lamb	English	N/A	9	6	42
Loughlin	Irish	N/A	3	3	28
Loughman	Irish	N/A	10	8	58

(continued)

McArdle	Irish	N/A	17	13	33
McAuliffe	Norse	N/A	8	7	40
McCabe	Irish	N/A	14	9	16
McCarthy	Irish	MacCarthy	21	16	4
McCauley	Norse	N/A	8	5	36
McDowell	Norse	N/A	10	9	26
McGettigan	Irish	N/A	7	4	51
McGinty	Irish	N/A	5	4	41
McGorry	Norse	N/A	2	1	62
McKeever	Irish	N/A	7	6	43
McLoughlin	Norse	McLaughlin	18	16	8
McMahon	Irish	N/A	6	6	15
McManus	Irish	MacManus	9	9	22
McNea	Irish	N/A	2	1	63
McSorley	Norse	N/A	4	2	50
McSweeney	Irish	MacSweeney, McSwiney	6	6	11
Murphy	Norse	N/A	22	19	1
O'Donovan	Irish	N/A	6	6	12
O'Loughlin	Irish	N/A	7	6	44
O'Mahony	Irish	N/A	5	4	13
O'Rourke	Norse	N/A	6	6	21
O'Toole	Irish	N/A	12	11	23
Reilly	Irish	O'Reilly	24	13	5
Reynolds	Norse	N/A	15	14	25
Somers	Norse	N/A	5	4	45
Sullivan	Irish	O' Sullivan	14	9	2
Taylor	English	N/A	1	1	19
Thompson	English	N/A	14	13	10
Toner	Norse	N/A	6	3	38
Turley	Norse	N/A	7	4	56

APPENDIX C

The Excel sheet for the R1b-L11 Multiplex Assay.

No.	SNP	Primer set (all 50um stock)	PCR Ps d STOCK(50um)	10	Dir	SNPs	Conc p	20um(ul)	of 50um(ul)	10
1	L21	F&R	0.1 each F&R =0.4um	1	L21	C/G	1.30 µm		0.13	**1.30**
2	M222	F&R	0.1 each F&R =0.4um	1	M222	G/A	0.04 µm	0.01		**0.10**
3	L193	F&R	0.2 each F&R =0.4um	2	L193	A/G	2.30 µm		0.2	**2.00**
4	DF49	F&R	0.2 each F&R =0.4um	2	DF49-OLD	A/G	0.30 µm		0.23	**2.30**
5	Z2534	F&R	0.1 each F&R =0.4um	1	Z2534	C/T	0.50 µm		0.03	**0.30**
6	L513	F&R	0.1 each F&R =0.4um	1	L513	C/T	2.00 µm		0.05	**0.50**
7	Z251	F&R	0.1 each F&R =0.4um	1	Z251	A/G	0.50 µm		0.05	**0.50**
8	DF41	F&R	0.2 each F&R =0.4um	2	DF41	T/C	2.50 µm		0.25	**2.50**
9	P312	F&R	0.1 each F&R =0.4um	1	P312	A/C	1.00 µm		0.1	**1.00**
10	L1335	F&R	0.1 each F&R =0.4um	1	L1335	T/C	0.60 µm		0.06	**0.60**
11	Z255	F&R	0.1 each F&R =0.4um	1	Z255	G/A	0.20 µm		0.02	**0.20**
12	L11	F&R	0.1 each F&R =0.4um	1	L11	C/T	0.80 µm		0.08	**0.80**
13	DF21	F&R	0.1 each F&R =0.4um	1	DF21	G/A	0.30 µm		0.07	**0.70**
14	DF63	F&R	0.1 each F&R =0.4um	1	DF63	C/T	0.70 µm		0.03	**0.30**
15	FGC11134	F&R	0.1 each F&R =0.4um	1	FGC11134	A/T	0.04 µm	0.01		**0.10**
16	Z195	F&R	0.1 each F&R =0.4um	1	Z195	A/G	0.30 µm		0.03	**0.30**
17	S764	F&R	0.1 each F&R =0.4um	1	S764	G/A	0.80 µm		0.08	**0.80**
18	DF5	F&R	0.1 each F&R =0.4um	1	DF5	A/G	0.80 µm		0.08	**0.80**
19	S588	F&R	0.1 each F&R =0.4um	1	S588	C/T	2.00 µm		0.2	**2.00**
20	Z3017	F&R	0.1 each F&R =0.4um	1	Z3017	C/T	0.80 µm		0.08	**0.80**
		Total primers (for both F&R)	**46**	**23**				**total**	**1.79**	17.90

				no.samples				no. of samples		
Stocks		**PCR 1st set up (ul)**	Final conc	**10**		**SBE reaction**	**(ul)**	**10**		
primers		**4.6**		46		primers	**1.79**	17.9		
10x PCR gold buffer (no mg)		**1**	1X	10		snapshot rxn mix	**1**	10		
Mgcl2 (25mM)		**1**	2.5mM	10		H20	**0.21**	2.1		
dNtps (10mM each,combo 40mM)		**0.22**	220 uM	2.2		1st PCR product	**2**			
taq gold (5U/ul)		**0.3**	1.75U	3			5ul/rxn	30	total	
H2O		**1.88**		18.8		**Thermo conditions: SBE rxn**				
Total reagents		**9**		90		96 °C for 2 min and 25 cycles of 96 °C for 10 s, 50 °C for 5 s and 60 °C for 30 s				
DNA		**1**								
		10ul total				**Clean up after SBE**				
Thermo conditions: 1st PCR						Sap(1U/ul)		(1)37°C - 45min, (2) 75°C - 15min		
(1) 95 °C for 10 min, (2) 33 cycles of 95 °C for 30 s and 61 °C for 30 s, (3) 5 min at 61 °C.										
						CE run				
Clean up after 1st PCR						POP-7 on a 36 cm capillary length array.				
Exo Pro Star	2ul used per 5ul product of each sample as per des		(1)37°C - 15min, (2) 80°C - 15min			Run parameters: injection voltage of 2.5 kV for 10 s, and run time of 500 s at 60				

