5-2012

Vikings, merchants, and pirates at the top of the world : Y-chromosomal signatures of recent and ancient migrations in the Faroe Islands.

Allison Mann
University of Louisville

VIKINGS, MERCHANTS, AND PIRATES AT THE TOP OF THE WORLD: Y-CHROMOSOMAL SIGNATURES OF RECENT AND ANCIENT MIGRATIONS IN THE FAROE ISLANDS

By

Allison Mann
B.S., B.A., University of Louisville, 2009

A Thesis
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Master of Arts

Department of Anthropology
University of Louisville
Louisville, Kentucky

May 2012

VIKINGS, MERCHANTS, AND PIRATES AT THE TOP OF THE WORLD: Y-CHROMOSOMAL SIGNATURES OF RECENT AND ANCIENT MIGRATIONS IN THE FAROE ISLANDS

By

Allison Mann
B.S., B.A., University of Louisville, 2009

A Thesis Approved on

4/20/2012
Date

by the following Thesis Committee:

Christopher Tillquist, Thesis Director

Fabian Crespo

Michael Perlin

ACKNOWLEDGEMENTS

It is my pleasure to thank those who made this thesis possible. In particular I am indebted to my advisor, Dr. Christopher Tillquist. His advice and support has been invaluable during this process. I would also like to thank the members of my graduate committee, Dr. Fabian Crespo and Dr. Michael Perlin, for their guidance and suggestions. Finally I must express my gratitude to my family and friends, without whose unfailing encouragement this thesis would not have been possible.

ABSTRACT

VIKINGS, MERCHANTS, AND PIRATES AT THE TOP OF THE WORLD: Y-CHROMOSOMAL SIGNATURES OF RECENT AND ANCIENT MIGRATIONS IN THE FAROE ISLANDS

Allison Mann

April 20, 2012

The Faroe Islands are a small archipelago in the North Atlantic. With a current population of approximately 49,000 individuals and evidence of high levels of genetic drift, the Faroese are thought to have remained highly homogeneous since their settlement by Vikings around 825 CE. Despite their geographic isolation, however, there is historical evidence that the Faroese experienced sporadic contact with other populations since the time of founding. This study set out to distinguish the signal of the original founders from later migrants. Twelve Y-chromosomal STR markers were scored for 139 Faroese males. Median-joining networks were constructed to determine the phylogenetic relationships within the Faroese and between likely parental populations. Dispersal patterns of individuals around Faroese haplogroups suggest different times of haplogroup introduction to the islands. The most common haplogroup, R1a, consists of a large node with a tight network of neighbor haplotypes, such that 62.06% of R1a individuals are $\leq$ two mutational steps away. This pattern may represent the early founder event of R1a in the Faroes. Other distributions document more recent introductions to the islands. The overall pattern is one of a strong founder effect followed by minor instances of later migrations.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The Faroes are a small volcanic chain of islands in the North Atlantic Ocean situated approximately midway between the southern coast of Iceland and the northern tip of Scotland. There is ample evidence that the population living on the Faroe Islands has remained relatively small and isolated due to geographic and socio-political pressures.

Because of their unique demographic history and secluded location, the Faroese, along with other North Atlantic island populations like Iceland, are thought to be uniquely suited for the study of specific genetic disorders. An example of this is Bardet-Biedl syndrome, a congenital disease that affects the eyes, which presents itself at unusually high frequencies in the Faroes (Hjortshoj et al., 2009). In fact, the disease is at its highest prevalence in the Faroe Islands while it remains fairly rare among the rest of the world's population (Hjortshoj et al., 2009, p. 412).

Despite their geographic isolation, however, there is historical and genetic evidence of sporadic contact between the Faroese and other European and non-European groups since the islands were first colonized by Viking migrants in the ninth century. This is due in part to the flourishing maritime culture found in the Faroes which has long attracted people to the North Atlantic. This study uses Y-chromosomal haplotype data in tandem with published historical, literary, and archaeological evidence to distinguish between ancient and recent population events in the Faroe Islands.

A brief history of the Faroes: The Faroe Islands were settled around 825 CE by Viking settlers from western Norway (Jorgensen et al., 2002, p. 382). There is some historical and paleo-botanical evidence that before this time the Faroes were inhabited by Celtic monks from as early as 650 CE (Jorgensen et al., 2004, p. 19). However, this remains controversial, and the existing archaeological and documentary evidence is very slight (Magnusson, 2010, p. 183). The Viking colonization of the Faroe Islands represents a larger pattern of mass emigration of Viking settlers and raiders around this time. The three centuries starting around 800 CE are known as the 'Viking Age' and the reach of this population is well documented by contemporary witnesses (Magnusson, 2010, p. 7). Potential reasons for this Viking diaspora include population and political pressures as well as reasons of personal gain (Arge et al., 2005, p. 599). Certainly the Faroe Islands at the time must have had sufficiently enticing marine and terrestrial resources for the new settlers (Arge et al., 2005, p. 599).

The earliest written evidence of the Viking settlement in the Faroe Islands comes from the *Færeyinga Saga* (Faroese Saga). The Saga names the first man to settle the islands, Grim Kamban, who was later given a wife by a woman on her way to Iceland, thus beginning the first family of the Faroe Islands (Johnston, 1975, p. 19). From this first story in the Saga there are tantalizing clues to the origins of the initial settlers. Grim Kamban's name has been described by some scholars as attributable to both Norse and Gaelic origins (Arge et al., 2005, p. 602). The woman traveling to Iceland, Aud the Deep-minded who bestowed a wife upon Grim, had connections to the Celto-Norse world including the Sudreys (the Hebrides, Isle of Man and the Islands of the Firth of Clyde) (Young, 1979, p. 3).

Others have suggested that the name Grim Kamban is not Norse at all and is instead evidence that the original Faroese settlers had their origins in the Hebrides or Ireland (Young, 1979, p. 2). Nauerby (1996) suggests that the original settlers of the islands were Norsemen living in the British Isles around 800 CE followed by a

second migration directly from Norway (Nauerby, 1996, p. 29). This is a possibility as it is well known that people of Norse descent were living far from Norway at this time, including those who had for many years resided in Ireland or Scotland (Arge et al., 2005, p. 604). Conjecture that some populations living in the North Atlantic are the result of an admixed Norse-Celtic founder group is strengthened by various genetic studies demonstrating this fact. It is certainly the case among Icelandic and Faroese women who are much more likely to have their origins in the British Isles than Norway (Helgason et al., 2001)(Als et al., 2006). Additionally, a Celtic runic stone dated between 865 and 1,000 CE has been found in Kirkjubøur on Streymoy Island in the Faroes (Young, 1979, p. 107). Even the Faroese language, which emerged in the Middle Ages, has notable Celtic influences (Jorgensen et al., 2004, p. 19)(Young, 1979, p. 107). Certain place names within the Faroes seem to suggest a Celtic imprint including Argisa, Argisfossur, Ergibyrgi and Ergidalur (Young, 1979, p. 3). Importantly, this also is an occurrence with Icelandic place names (Arge, 1991, p. 103). This Celtic connection may have been influenced by slaves who very likely accompanied the earliest settlers. Slaves were not uncommon in Europe in the ninth century and Irish, Slavic and even Scandinavians who had lost their freedoms played an important role for the early Vikings (Johnston, 1975, p. 127)(Jones, 1984, p. 148). However, the vast majority of place names in the Faroes are Norse in origin (Arge et al., 2005, p. 604), and it remains unclear what role and how extensive an influence the Celtic population had on the original founding group. The Faroese Saga itself is questionable in its accuracy and the author, most likely from Iceland, often reports inconsistent assertions surrounding family histories and errors in geography (Johnston, 1975, p. 121). Regardless, the Faroese people are protective of their Saga (Johnston, 1975, p. 121) and it remains the most complete written record of early Faroese life.