APPENDIX D

Allele frequency distribution of 23 Y-STR markers in the Irish male population

Alleles	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439
8										
9									0.007	0.002
10					0.362				0.116	0.024
11					0.61	0.082			0.017	0.261
12		0.071			0.026	0.056	0.031		0.838	0.51
13	0.029	0.768			0.002	0.609	0.898		0.021	0.164
14	0.829	0.158				0.249	0.05	0.135		0.036
15	0.135	0.002				0.005	0.019	0.803		0.002
16	0.005						0.002	0.062		
17	0.002									
18										
19										
20				0.002						
21				0.005						
22				0.031						
23				0.117						
24				0.53						
25				0.291						
26				0.021						
27			0.002	0.002						
28			0.1							
29			0.602							
30			0.219							
31			0.05							
32			0.024							
33			0.002							

(continued)

Alleles	DYS448	DYS456	DYS458	DYS481	DYS533	DYS549	DYS570	DYS576	DYS635	DYS643	GATA-H4
8										0.007	
9						0.005				0.039	
10										0.812	0.015
11					0.067	0.059				0.037	0.169
12					0.479	0.383				0.088	0.763
13		0.002			0.393	0.409				0.015	0.054
14		0.056	0.007		0.055	0.121	0.002			0.002	
15		0.286	0.05		0.005	0.024	0.012	0.005			
16		0.313	0.164		0.002		0.092	0.057			
17	0.002	0.299	0.558				0.537	0.286			
18	0.312	0.044	0.19	0.005			0.208	0.355			
19	0.55		0.024				0.113	0.253	0.014		
20	0.126		0.007	0.007			0.019	0.038	0.002		
21	0.007			0.057			0.012	0.005	0.054		
22				0.476			0.005		0.031		
23				0.113				0.002	0.719		
24	0.002			0.057					0.158		
25				0.233					0.014		
26				0.035					0.007		
27				0.017							
28											
29											
30											
31											
32											
33											

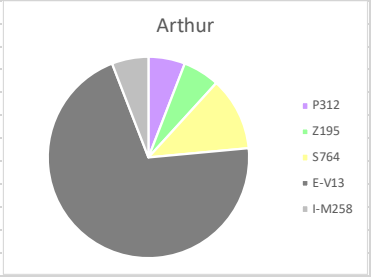
(continued)

Allele	DYS385
9,11	0.002
9,14	0.002
10,11	0.007
10,14	0.005
11,11	0.009
11,12	0.021
11,13	0.232
11,14	0.407
11,15	0.144
11,16	0.019
12,14	0.009
12,15	0.038
12,16	0.028
13,16	0.019
13,17	0.002
14,17	0.012
15,17	0.002
15,18	0.017
16,16	0.002
16,18	0.002
16,19	0.005
17,17	0.007
17,18	0.002
18,18	0.005

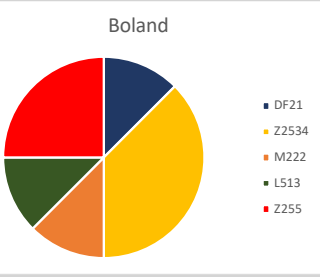
APPENDIX E

Y-Haplogroup and R1b subclade frequency of the Irish population by surname

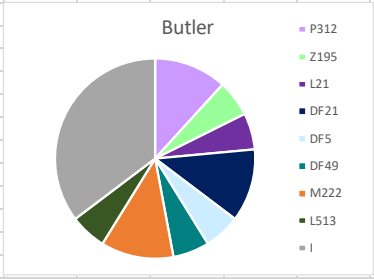
Arthur	
P312	1
Z195	1
S764	2
E-V13	12
I-M258	1



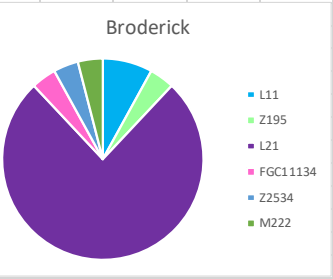
Boland	
DF21	1
Z2534	3
M222	1
L513	1
Z255	2



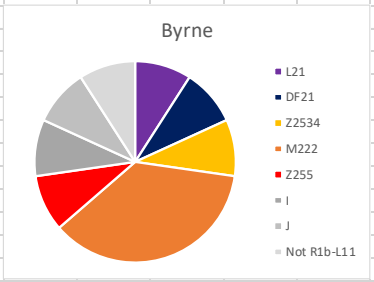
Butler	
P312	2
Z195	1
L21	1
DF21	2
DF5	1
DF49	1
M222	2
L513	1
I	6



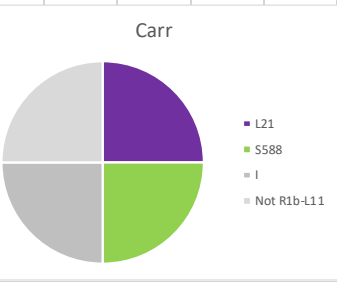
Broderick	
L11	2
Z195	1
L21	19
FGC11134	1
Z2534	1
M222	1



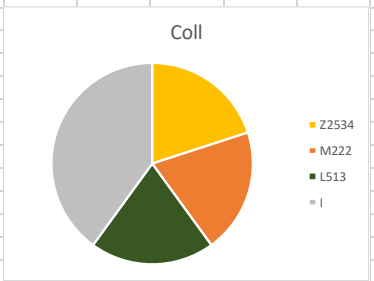
Byrne	
L21	1
DF21	1
Z2534	1
M222	4
Z255	1
I	1
J	1
Not R1b-L	1



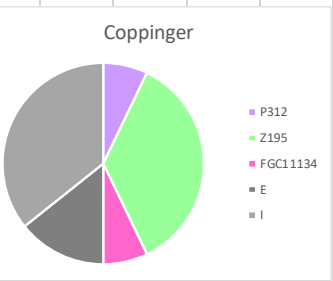
Carr	
L21	1
S588	1
I	1
Not R1b-L	1



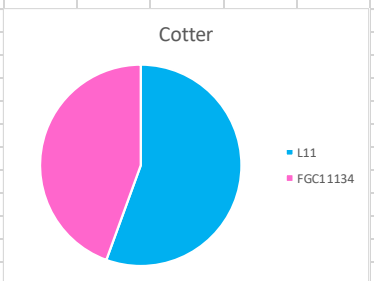
Coll	
Z2534	1
M222	1
L513	1
I	2



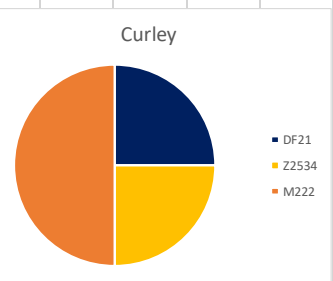
Coppinger	
P312	1
Z195	5
FGC11134	1
E	2
I	5



Cotter	
L11	5
FGC11134	4

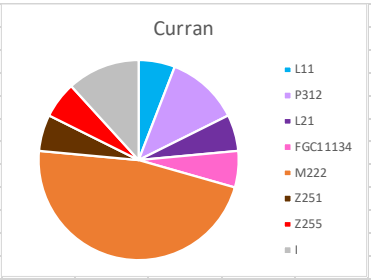


Curley	
DF21	1
Z2534	1
M222	2

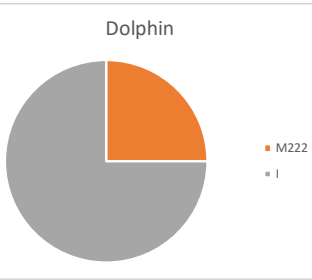


(continued)

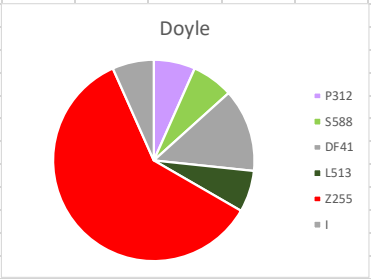
Curran	
L11	1
P312	2
L21	1
FGC11134	1
M222	8
Z251	1
Z255	1
I	2



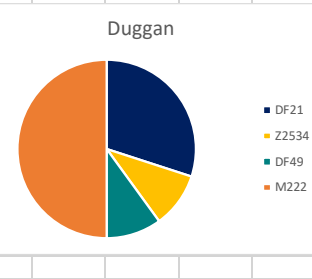
M222	1
I	3



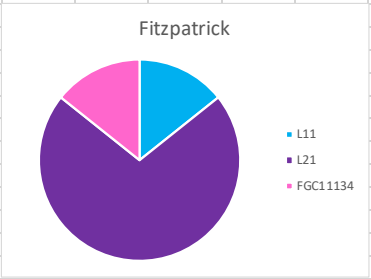
Doyle	
P312	1
S588	1
DF41	2
L513	1
Z255	9
I	1



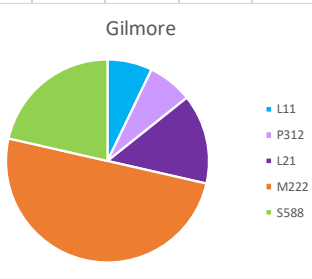
Duggan	
DF21	3
Z2534	1
DF49	1
M222	5



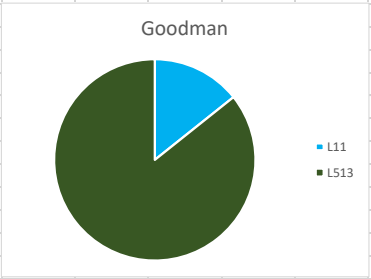
Fitzpatrick	
L11	1
L21	5
FGC11134	1



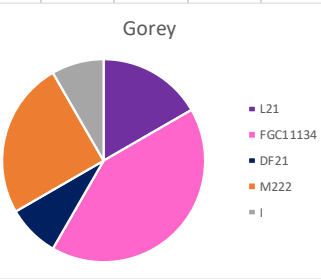
Gilmore	
L11	1
P312	1
L21	2
M222	7
S588	3



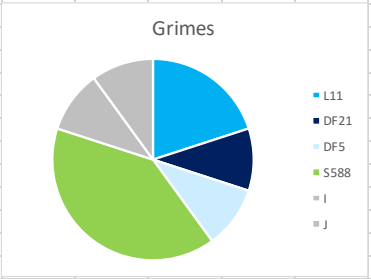
Goodman	
L11	1
L513	6



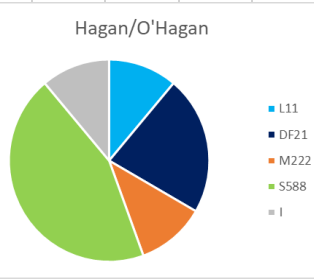
Gorey/Gorry	
L21	2
FGC11134	5
DF21	1
M222	3
I	1



Grimes	
L11	2
DF21	1
DF5	1
S588	4
I	1
J	1

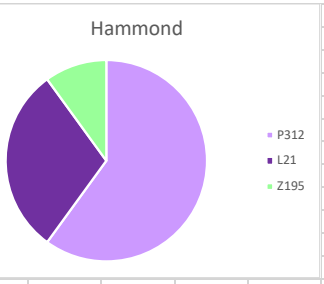


Hagan/O'Haggan	
L11	1
DF21	2
M222	1
S588	4
I	1

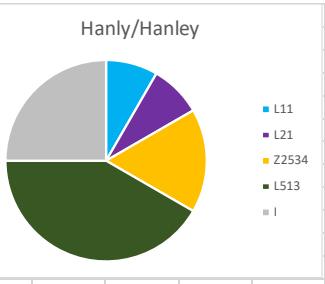


(continued)