It appears that the Faroes and Iceland were settled at approximately the same time. Archaeological data and historical documents point to the settlement of

Iceland around 870 CE, approximately 45 years after the first Faroese colonist (Als et al., 2006, p. 497). Although the settlement of the Faroes remains somewhat obscure, it has been assumed that the settlement pattern in the Faroe Islands was similar to that found in Iceland but on a smaller scale (Arge et al., 2005, p. 615). Icelandic Y-chromosome evidence suggests an admixed founder population of Gaelic and Norse ancestry (Helgason et al., 2000a, p. 697). It has been suggested that like the Faroe Islands, the original colonizers of Iceland were family groups of Scandinavian men and women native to the British Isles living in coastal settlements outside of Norway (Helgason et al., 2000a, p. 697). The size of the original founding population of Iceland is estimated at around 8,000 to 20,000 individuals over a sixty year period, followed by dramatic reductions in population size due to various diseases, famines and natural disasters. One of these events reduced the size of the population by 20% after a volcanic eruption in the late 1700s (Helgason et al., 2000b, p. 1000).

Historical population size: As there is little information available on the original founders of the Faroe Islands, it is unknown how large the population was at its beginnings. What is clear is that the Faroese population remained rather small until very recently when industry and commerce triggered an expansion. Estimates of population size before the 1800s in the Faroes come from land usage appraisals. With a traditional economy and the amount of arable land found in the Faroes, it is estimated that the islands could have supported around 4,500 people, or 8.3 people per square mile (Wylie, 1987, p. 174). The population may have been smaller than this originally, being around 4,000 in the late 1300s (Jorgensen et al., 2002, p. 382), reaching into the 9,000s in the late nineteenth century and finally peaking around 49,000 today (Table 1). The historical growth rate in the Faroes was irregular, and while there were some signs of population growth in the 1800s, the Faroes only saw a rapid expansion in the nineteenth century, concordant with

the development of trade and the fishing industry (Coull, 1966, p. 162). Despite recent growth, migration from the Faroe Islands remained quite small in the nineteenth century; on average only around a dozen people emigrated from the islands per year (Coull, 1966, p. 163).

The demographic history of the Faroe Islands can be characterized by a small founder population, slow population growth over much of its early history and intermittent population contractions due to disease epidemics (Jorgensen et al., 2004, p. 26). One of the most catastrophic of these events was the arrival of the bubonic plague, and like much of Europe, the Faroe Island's population was decimated. The disease appeared on the islands in the 1350s and it is estimated that mortality rates may have reached over 70 percent, leaving whole villages eradicated (Wylie, 1987, p. 14). After the plague passed, legends suggest that the islands were partially resettled from Norway to replenish the dwindling population (Wylie, 1987, p. 14). The early history of the Faroes is in accordance with the highly genetic homogeneity of the modern Faroese population.

Trade: In spite of the geographic isolation of the Faroese archipelago, there is ample historical and archaeological evidence that from the earliest days of settlement the Faroese had strong cultural and commercial connections to the outside world. One of the earliest archaeological sites in the Faroes, a silver coin hoard buried in 1090 CE on the island of Sandur, indicates that at least some early Faroese were fairly eminent and had strong connections to Europe and Scandinavia (Arge et al., 2005, p. 613).

It appears that early on the Faroese were in need of imports from other countries because of the lack of arable land and absence of trees on the islands. The unavailability of trees in the Faroes was apparently a feature of the islands pre-settlement, and not due to human deforestation as was the case with the Norse settlement of Iceland (Lawson et al., 2005, p. 652). Tree birch macro fossils are

TABLE 1

Population growth in the Faroes from 1801 to 2011. Table adapted from (Wylie, 1987, p. 114). *Data for 2011 from (CIA, 2011)

Year	**Population**	**Est. yearly growth**
1801	5265	n/a
1834	6928	0.83
1840	7314	0.91
1845	7781	1.24
1850	8137	0.9
1855	8651	1.22
1860	8922	0.61
1870	9992	1.13
1880	11220	1.16
1890	12955	1.48
1901	15230	1.47
1906	16349	1.4
1911	18000	1.92
1921	21352	1.71
1925	22835	1.69
1930	24200	1.17
1935	25744	1.24
1945	29178	1.25
1950	31781	1.71
1955	32505	0.45
1960	34596	1.24
1965	37205	1.45
1970	38612	0.75
1980	43609	1.22
2011	49267	0.42*

occasionally found in the peat that has accumulated on the islands (Lawson et al., 2005, p. 660); however, there are not enough to suggest sufficient forestation of the Faroes. Along with timber, the most important imports included luxuries, corn, and malt (Young, 1979, p. 95). The Faroes had products that were of value as exports as well, and it is through trade that the Faroese kept in contact with populations beyond the islands (Young, 1979, p. 94). Among the main exports from the Faroes in the Middle Ages, wool, butter, cod and yarn were the most important, mostly entering the trading post of Bergen, Norway (Young, 1979, p. 94). The Faroese traded not only with Norway but also with the British Isles and very likely populations much further abroad (Wylie, 1987, p. 9). An archaeological excavation of a classic Viking hall in Kvivik unearthed evidence of the importance of trade and fishing in early Faroese society (Magnusson, 2010, p. 183). Among the artifacts uncovered were tools for Faroese industry: "stone sinkers for fishing-lines [...] spindle whorls and loom weights to work wool into textiles" (Magnusson, 2010, p. 183) along with imported luxury items like: "beads of amber and silver-foiled glass" (Magnusson, 2010, p. 183). Finds like these illustrate the importance of commerce between the Faroes and the outside world.

The Hanseatic League: An important trading conduit between the Faroese and mainland Europe was implemented by The Hanseatic League. The Hanseatic League was a collective of merchants that played a major role in the economics of Europe starting in the Middle Ages, stretching to the 17th century in some regions (Mehler, 2009, p. 90). League members were from a collection of villages across Europe bridging Bruges to London in the western part of the continent, Cologne to Bergen in the northernmost part, and as far as the eastern European regions of Estonia, Livonia and Novgorod (Mehler, 2009, p. 90). Remnants of the material culture of the Hanseatic League still persists in much of the North Atlantic (Mehler, 2009, p. 89). Among the various island nations that the Hanseatics traded with are

Scotland, the Orkney and Shetland Islands, Iceland, the Faroes and perhaps even Greenland (Mehler, 2009, p. 90). Documentation from the period indicates that the Hanseatic League gained economic interest in the Faroes as early as the late 14th century and set up trading posts on two of the islands: the village of Tórshavn on Streymoy in the central islands and Krambatangi on Suduroy in the south (Mehler, 2009, p. 90). The extent of contact between the Faroes and the Hanseatic League is made apparent by archaeological finds on the islands. Included in these are Hamburg ells (standardized measuring implements) that replaced the old Norwegian ell (*stikka*) and were in use until the 1680s in the Faroes (Mehler, 2009, p. 92).

Trade and politics: Because of their strategic trading value, the Faroes were subject to multiple trading restrictions as they came under the rule of different monarchies. The stranglehold over the internal affairs of the Faroes, including their church, economy and local government (Bronner, 1972, p. 6), was first imposed by Norway, which claimed the islands as a Christian colony in the early 11th century (Wylie, 1987, p. 18), and then by Denmark, which appropriated Norway and all of its territory in 1380 (Jackson, 1979, p. 35). The trading system set up by the Faroese and Hanseatic merchants was replaced by a trading monopoly instituted by the Danish monarchy. This monopoly prohibited trade between the Faroes and anyone not affiliated with the crown (Jackson, 1979, p. 36). Faroese who wished to trade with foreign vessels were forbidden from doing so, even when they were well out of sight of land (Wylie, 1987, p. 31).

This trade embargo was in place until 1856 (Jackson, 1979, p. 36) but was apparently not strictly adhered to, despite the threat of severe punishment for those who disobeyed. Among those who continued to trade with the Faroes was the Hanseatic League, which had been forbidden to trade with any site north of Bergen, including the Faroes, even earlier than the Danish proscription (as early as 1294) (Wylie, 1987, p. 18)(Mehler, 2009, p. 90). Hanseatic merchants are documented as

trading with the Faroes in 1416, in direct defiance of the ban on their movements (Mehler, 2009, p. 90). Trading ships who wanted to skirt the ban on commerce merely found an isolated port (of which there are an abundance in the Faroes) and exchanged Faroese goods for staples such as flour (Wylie, 1987, p. 29). The physical distance from Denmark (785 miles separated Copenhagen from the islands (Coull, 1966, p. 159)) may have facilitated the proliferation of illegal trading on the islands, which became very important for the economic development of the Faroes in subsequent years (Young, 1979, p. 95).