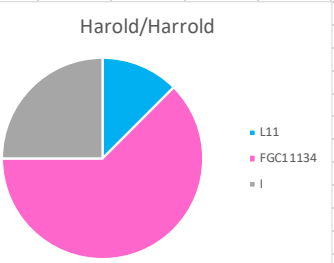
Hammond/Hamond	
P312	6
L21	3
Z195	1



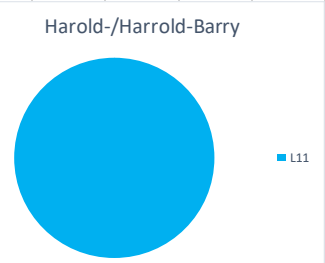
Hanly/Hanley	
L11	1
L21	1
Z2534	2
L513	5
I	3



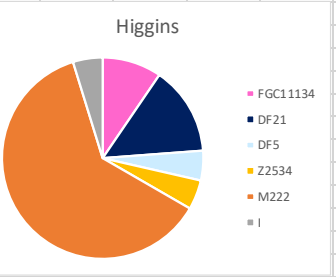
Harold/Harrold	
L11	1
FGC11134	5
I	2



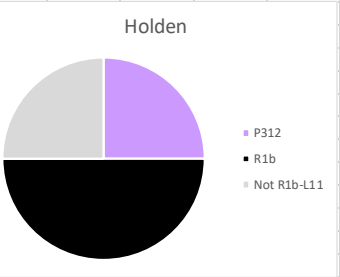
Harold-/Harrold-Barry	
L11	3



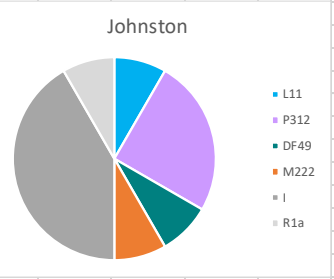
Higgins	
FGC11134	2
DF21	3
DF5	1
Z2534	1
M222	13
I	1



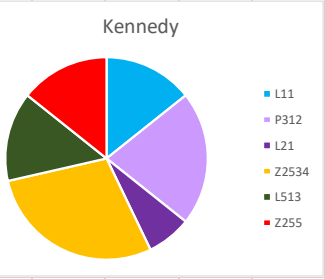
Holden	
P312	1
R1b	2
Not R1b-L	1



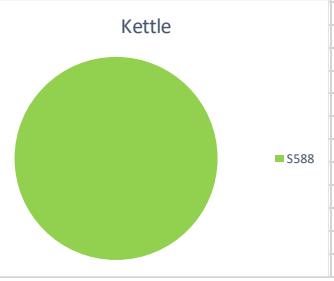
Johnston	
L11	1
P312	3
DF49	1
M222	1
I	5
R1a	1



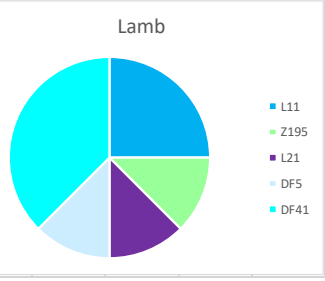
Kennedy	
L11	2
P312	3
L21	1
Z2534	4
L513	2
Z255	2



Kettle	
S588	3

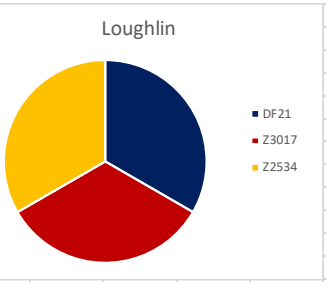


Lamb	
L11	2
Z195	1
L21	1
DF5	1
DF41	3

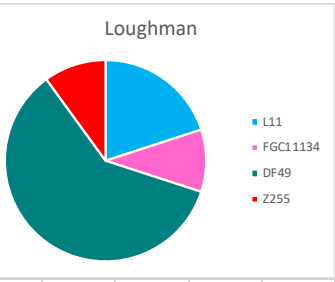


(continued)

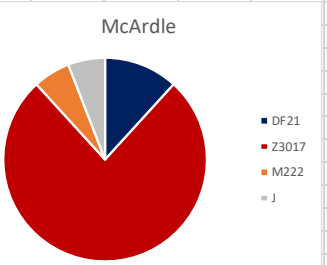
Loughlin	
DF21	1
Z3017	1
Z2534	1



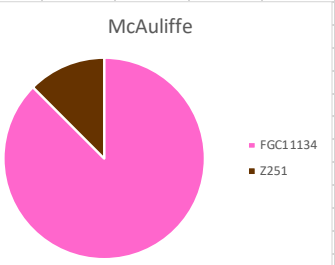
Loughman	
L11	2
FGC11134	1
DF49	6
Z255	1



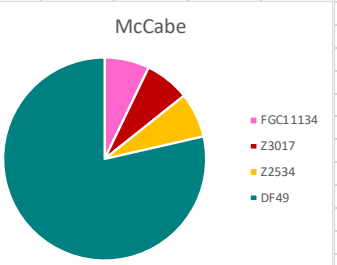
McArdle	
DF21	2
Z3017	13
M222	1
J	1



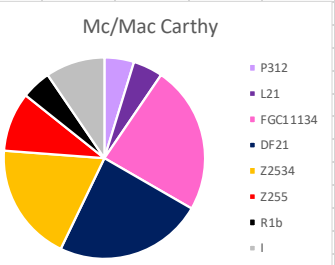
McAuliffe	
FGC11134	7
Z251	1



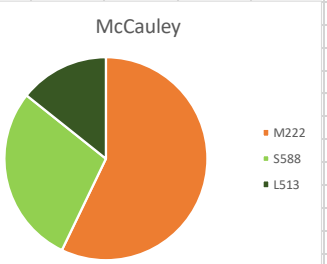
McCabe	
FGC11134	1
Z3017	1
Z2534	1
DF49	11