Maritime culture: A key component of modern Faroese culture is the emphasis on fishing as both an economic and cultural venture. The islands themselves are situated in an area rich in marine resources. From the ocean cod, haddock, coal fish, and marine mammals like seal and pilot whale are plentiful, while trout and salmon are found in terrestrial waters (Arge et al., 2005, p. 601). Despite this, there is no mention of fishing in the Faroese Saga, which has led some to speculate that the earliest Faroese were focused on agriculture and animal husbandry (Young, 1979, p. 103).

Later on fishing gained increasing importance for the Faroese, spurred by the Catholic Church's ban on meat eating on Fridays and during Lent (Young, 1979, p. 104). There is archaeological evidence that fish were consumed as food during the Viking period, but it doesn't appear to have reached the economic importance it enjoys today until the late twelfth century, when cod became the main export of the Faroes (Young, 1979, p. 104). Another important fishing tradition in the Faroes, the pilot whale hunt, likely began early in their history, as the whales themselves provided not only food but fuel and their bones provided raw materials to create a number of items (Young, 1979, p. 104). Today this tradition continues with the pilot whale hunt representing the last of the collective fishing traditions in the Faroes (Jackson, 1979, p. 60).

Perhaps the most important event to increase the fishing trade in the Faroes was the repeal of the oppressive trade laws in 1856 (Wylie, 1987, p. 39). Among other commercial ventures, the Faroese fish trade expanded greatly and its main export, dried fish, enjoyed the number one spot in trade until the 1960s (Jackson, 1979, p. 53). Like their Icelandic neighbors, the Faroese economy today is based almost entirely on the fishing trade (Coull, 1966, p. 159)(Jackson, 1979, p. 32). In fact, the Faroes held a world record for the most fish caught per person living on the islands in 1958 (at 3.5 tons per individual) (Jackson, 1979, p. 54). In more recent years, the rich fishery of the Faroes has attracted men and women from all over the world who wish to exploit the marine resources of the North Atlantic. Up until the 1960s, foreign fishing vessels were allowed to fish in close waters around the Faroes (Zeller and Reinert, 2004, p. 404). Since 1977 foreign fishing in the Faroes has declined, yet the vast majority of open ocean fish caught is still largely by foreign vessels (Zeller and Reinert, 2004, p. 404).

Pirates: Unfortunately for the residents of the islands, the Faroes also attracted visitors of a more unsavory nature: pirates. Evidence suggests that English, Irish and French pirates began raids on the Faroes around the 1500s (Young, 1979, p. 95). Attacks didn't stop at assaults to Faroese merchant ships but often included land raids, in which invaders would steal cattle and the Faroese people themselves (Young, 1979, p. 95). The vehemence reserved for Irish pirates is preserved in the Faroese language. The word *iri* historically means an Irish or Gaelic speaking pirate (Young, 1979, p. 95). From this root there are a multitude of opprobrious words "meaning Irish rabble, riff-raff or scum" (Young, 1979, p. 95). Because of their connection to larger trade groups like the Hanseatic League, trading posts on the Faroes were often a target. In 1580 the trading post in Tórshavn was attacked while the post in Krambatangi had to be abandoned because of the pirate threat (Mehler, 2009, p. 96).

One particularly harrowing account of a pirate attack preserved in traditional Faroese oral histories was the attack of Barbary corsairs on Suduroy in the 1600s. The Barbary corsairs were Ottoman empire privateers who traveled as far north as Iceland to capture people to be sold as slaves. In 1627 corsairs from a port in Algiers attacked different coastal settlements in Iceland and carried off 400 captives (Lewis, 1973, p. 140). In the summer of 1629 the Faroes were attacked at Hvalba on Suduroy (Wylie, 1987, p. 28). The invasion is documented in a petition to King Christian IV of Denmark where it was reported that six people were killed and thirty carried away (West, 1980, p. 46). The document also notes that one of the two ships that landed on the islands was wrecked (West, 1980, p. 47), suggesting that part of the crew may have been forced to stay on the islands and met an undetermined fate.

The genetics of the Faroes: Previous population genetic studies of the Faroe Islands indicate a mixed Scandinavian and Celtic founder group (Jorgensen et al., 2004)(Als et al., 2006). Male and female origins are distinct in that the Faroese mtDNA is overwhelmingly Celtic in origins while the vast majority of Y-chromosomes are attributable to Scandinavian populations (Als et al., 2006). The Faroese are homogeneous relative to other European populations, largely due to the effects of genetic drift on the archipelago (Als et al., 2006, p. 503).

Faroese Y-chromosomes, while bearing no significant similarity to any one particular British or Scandinavian group (Jorgensen et al., 2004, p. 24), do have probable provenance in Norway, Sweden, Iceland and to a lesser extent, the British Isles (Jorgensen et al., 2004). Previous investigations suggest that the majority of Faroese Y-chromosomes originated in Norway and Sweden, however, historical evidence suggests that Norway is a much more likely progenitor for the first Faroese colonizers (Jorgensen et al., 2004, p. 26).

The most common haplogroup among Faroese males, R1a, is also the least variable at the haplotype level when compared to other common groups (Jorgensen

et al., 2004, p. 23). This highlights the high level of homogeneity in the population and supports an interpretation of strong genetic drift (Jorgensen et al., 2004, p. 23). Jorgensen et al. (2004) found that haplotypes within this group have a higher affinity to Iceland and Norway. Both mtDNA and Y-chromosome data suggest a high degree of similarity between the Faroese and Icelandic populations, lending credence to hypotheses that the two populations had similar ancestral populations and that later migrations between the Faroes and Iceland occurred (Als et al., 2006)(Jorgensen et al., 2004). Finally, despite close connections between the Faroes and Denmark, previous study of Faroese Y-chromosome haplotypes found little evidence of admixture between the two groups (Jorgensen et al., 2004, p. 26). In the current inquiry it should be expected that Faroese Y-chromosome haplogroup and haplotype frequencies display evidence of homogeneity and strong genetic drift, a high affinity to some Scandinavian populations (i.e., Norway and Sweden), but scant evidence of admixture between the Faroes and Denmark.

CHAPTER II

THE Y-CHROMOSOME: RELEVANCE AND EVOLUTION

The Y-chromosome is particularly well-suited to these types of genetic questions because of the hierarchical nature of types of polymorphic markers. Unlike autosomal DNA, the Y-chromosome has a fourfold effective population size and patrilineal transmission, making it extremely sensitive to genetic drift and bottlenecks (Roewer et al., 2005, p. 286). The majority of the Y-chromosome is non-recombining and as a result is obstensively inherited as a single unit paternally, providing an analogous DNA sequence for the maternally inherited mtDNA (Linares, 1999, p. 192). The Y-chromosome is inherited this way because it does not undergo recombination over the majority of its length (Jorgensen et al., 2004). This non-recombining portion (NRY) of the human Y-chromosome is a linked haplotype (Karafet, 1999, p. 250).

Within the Y-chromosome, different mutational events contribute different levels of resolution in the temporal and spatial origins of an individual or localized population. Single nucleotide polymorphisms (SNPs) on the Y-chromosome inform haplogroup association. The relatively slow rate of mutation of these markers conveys information on more ancient demographic history events. Linked haplotypes of microsatellites within haplogroups representative of lineages communicate more recent demographic events. At a more refined level, differences in mutation rates between individual microsatellite (STRs) loci that form haplotypes can be used to document the impact of contemporary evolutionary and demographic events (deKnijff, 2000).

Evolution of the NRY occurs during meiosis when realignment is amiss and results in a gain or loss of one repeat unit at a particular locus (Buschiazzo and Gemmell, 2006, p. 1043). This gain or loss of one repeat unit is interpreted with the stepwise mutation model (Ohta and Kimura, 1973), which has two main assumptions; (1) the mutation rate at a particular locus is the same for all alleles possible at that locus; and, (2) the way in which a mutation changes the length of a particular locus does not depend on the length of the original allele (Donnelly, 1999, p. 121). However, this is a simplification of how autosomal and Y-chromosome mutational processes occur. Contradictions to the original assumptions of the stepwise mutation model include occasional multistep mutations in natural populations (Rienzo et al., 1994) and evidence that mutation rates may be dependent on the progenitor allele length (Xu et al., 2000). Regardless, the stepwise mutation model is a useful paradigm that allows researchers to make important observations in population genetics and therefore is assumed for the purposes of this investigation.

Because microsatellites mutate at relatively high rates as compared to other genetic markers commonly used in population genetics, they tend to communicate the level of affiliation between very closely related populations (Goldstein et al., 1995, p. 463). Within Y-chromosomal microsatellites, however, there is evidence of significant differences between the mutation rate of one locus versus the other (deKnijff, 2000, p. 1059). A mechanism also appears to be in place to prevent unbridled expansion of microsatellites with longer alleles tending to mutate to shorter lengths (Ellegren, 2000, p. 400). Further complicating the stepwise mutation model, expansions seem to be more common among dinucleotide repeats while trinucleotide repeats exhibit no such trend (Ellegren, 2000, p. 400). Structural differences between microsatellites located on the Y-chromosome affect the rate and mode of evolution of a particular type of locus. This must be considered when determining the age of haplogroups within a population. To estimate the age of the

haplogroups present in the Faroese, this paper considers all three levels of mutational age; the haplogroups themselves, associations between linked haplotypes, and mutation rates at the level of individual loci.

CHAPTER III

OBJECTIVE & HYPOTHESES

The geographic isolation of the Faroes suggests that since the initial settlement of the islands there has been little migration (Als et al., 2006, p. 498). The potential for admixture from sailors and traders from other European countries has been noted in previous studies on the genetic makeup of the Faroese (Als et al., 2006)(Jorgensen et al., 2004). However, these admixtures have not been fully documented. This study attempts to distinguish between recent admixture in the Y-chromosome of the Faroes and those Y-chromosomal haplotypes that are representative of the original founding population. Three main hypotheses will be investigated; (1) the majority of the Faroese samples is highly homogeneous as a function of strong genetic drift, (2) admixture from other regional European populations is evident; and, (3) there is evidence of admixture from populations outside of the region.

CHAPTER IV

METHODS

Y-chromosomal DNA samples from 139 Faroese males were extracted and scored for 12 different STR markers (DYS19, DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426 and DYS438). Because of the highly variable nature of DYS389I and DYS389II, DYS389b was calculated by subtracting DYS389I from DYS389II to infer the true length of the loci (Roewer et al., 1996). Since DYS385 is characterized by a duplication, the shorter allele was assigned to DYS385a while the longer was assigned to DYS385b within each individual haplotype (Pacheco et al., 2005, p. 147). The samples themselves were collected from three geographically disparate locations within the island group; the city of Klaksvik in the northern islands; Tórshavn in the central islands; and Tvoroyri in the south (Figure 1).

Putative parental population – Irish, Norse, Danish and Swedish – haplotype data were compared to the Faroese data set presented here. Icelandic haplotype data was also included in this study because of their reportedly close relationship to the Faroese population (Jorgensen et al., 2004). The Icelandic data were found in Helgason et al., 2000a. Danish haplotype frequencies are based on haplotype data reported by Hallenberg et al., 2005. Irish haplotypes are based on data reported by Ballard et al., 2006. Norwegian samples were previously sampled ranging from the southernmost tip of Norway to above the Arctic circle, and scored by C. Tillquist. Swedish haplotypes were also sampled and scored by C. Tillquist. Any individuals in the parental data sets that had non-discrete values reported were not included for the purposes of compatibility with the Faroese sample.

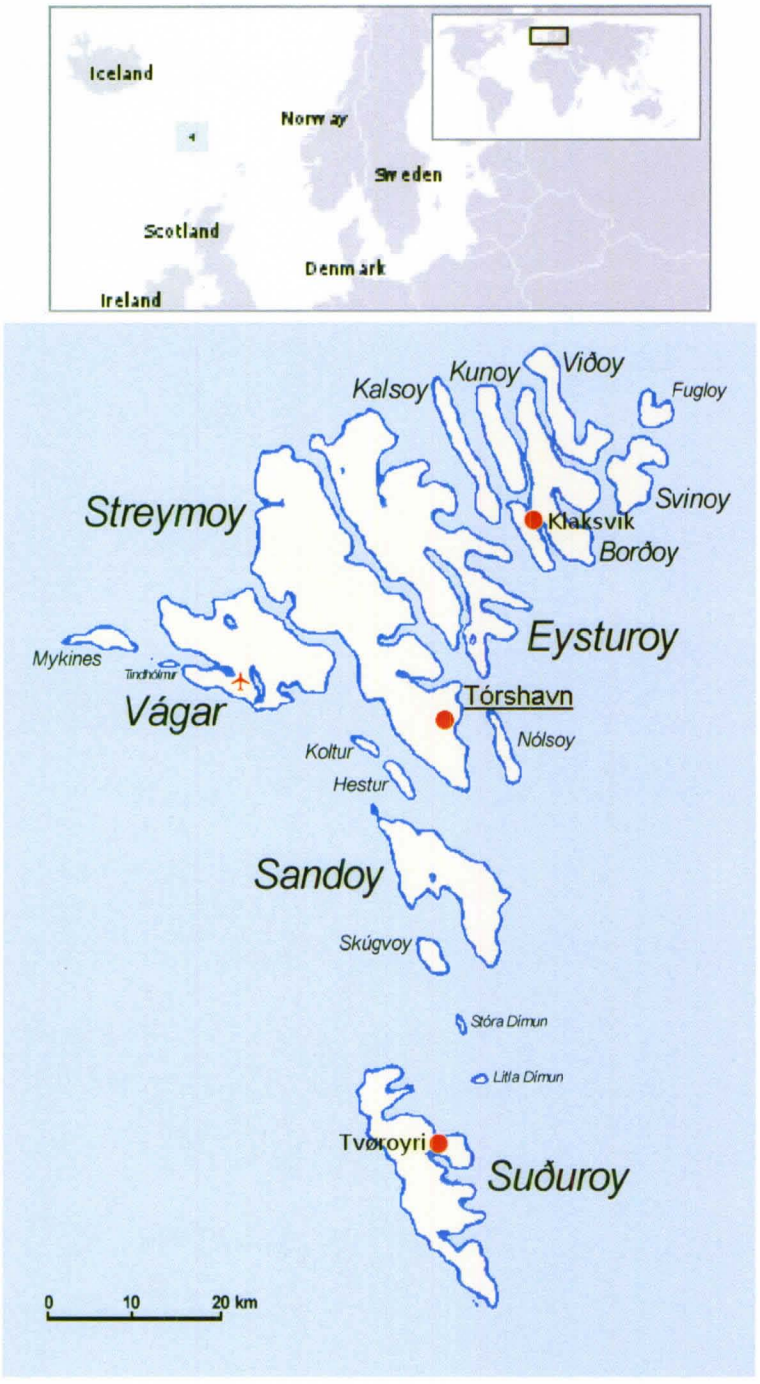

Figure 1. Maps of the Faroe Islands. Upper map shows the relationship of the Faroes (boxed in green) to other Northern European countries. Lower map shows island names and major cities. Original images under GPL. (Wikimedia, 2012)

Haplogroups were inferred from haplotype data using an allele-frequency Bayesian method (Athey, 2005), and nomenclature for haplogroups is that defined by the Y-Chromosome Consortium (YCC) (Consortium, 2002). For each population in this study, all available loci were used to determine haplogroup even if they were not used in further analyses. The online module (http://www.hprg.com/hapest5) allows the user to input specific marker values, returns a goodness of fit and probability score which assigns an individual haplotype to one of ten common European haplogroups (Athey, 2005, p. 1). Reliability of the Athey method was tested on the Norwegian and Swedish data, both of which had been typed for SNP haplogroup.

In cases where individual haplotypes did not fall into expected Scandinavian or British haplogroups, potential origins were determined using the Y-Chromosome Haplotype Reference Database (YHRD)(www.yhrd.org). Matching or neighbor haplotypes located by the YHRD were recorded for rarer Faroese haplotypes.

Each sample was tested for shared haplotypes between populations. Since all loci scored in the Faroese sample were not reported in all source samples a curtailed list of eight loci were used to determine identical haplotypes across populations. This method was also used to determine the proportion of unique haplotypes in each population.