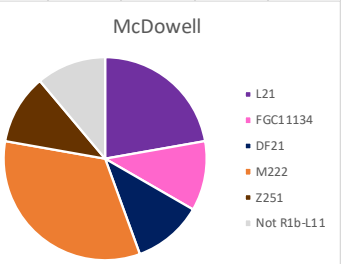
Mc/Mac Carthy	
P312	1
L21	1
FGC11134	5
DF21	5
Z2534	4
Z255	2
R1b	1
I	2



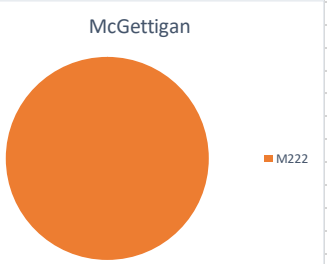
McCauley	
M222	4
S588	2
L513	1



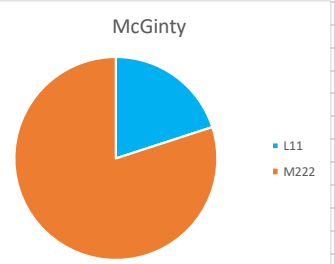
McDowell	
L21	2
FGC11134	1
DF21	1
M222	3
Z251	1
Not R1b-L	1



McGettigan	
M222	7

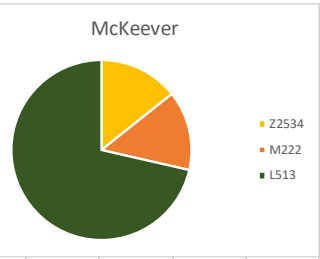


McGinty	
L11	1
M222	4

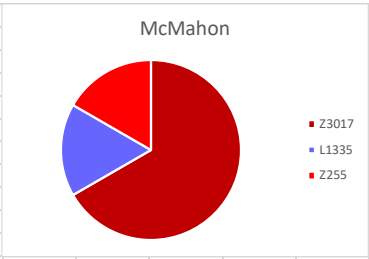


(continued)

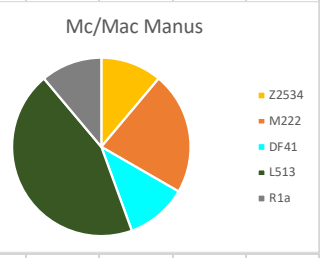
Mc Keever	
Z2534	1
M222	1
L513	5



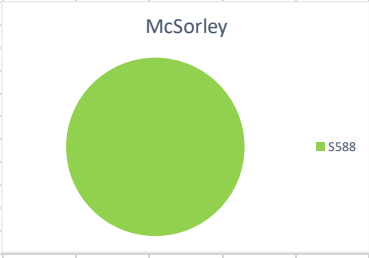
McMahon	
Z3017	4
L1335	1
Z255	1



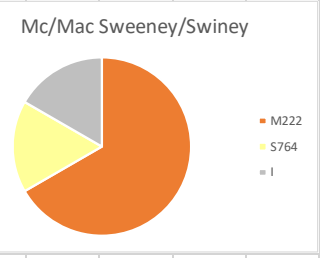
Mc/Mac Manus	
Z2534	1
M222	2
DF41	1
L513	4
R1a	1



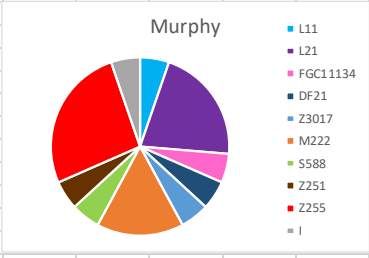
McSorley	
S588	4



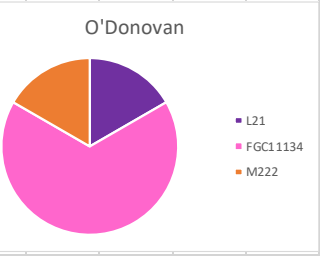
Mc/Mac Sweeney/Swiney	
M222	4
S764	1
I	1



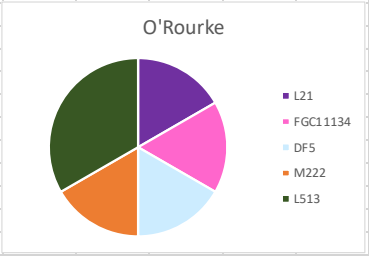
Murphy	
L11	1
L21	4
FGC11134	1
DF21	1
Z3017	1
M222	3
S588	1
Z251	1
Z255	5
I	1



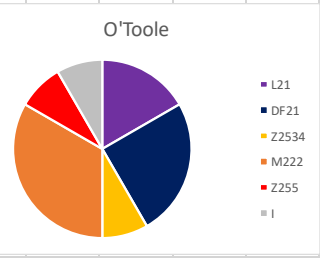
O'Donovan	
L21	1
FGC11134	4
M222	1



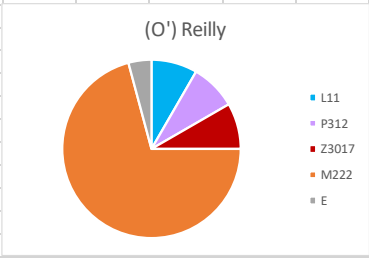
O'Rourke	
L21	1
FGC11134	1
DF5	1
M222	1
L513	2



O'Toole	
L21	2
DF21	3
Z2534	1
M222	4
Z255	1
I	1

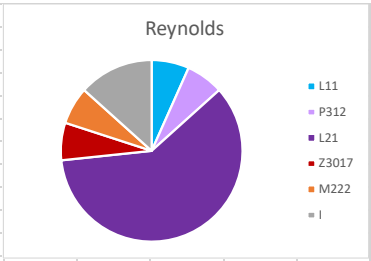


(O') Reilly	
L11	2
P312	2
Z3017	2
M222	17
E	1

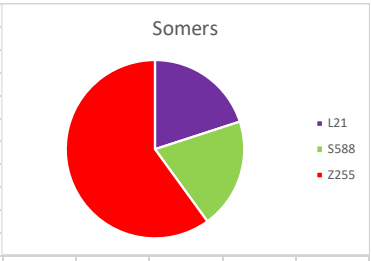


(continued)

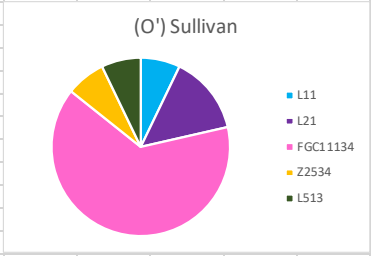
Reynolds	
L11	1
P312	1
L21	9
Z3017	1
M222	1
I	2



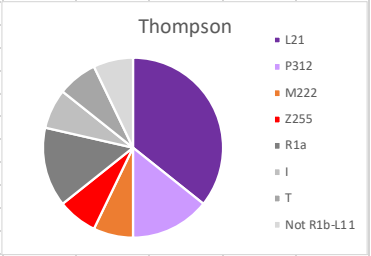
Somers	
L21	1
S588	1
Z255	3



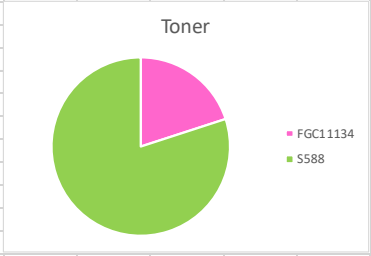
(O') Sullivan	
L11	1
L21	2
FGC11134	9
Z2534	1
L513	1



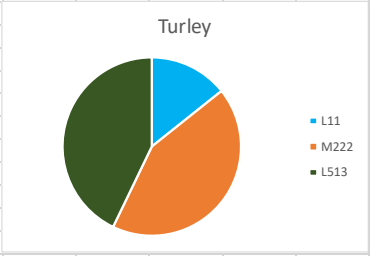
Thompson	
L21	5
P312	2
M222	1
Z255	1
R1a	2
I	1
T	1
Not R1b-L	1



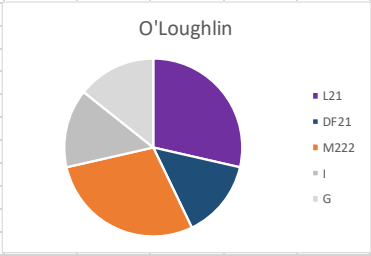
Toner	
FGC11134	1
S588	4



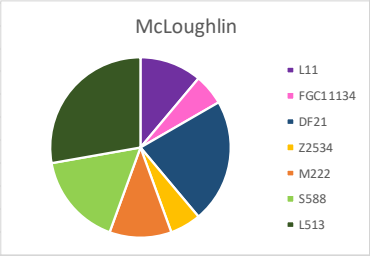
Turley	
L11	1
M222	3
L513	3



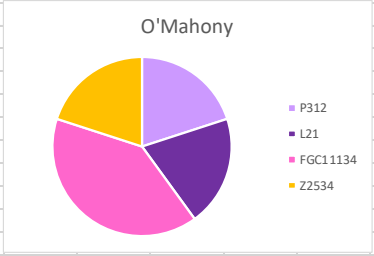
O'Loughlin	
L21	2
DF21	1
M222	2
I	1
G	1



McLoughlin	
L11	2
FGC11134	1
DF21	4
Z2534	1
M222	2
S588	3
L513	5



O'Mahony	
P312	1
L21	1
FGC11134	2
Z2534	1



Costin	
L11	2

Kilpatrick	
E	2

McNea	
M222	2

Taylor	
P312	1

McGorry	
M222	2

REFERENCES

1. Steinsland, G., *Ideology and Power in the Viking and Middle Ages: Scandinavia, Iceland, Ireland, Orkney and the Faeroes*, in *Ideology and Power in the Viking and Middle Ages*. 2011, Brill. p. 1-14.
2. McEvoy, B., et al., *The scale and nature of Viking settlement in Ireland from Y-chromosome admixture analysis.* European Journal of Human Genetics, 2006. **14**(12): p. 1288.
3. Bowden, G.R., et al., *Excavating past population structures by surname-based sampling: the genetic legacy of the Vikings in northwest England.* Molecular Biology and Evolution, 2007. **25**(2): p. 301-309.
4. Scully, M., T. King, and S.D. Brown, *Remediating Viking origins: Genetic code as archival memory of the remote past.* Sociology, 2013. **47**(5): p. 921-938.
5. Roche, R., *The Norman Invasion of Ireland*. 1995: Anvil Books.
6. King, T.E. and M.A. Jobling, *Founders, drift, and infidelity: the relationship between Y chromosome diversity and patrilineal surnames.* Molecular Biology and Evolution, 2009. **26**(5): p. 1093-1102.
7. King, T.E. and M.A. Jobling, *What's in a name? Y chromosomes, surnames and the genetic genealogy revolution.* Trends in Genetics, 2009. **25**(8): p. 351-360.
8. Hill, E.W., M.A. Jobling, and D.G. Bradley, *Y-chromosome variation and Irish origins.* Nature, 2000. **404**(6776): p. 351.
9. Moore, L.T., et al., *A Y-chromosome signature of hegemony in Gaelic Ireland.* The American Journal of Human Genetics, 2006. **78**(2): p. 334-338.
10. Maglio, M.R., *Identifying Y-Chromosome Dynastic Haplotypes: The High Kings of Ireland Revisited.* 2015
11. L'Allemagne Dynastique, T.V. and B.-M. Le, *House of Medici.* Ulwencreutz's The Royal Families in Europe V, 2013: p. 54.
12. Laurence, R. and J. Berry, *Cultural identity in the Roman Empire*. 2001: Psychology Press.
13. Wright, D.M., *A Set of Distinctive Marker Values Defines a Y-STR Signature for Gaelic Dalcassian Families.* Journal of Genetic Genealogy, 2009. **5**(1): p. 1-7.
14. Meyer, K., *Brian Borumha*. 1908.