Genetic distances (R_{st}, D_c) were calculated using the MICROSAT software package, (*MICROSAT Ver. 1.5*). R_{st} is a measure of genetic distance between populations that is suitable for microsatellite data since it takes into account both the higher mutation rates of these sequences (Slatkin, 1995, p. 457) and assumes a generalized stepwise mutation model (Slatkin, 1995, p. 461). R_{st}, however, tends to be biased when sample sizes are small (Ruzzante, 1998). Slatkin's R_{st} is calculated as: $R_{st}= \frac{\bar{s}-s_w}{\bar{s}}$ (Slatkin, 1995). Because R_{st} is sensitive to sample size, chord distances (D_c) (Cavalli-Sforza and Edwards, 1967) were also calculated to provide a comparative measure of genetic distance to Slatkin's R_{st}. The chord distance

measure is less sensitive to population size differences. This measure assumes that frequencies of alleles have changed via genetic drift alone as well as allowing the population to change in size over time (Jorgensen et al., 2004, p. 21). D_c is calculated as: $D_c = \frac{1}{2l}\sum_{j}^{l}(2(1-\sum_{i}^{mj}(x_{ij}-y_{ij})^{\frac{1}{2}}))^{\frac{1}{2}}$

Statistical significance of R_{st} and D_c values were determined by performing 1,000 bootstrap (Felsenstein, 1985) analyses using the MICROSAT software package. Bootstrapping re-samples the original data and estimates the distribution of the statistics (Goodman, 1997, p. 883). The mean and standard error were determined from the resulting R_{st} and D_c scores for each haplogroup and in the aggregate. 95% confidence intervals were calculated using bootstrapped means and standard error of each R_{st} and D_c value (See Appendix for C.I. scores).

Euclidean distance (metric) multidimensional scaling plots using R_{st} and D_c data were created with the R statistical package (R-project.org) to model the genetic dissimilarity between the Faroese and source populations (See Appendix for R scripts).

Gene diversity was calculated in the Faroese sample according to the allele frequency of a particular locus in a particular haplogroup. Gene diversity (or heterozygosity) is a measure of divergence used to determine distances within and between closely related populations (Nei, 1973, p. 3321). Gene diversity of the Y-chromosome has been used to demonstrate higher levels of diversity among African populations as compared to non-African populations, as well as support the notion of a rapid population expansion after the Pleistocene (Jorde et al., 2000, p. 983). In the same study it was found that Northern Europeans have exceptionally low gene diversity levels as compared to other European populations (Jorde et al., 2000, p. 981). It can therefore be assumed that gene diversity values for the Faroese will be very low. Gene diversity is calculated as: $1\text{-}\sum_{i=1}^{k} p_i^2$

Finally, phylogenetic relationships within the Faroese and between parental populations were inferred using the NETWORK 4.6.1.0 software package

(http://www.fluxus-engineering.com)(*Network 4.6.1.0. User Guide* 2011). Individual loci within median-joining networks (Bandelt, Forster, and Rohl, 1999) were weighted according to mutation rate (Table 2) using the following scheme.

Each of the 12 loci was given a weight ranging from zero to 11 based on its respective mutation rate. A decrease in weight (-1) correlates to each increase of 0.025% in mutation rate such that the most mutable loci will have little influence on the basal structure of the network (A score of zero essentially tells the Network software to ignore that locus). The loci DYS388, DYS426 and DYS438 were all scored as zero in the macro analyses between Faroese and parental populations because one or more were not reported in the Norse, Irish, Swedish, or Danish data sets. This allows for maximum comparability between data sets and decreases the likelihood for idiosyncrasies in the networks themselves caused by "not applicable" scores for loci. The remaining loci were weighted as follows: DYS19(4), DYS385a/b(5), DYS389I(4), DYS389b(1), DYS390(4), DYS391(2), DYS392(11), DYS393(10). In networks showing the relationship within the Faroese population, DYS438 was weighted (11) according to its fact sheet in the National Institute of Standards and Technology (NIST) Standard Reference Database (*NIST Standard Reference Database* 2012). Since no mutation rate is currently cited for DYS388 and DYS426 in the NIST Database, these two loci were weighted as (10) and (11) respectively, according to their estimated mutation rates reported in (Ravid-Amir and Rosset, 2010). Median-joining networks representing the Faroese and their parental populations were further processed using the maximum parsimony (MP) calculation option in the NETWORK software. This process determines the shortest possible trees from the original network (*Network 4.6.1.0. User Guide* 2011, p. 22). The MP algorithm facilitates the assessment of neighbor haplotype frequencies, time to most recent common ancestor estimation, and visual clarity of the network.

Mutational distances from calculated modal haplotypes were determined by counting the number of mutations separating each node in haplotype networks. In

TABLE 2

Individual loci mutation rates (probability that a mutation will occur in any given generation), weight, and reference

Loci	Weight	μ	Ref.
DYS19	4	0.25%	*NIST Standard Reference Database* 2012
DYS385a	5	0.21%	*NIST Standard Reference Database* 2012
DYS385b	5	0.21%	*NIST Standard Reference Database* 2012
DYS388	10	0.10%	Ravid-Amir and Rosset, 2010
DYS389I	4	0.24%	*NIST Standard Reference Database* 2012
DYS389b	1	0.35%	*NIST Standard Reference Database* 2012
DYS390	4	0.25%	*NIST Standard Reference Database* 2012
DYS391	2	0.28%	*NIST Standard Reference Database* 2012
DYS392	11	0.07%	*NIST Standard Reference Database* 2012
DYS393	10	0.08%	*NIST Standard Reference Database* 2012
DYS426	11	0.07%	Ravid-Amir and Rosset, 2010
DYS438	11	0.07%	*NIST Standard Reference Database* 2012

cases where no individual in the sample held the calculated modal haplotype, mutational distance was estimated using raw haplotype data: $\bar{x} = \mid \bar{x} \pm x \mid$

Estimates for the time to most recent common ancestor (TMRCA) between resulting modal and satellite nodes within median-joining networks were calculated using the optional time estimation algorithm included in network software package. These approximate time estimates were determined using the mean effective mutation rate for Y-chromosome microsatellite loci as reported by (Zhivotovsky et al., 2004) (approximately 1 mutation per 2,131 years).

CHAPTER V

RESULTS

Y-chromosomal haplogroup and haplotypes tend to be more highly variable among mainland European samples than those from islands (Table 5). The lowest levels of inner haplotype variability are found in the Faroese and Icelanders while the Irish sample is the most constrained at the haplogroup level (Table 5). The eight STR markers reported in all samples formed a total of 572 distinct haplotypes. A total of 11 haplogroups were found in the Faroese sample, nine of which were also found in one or more source population.

Multidimensional scaling plots of estimated distances (R_{st}, D_c) between populations by haplogroup are found in Figures 3 and 2. Because of the extreme low frequency of R1a in the Irish sample (n=2) it was excluded from the R1a R_{st} MDS plot. Since inclusion in the R1a MDS plot obscured the relationship between the other populations, this allowed for better modeling of the relationship between the remaining samples. Distances between aggregate R_{st} values illustrate a cluster of Faroese, Irish and Danish haplotypes with some divergence of Icelandic, Swedish and Norwegians. Among R1a R_{st} estimates there is a clear cluster of Icelandic and Norwegians. Haplogroup R1b R_{st} measures show a clustering of Irish and Danish haplotypes and a closer relationship between Swedish and Norwegian R1b haplotypes. Faroese I1 R_{st} values have a close relationship to those found in the Danish and Swedish sample. Chord distances in the aggregate exhibit a clustering of Swedish and Norwegian samples. In R1a a cluster of Swedish and Faroese as well as Danish and Norse is evident, while in R1b a cluster of Irish and Danish and a closer relationship between the Faroes and Norway is apparent. Chord distances in the I1