15. Curtis, E., *TURLOCH O'CONNOR AND THE PRELUDE TO THE NORMAN INVASION*, in *A History of Medieval Ireland (Routledge Revivals)*. 2013, Routledge. p. 54-79.
16. McCartney, D., *North Munster Studies: Essays in Commemoration of Monsignor Michael Moloney*. 1967, JSTOR.
17. Duffy, S., *Brian Boru and the battle of Clontarf*. 2013: Gill & Macmillan Ltd.
18. Sheehan, J., *The longphort in Viking Age Ireland.* Acta Archaeologica, 2008. **79**(1): p. 282-295.
19. Downham, C., *Viking kings of Britain and Ireland: the dynasty of Ívarr to AD 1014*. 2010: Dunedin Academic Press.
20. Floinn, Ó., *R. 1998. The archaeology of the early Viking Age in Ireland.* Ireland and Scandinavia in the early Viking Age, 1998: p. 131-165.
21. Valante, M.A., *The Vikings in Ireland: settlement, trade and urbanization*. 2008: Four Courts Press Dublin.
22. Foster, R.F., *The Oxford illustrated history of Ireland*. 1989: Oxford Paperbacks.
23. MacLysaght, E., *Irish families: their names, arms, and origins*. 1985: Irish Academic Press.
24. Smith, M.T. and D.M. Macraild, *Nineteenth-century population structure of Ireland and of the Irish in England and Wales: An analysis by isonymy.* American Journal of Human Biology: The Official Journal of the Human Biology Association, 2009. **21**(3): p. 283-289.
25. Jesch, J., *The Viking Diaspora*. 2015: Routledge.
26. Cooke, H.J. and B. Smith. *Variability at the telomeres of the human X/Y pseudoautosomal region*. in *Cold Spring Harbor symposia on quantitative biology*. 1986. Cold Spring Harbor Laboratory Press.
27. Jobling, M.A. and C. Tyler-Smith, *The human Y chromosome: an evolutionary marker comes of age.* Nature Reviews Genetics, 2003. **4**(8): p. 598.
28. Cruciani, F., et al., *A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes.* The American Journal of Human Genetics, 2002. **70**(5): p. 1197-1214.
29. Weber, J.L. and C. Wong, *Mutation of human short tandem repeats.* Human molecular genetics, 1993. **2**(8): p. 1123-1128.
30. Genetica. *What is STR DNA Profiling?* [cited 2018 September 29]; Available from: https://www.celllineauthentication.com/what-is-str-profiling-.html.

31. Kayser, M., *Forensic use of Y-chromosome DNA: a general overview*. Human genetics, 2017. **136**(5): p. 621-635.

32. Mulero, J.J., et al., *Development and Validation of the AmpFℓSTR® Yfiler™ PCR Amplification Kit: A Male Specific, Single Amplification 17 Y-STR Multiplex System.* Journal of forensic sciences, 2006. **51**(1): p. 64-75.

33. Rolf, B., et al., *Paternity testing using Y-STR haplotypes: assigning a probability for paternity in cases of mutations.* International journal of legal medicine, 2001. **115**(1): p. 12-15.

34. Kayser, M. and A. Sajantila, *Mutations at Y-STR loci: implications for paternity testing and forensic analysis.* Forensic Science International, 2001. **118**(2-3): p. 116-121.

35. Ballantyne, K.N., et al., *Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications.* The American Journal of Human Genetics, 2010. **87**(3): p. 341-353.

36. Ballantyne, K.N., et al., *A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages.* Forensic Science International: Genetics, 2012. **6**(2): p. 208-218.

37. Butler, J., C.R. Hill, and M. Coble, *Variability of new STR loci and kits in US population groups.* Profiles in DNA, 2012.

38. Kwok, P.-Y., *Single nucleotide polymorphisms: methods and protocols*. Vol. 212. 2003: Springer Science & Business Media.

39. Tiirikka, T. and J.S. Moilanen, *Human chromosome Y and haplogroups; introducing YDHS database.* Clinical and translational medicine, 2015. **4**(1): p. 20.

40. Wetton, J.H., K.W. Tsang, and H. Khan, *Inferring the population of origin of DNA evidence within the UK by allele-specific hybridization of Y-SNPs.* Forensic science international, 2005. **152**(1): p. 45-53.

41. Balanovsky, O., *Toward a consensus on SNP and STR mutation rates on the human Y-chromosome.* Human genetics, 2017. **136**(5): p. 575-590.

42. Thomson, R., et al., *Recent common ancestry of human Y chromosomes: evidence from DNA sequence data.* Proceedings of the National Academy of Sciences, 2000. **97**(13): p. 7360-7365.

43. Y Chromosome Consortium, *A nomenclature system for the tree of human Y-chromosomal binary haplogroups.* Genome research, 2002. **12**(2): p. 339-348.
44. Ashraf, Q. and O. Galor, *The'Out of Africa'hypothesis, human genetic diversity, and comparative economic development.* American Economic Review, 2013. **103**(1): p. 1-46.
45. Hammer, M.F., et al., *Out of Africa and back again: nested cladistic analysis of human Y chromosome variation.* Molecular biology and evolution, 1998. **15**(4): p. 427-441.
46. Chiaroni, J., P.A. Underhill, and L.L. Cavalli-Sforza, *Y chromosome diversity, human expansion, drift, and cultural evolution.* Proceedings of the National Academy of Sciences, 2009. **106**(48): p. 20174-20179.
47. Underhill, P.A., et al., *The phylogenetic and geographic structure of Y-chromosome haplogroup R1a.* European Journal of Human Genetics, 2015. **23**(1): p. 124.
48. Myres, N.M., et al., *A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe.* European Journal of Human Genetics, 2011. **19**(1): p. 95.
49. Larkin, D., *Sampling in Ulster and Wexford.* 2012.
50. Larkin, B.T., *Larkin DNA Project-Ancestral Parish Sampling on the Shannon River.* Journal of Genetic Genealogy, 2010. **6**: p. 1.
51. Sawyer, P.H., *Kings and vikings: Scandinavia and Europe AD 700–1100*. 2013: Routledge.
52. Wilson, D.S., et al., *The Norway Project*. 2016.
53. Gilbert, E., et al., *The Irish DNA Atlas: Revealing Fine-Scale Population Structure and History within Ireland.* Nature, 2017.
54. MacLysaght, E., *The surnames of Ireland*. 1985: Irish Academic Press.
55. van Oven, M., A. Ralf, and M. Kayser, *An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups.* International journal of legal medicine, 2011. **125**(6): p. 879.
56. Untergasser, A., et al., *Primer3Plus, an enhanced web interface to Primer3.* Nucleic acids research, 2007. **35**(suppl_2): p. W71-W74.
57. Vallone, P.M. and J.M. Butler, *AutoDimer: a screening tool for primer-dimer and hairpin structures.* Biotechniques, 2004. **37**(2): p. 226-231.
58. Consortium, G.P., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68.

59. Aliferi, A., et al., *UK and Irish Y-STR population data—A catalogue of variant alleles.* Forensic Science International: Genetics, 2018. **34**: p. e1-e6.
60. Purps, J., et al., *A global analysis of Y-chromosomal haplotype diversity for 23 STR loci.* Forensic Science International: Genetics, 2014. **12**: p. 12-23.
61. FamilyTreeDNA. *The Norway DNA Project*. [cited 2018 Spetember 27]; Available from: https://www.familytreedna.com/public/Norway/default.aspx?section=yresults.
62. Meirmans, P.G. and P.W. Hedrick, *Assessing population structure: FST and related measures.* Molecular ecology resources, 2011. **11**(1): p. 5-18.
63. Schneider, S., D. Roessli, and L. Excoffier, *Arlequin: a software for population genetics data analysis, version 2.000.* Genetics and biometry laboratory, department of anthropology, University of Geneva, Switzerland, 2000.
64. XLSTAT, *XLStat version 2018-01-29*. 2018, Addinsoft SARL Paris.
65. Gouy, A. and M. Zieger, *STRAF—A convenient online tool for STR data evaluation in forensic genetics.* Forensic Science International: Genetics, 2017. **30**: p. 148-151.
66. Nevski, A. and M. Gentula. *NEVGEN Y-DNA Haplogroup Predictor*. [cited 2018 Spetember 19] Available from: http://www.nevgen.org/.
67. Cruciani, F., et al., *Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa.* The American Journal of Human Genetics, 2004. **74**(5): p. 1014-1022.
68. Di Giacomo, F., et al., *Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe.* Human genetics, 2004. **115**(5): p. 357-371.
69. Rootsi, S., et al., *Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus.* European journal of human genetics, 2012. **20**(12): p. 1275.
70. Mendez, F.L., et al., *Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa.* Human biology, 2011. **83**(1): p. 39-53.
71. Poznik, G.D., et al., *Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences.* Nature genetics, 2016. **48**(6): p. 593.
72. Dupuy, B.M., et al., *Geographical heterogeneity of Y-chromosomal lineages in Norway.* Forensic science international, 2006. **164**(1): p. 10-19.

73. Broich, G., M. Fabbri, and F.M. Avato, *Specific European Y-Chromosome Haplotype I and its subclasses: migrations and modern prevalence.* 2015
74. Solé-Morata, N., et al., *Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ.* Scientific Reports, 2017. **7**(1): p. 7341.
75. Wainwright, F.T., *The Scandinavians in Lancashire*. 1948: Burleigh Press.
76. Guo, X. and R. Elston, *Linkage information content of polymorphic genetic markers.* Human Heredity, 1999. **49**(2): p. 112-118.
77. Kruskal, J.B., *Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis.* Psychometrika, 1964. **29**(1): p. 1-27.
78. Holland, S.M., *Principal components analysis (PCA).* Department of Geology, University of Georgia, Athens, GA, 2008: p. 30602-2501.
79. Martinez-Cadenas, C., et al., *The relationship between surname frequency and Y chromosome variation in Spain.* European Journal of Human Genetics, 2016. **24**(1): p. 120.
80. Brink, S. and N. Price, *The Viking World*. 2008: Routledge.
81. Etchingham, C., *A New History of Ireland. Vol. I: Prehistoric and Early Ireland*. 2007, Oxford University Press.
82. Myhre, B., et al., *Ireland and Scandinavia in the early Viking age*. 1998: Dublin; Portland; Four Courts Press.
83. YOUNG, J.I., *A NOTE ON THE NORSE OCCUPATION OF IRELAND 1*. History, 1950. **35**(123): p. 11-33.
84. Robbins, D.D., *The Oxford Handbook of Names and Naming*. 2018, Taylor & Francis.
85. nic Bryan, M.E., *Constructing SCA Household Names Using Irish Gaelic Family and Clan Name Models.* 2006
86. O'neill, E.B. and J.D. McLaughlin, *Insights Into the O'Neills of Ireland from DNA Testing.* Journal of Genetic Genealogy, 2006. **2**(2).
87. McEvoy, B., K. Simms, and D.G. Bradley, *Genetic investigation of the patrilineal kinship structure of early medieval Ireland.* American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 2008. **136**(4): p. 415-422.

88. MacShamhráin, A., *Brian Bóruma, Armagh and High Kingship.* Seanchas Ardmhacha: Journal of the Armagh Diocesan Historical Society, 2005. **20**(2): p. 1-21.
89. Ferguson, R., *English Surnames and their place in the Teutonic Family*. 1858.
90. Harding, S., *The Wirral carrs and holms.* Journal of the English Place-Name Society, 2007. **39**: p. 45-57.
91. Larmuseau, M., et al., *Increasing phylogenetic resolution still informative for Y chromosomal studies on West-European populations.* Forensic Science International: Genetics, 2014. **9**: p. 179-185.
92. Petrovic, V., et al., *Assessment of mutation rates for PPY23 Y chromosome STR loci in Serbian father-son pairs.* Forensic Science International: Genetics.
93. Davis, G., *The Thomas Jefferson paternity case.* Nature, 1999. **397**(6714): p. 32.