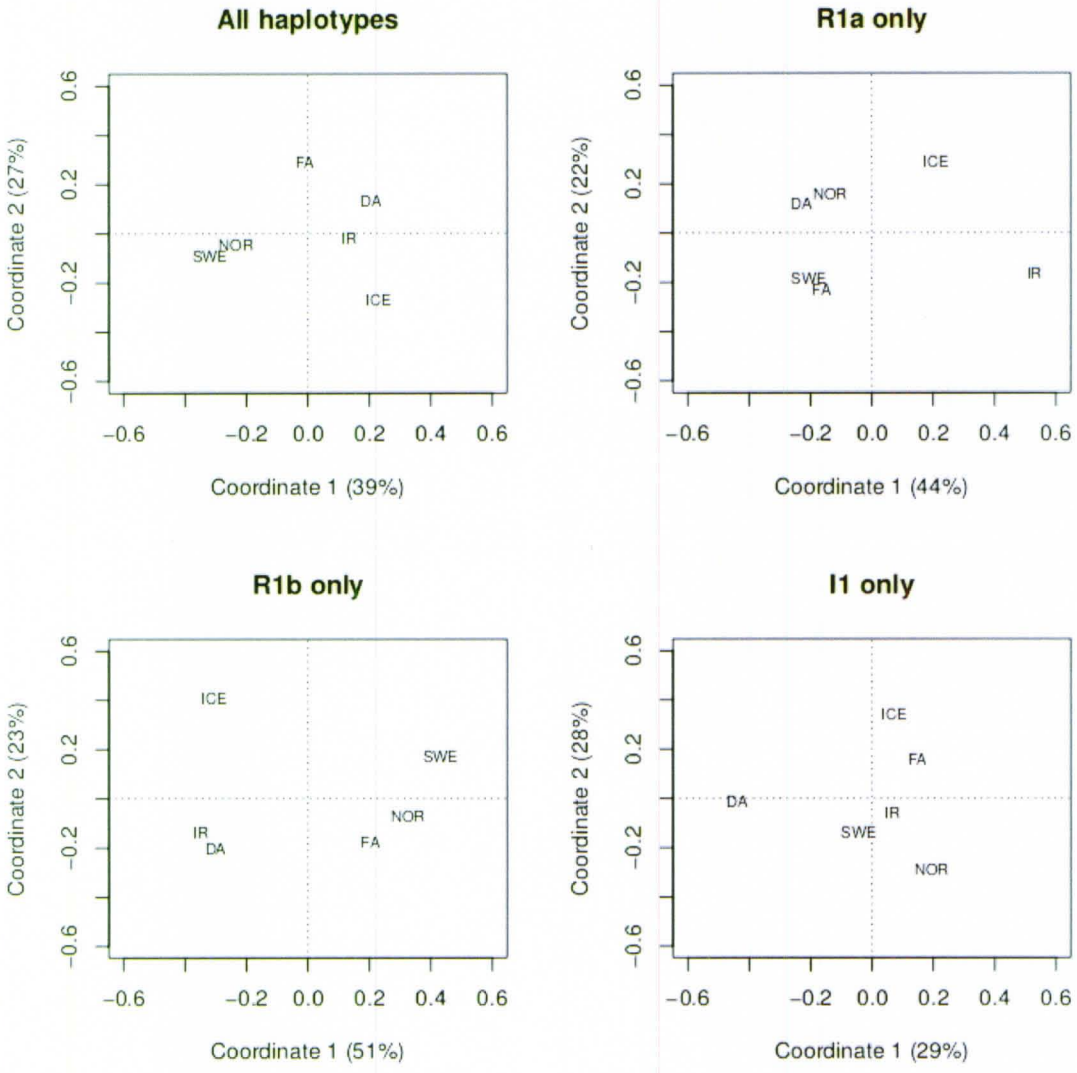

Figure 2. Multidimensional scaling plots of D_c values. FA:Faroes, DA:Denmark, ICE:Iceland, IR:Ireland, NOR:Norway, SWE:Sweden

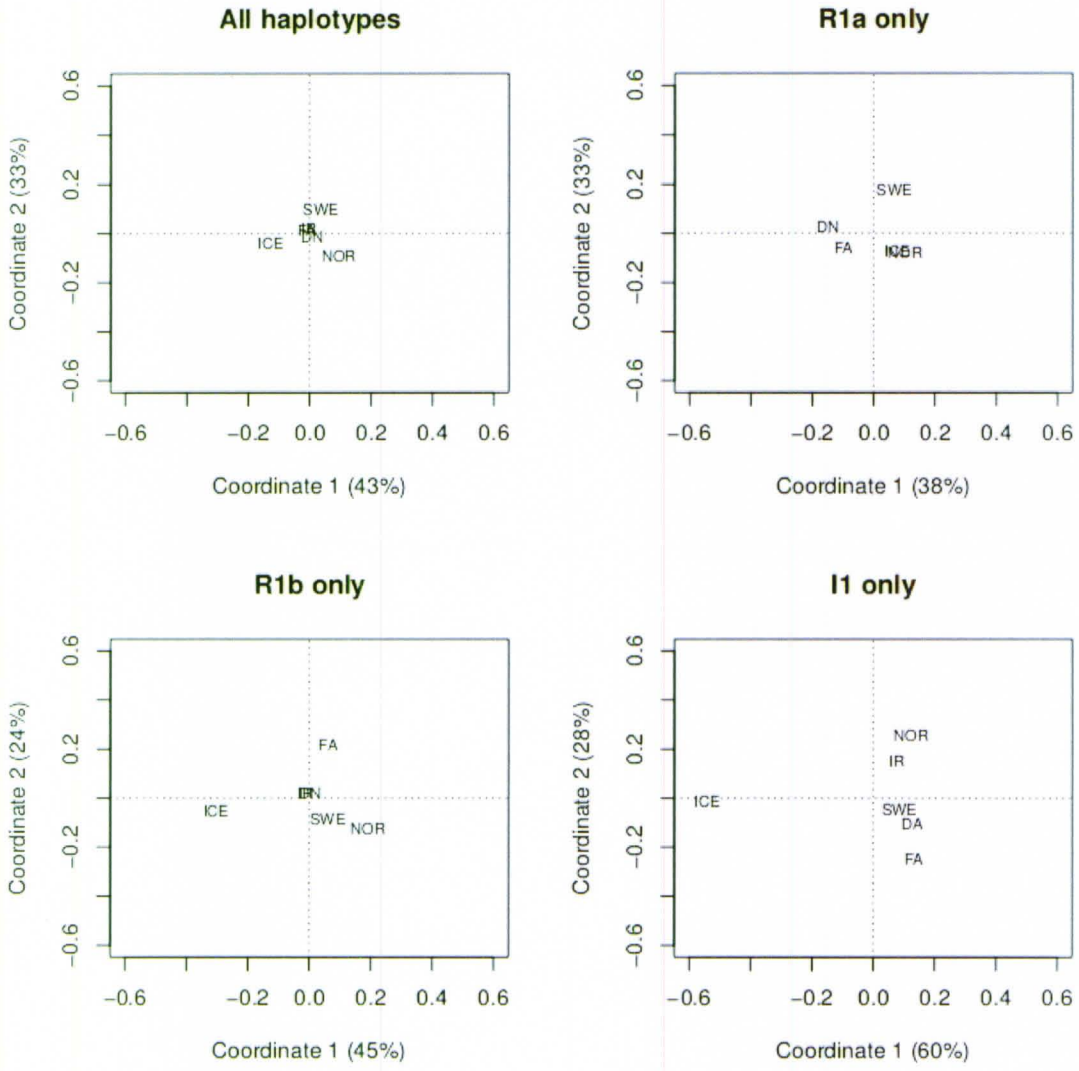

Figure 3. Multidimensional scaling plots of R_{st} values. FA:Faroes, DA:Denmark, ICE:Iceland, IR:Ireland, NOR:Norway, SWE:Sweden

haplogroup suggest a somewhat closer relationship between Iceland and the Faroes.

Calculated 95% confidence intervals for aggregate haplotypes show no statistical significance between the majority of populations ($0 \in$ CI). Those R_{st} values that returned confidence intervals that did not include 0 were negative and therefore should be interpreted cautiously. Additionally, the percent of variance explained by the two visible coordinates in each graph is fairly low except in the case of I1's horizontal distance. Results from D_c genetic distances all had statistically significant confidence intervals but again had very low percents of variance explained by both coordinates.

Haplogroup data: The most common haplogroups in the Faroese sample — R1a, R1b, and I1 — constitute 42%, 25%, and 21%, of the data set, respectively. Rare haplogroups are also found in the sample at low frequencies. Haplogroup J1 constitutes 4%, Q 3%, E1b1b 1%, and I2b1, I2b(xI2b1), I2a(xI2a1), L, and N combined make up 5% of the total data set (one individual per haplogroup) (Table 3) . Rare haplogroups found in source populations include G2a, J2b, J2a1b, and I2a1 (data not shown). Of the rare haplogroups found in the Faroese, six were found in one or more source population. Haplogroup E1b1b is at its highest frequency in the Icelandic sample (7%), Q in the Faroese and Swedish (3%), I2b1 in the Norse (6%), and I2a(xI2a1) in the Irish (3%). Faroese haplogroups J1 and L were not found in any of the source populations. Comparisons of Faroese haplogroup frequencies by population are found in Figure 4. While there are more rare haplogroups in the sample collected from the northern city of Klaksvik, no clear pattern emerges when haplogroup frequencies are compared by sampling location. Haplogroup frequencies by sampling location are found in Table 4.

Reliability tests of the Athey method of haplogroup inference were performed against the SNP-typed Norwegian and Swedish samples. All Swedish samples were placed in the correct haplogroup by the program. All Norwegian samples also scored

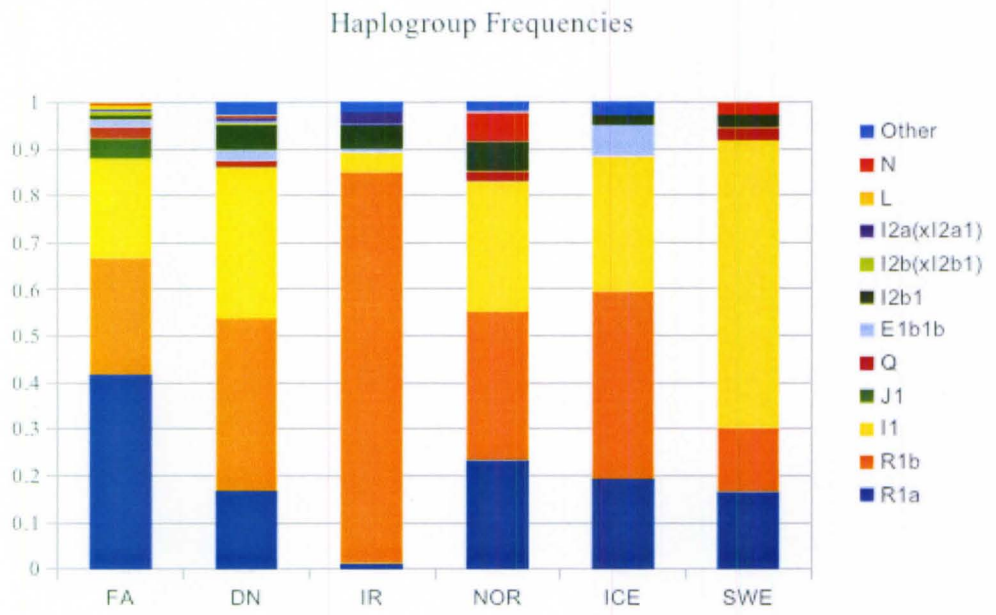

Figure 4. Comparison of haplogroup frequencies in all populations.

TABLE 3

Haplogroup Frequencies

Pop	R1a	R1b	I1	J1	Q	E1b1b	I2b1	I2b	I2a	L	N	Other
Faroes	0.42	0.25	0.21	0.04	0.03	0.01	0.01	0.01	0.01	0.01	0.01	n/a
Denmark	0.17	0.37	0.32	0.00	0.02	0.02	0.05	0.01	0.01	0.00	0.01	0.03
Ireland	0.01	0.84	0.04	0.00	0.00	0.01	0.05	0.00	0.03	0.00	0.00	0.02
Iceland	0.19	0.40	0.29	0	0	0.07	0.02	0	0	0	0	0.03
Norway	0.23	0.32	0.28	0.00	0.02	0.00	0.06	0.00	0.00	0.00	0.06	0.02
Sweden	0.17	0.14	0.61	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.03	0.00

TABLE 4

Haplogroup frequencies by sampling location

Loc	R1a	R1b	I1	J1	Q	E1b1b	I2b1	I2b	I2a	L	N
Tvoroyri	0.48	0.24	0.22	0.04	0.02	-	-	-	-	-	-
Klaksvik	0.37	0.27	0.15	0.10	0.02	0.05	-	0.02	0.02	-	-
Tórshavn	0.42	0.23	0.25	-	0.04	-	0.02	-	-	0.02	0.02

correctly save one SNP-tested R1a haplotype which earned a 62.3% probability for R1b and a 37.7% probability for R1a. For all further analyses this sample was grouped with R1a haplotypes.

Haplotype data: Summary statistics of all populations are found in Table 5. Percentages of unique haplotypes (k/n) are highest among the Danish, Irish, Norse and Swedish samples (88%, 81%, 85%, and 83%). The Danish, Norse, and Swedish samples also have the highest proportion of private haplotypes (i.e., they are unique to the data set). In both cases, the Faroese and Icelandic samples have more constrained diversity as evidenced by their inner haplotype variability. 48% of the Faroese and 55% of the Icelandic samples are comprised of unique haplotypes. Private haplotypes constitute 31% of the Faroese data set and 50% of the Icelandic data set. While the Irish sample has a high proportion of unique haplotypes (81%), the sample has a lower number of private haplotypes (45%). This pattern may be due to the unique demographic processes that characterize the Irish population. In particular, the population is at the end position of a cline of R1b across Europe and is thought to have had little migration since the island was founded (Moore et al., 2006, p. 334).

Haplotype diversity, as measured by proportion of private haplotypes, within the most common Faroese haplogroups is nearly identical for R1a and I1 (24%) while R1b is significantly higher (40%). Of the total number of haplotypes found in the Faroese sample, 26% are shared with one or more source population. Four

TABLE 5

Haplotype statistics

Population	n	k	k/n	%private
Faroes	139	68	48.92	30.93
Danish	185	163	88.10	60.54
Irish	148	120	81.08	45.27
Norse	47	40	85.10	74.46
Icelandic	181	100	55.24	49.72
Swedish	36	30	83.33	72.22

haplotypes were shared exclusively between the Faroes and Norway, four between the Faroes and Ireland, and five between the Faroes and Denmark. One haplotype was shared exclusively between the Faroes and Sweden.

Twenty-three Faroese individuals shared haplotypes with two or more Scandinavian populations — Denmark, Norway, and Sweden — while four individuals shared haplotypes common to Scandinavian and Irish samples (Table 6). While the Icelandic sample shared haplotypes with Denmark and Norway, no haplotypes were shared between the Faroes and Iceland.

Network data: The phylogenetic relationship between all Faroese haplotypes is found in Figure 5. Clear differentiation by haplogroup is present except in the case of the lone N haplotype which is included in the R1b haplogroup branch. This is possibly the result of haplogroup misassignment, a function of the network algorithm, or could be due to an error in the raw data itself. Overall cohesive haplotype branches representing different haplogroups support the accuracy of Athey's method.

TABLE 6

Faroese haplotypes that are shared with one or more parental population. ID is haplotype identifier. Number under parental population column represents how many individuals from each group share haplotypes with the Faroese. Percentage denotes the frequency of the haplogroup in the Faroese.

ID	n	Danish	Norse	Irish	Icelandic	Swedish	Haplogroup	%
59	3	-	1	-	-	-	Q	0.02
62	1	2	-	-	-	-	E1b1b	0.01
34	1	1	-	-	-	-	R1b	0.01
20	8	2	1	-	-	-	R1b	0.06
22	2	-	-	5	-	-	R1b	0.01
23	2	1	-	2	-	-	R1b	0.01
33	1	1	-	4	-	-	R1b	0.01
39	1	1	1	-	-	-	R1b	0.01
26	1	-	-	2	-	-	R1b	0.01
31	1	-	-	1	-	-	R1b	0.01
51	1	3	-	2	-	1	I1	0.01
48	2	3	1	-	-	-	I1	0.01
43	6	5	-	-	-	2	I1	0.04
44	5	6	4	-	-	1	I1	0.04
46	1	-	-	-	-	1	I1	0.01
24	2	-	1	-	-	-	R1b	0.01
9	2	1	-	-	-	-	R1a	0.01
2	10	-	-	1	-	-	R1a	0.07
1	14	2	-	-	-	-	R1a	0.10
5	4	-	2	-	-	-	R1a	0.03
47	2	-	1	-	-	-	I1	0.01
16	1	1	-	-	-	-	R1a	0.01

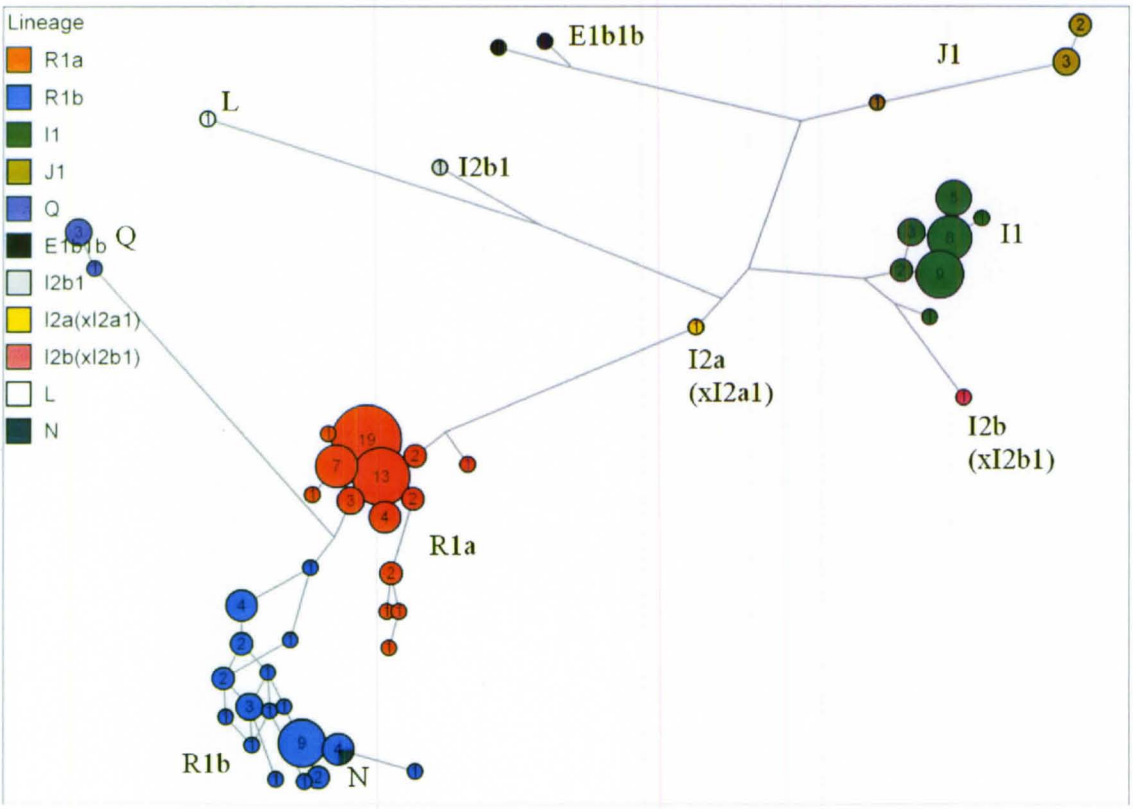

Figure 5. Simplified phylogenetic network of all Faroese haplotypes. Red:R1a, Blue:R1b, Green:I1, Orange:J1, Gray:E1b1b, Pink:I2b(xI2b1), Purple:I2a(xI2a1), Brown:I2b1, White:L, Black:Q, Dark Green:N

Population specific networks by common haplogroups — R1a, R1b, and I1 — are found in Figures 6, 7, and 8, the distributions of which are dependent on sample size and individual demographic histories. Generally, clustering of haplotypes around modal nodes is found among the Faroese and Icelandic networks, especially in haplogroups R1a and I1. Wider distributions are found among Scandinavian populations although small sample sizes render some individual networks less useful for analysis. The Faroese specific network representing haplogroup R1b is more diverse than R1a and I1. Specifically, there are more individuals at distal branches further from the mode, fewer large clusters of haplotypes, and a higher proportion of private haplotypes in the network.

Population specific networks by common haplogroups — R1a, R1b, and I1 — are found in Figures 6, 7, and 8, the distributions of which are dependent on sample size and individual demographic histories. Generally, clustering of haplotypes around modal nodes is found among the Faroese and Icelandic networks, especially in haplogroups R1a and I1. Wider distributions are found among Scandinavian populations although small sample sizes render some individual networks less useful for analysis. The Faroese specific network representing haplogroup R1b is more diverse than R1a and I1. Specifically, there are more individuals at distal branches further from the mode, fewer large clusters of haplotypes, and a higher proportion of private haplotypes in the network.

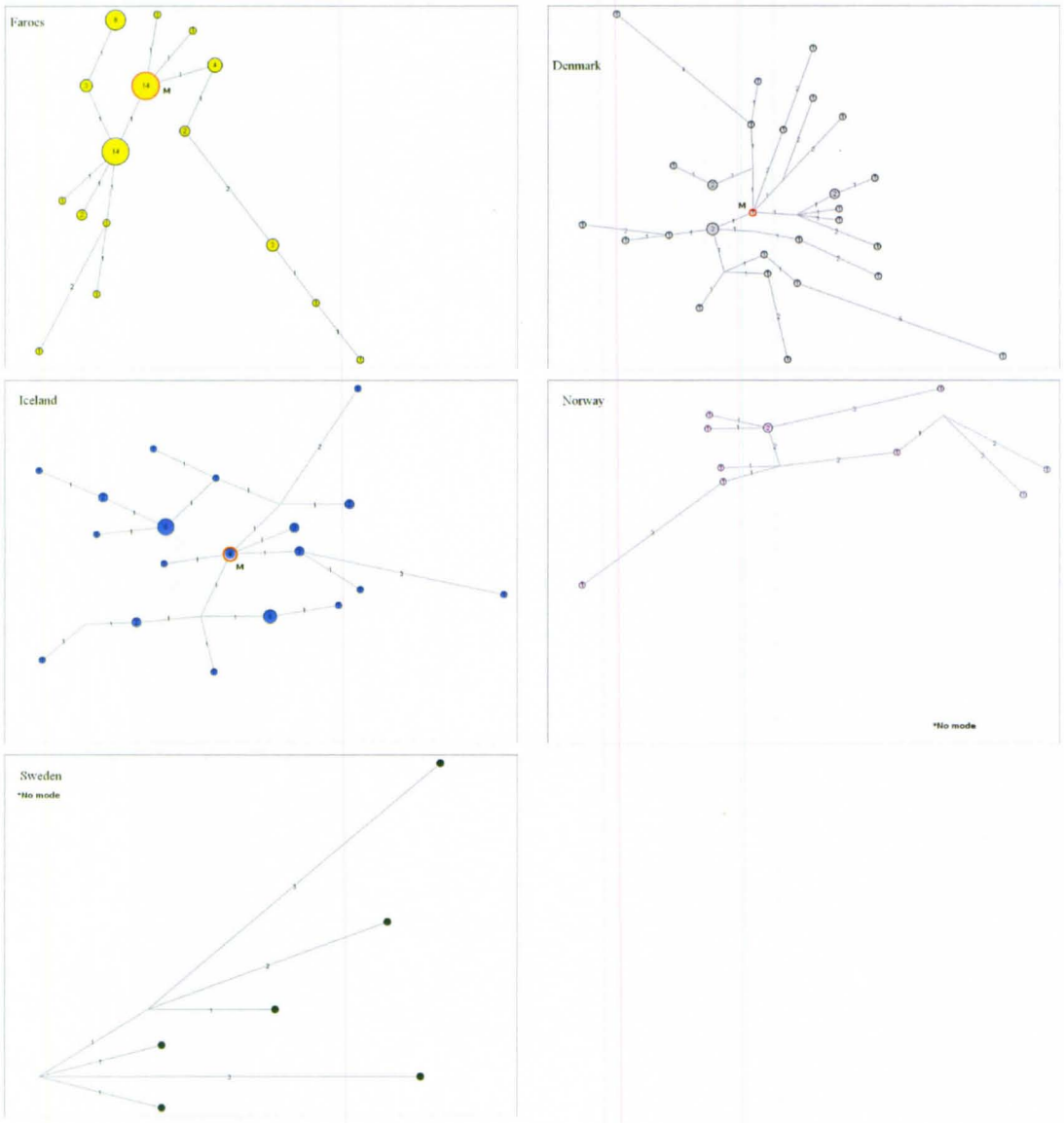

Figure 6. Median-joining network of R1a individuals. Faroes(yellow), Denmark(gray), Iceland(blue), Norway(pink) Sweden(green). Size of node is relative to number of individuals, length of branch between nodes is relative to number of mutational steps away. Modal node ("M") is circled in red.

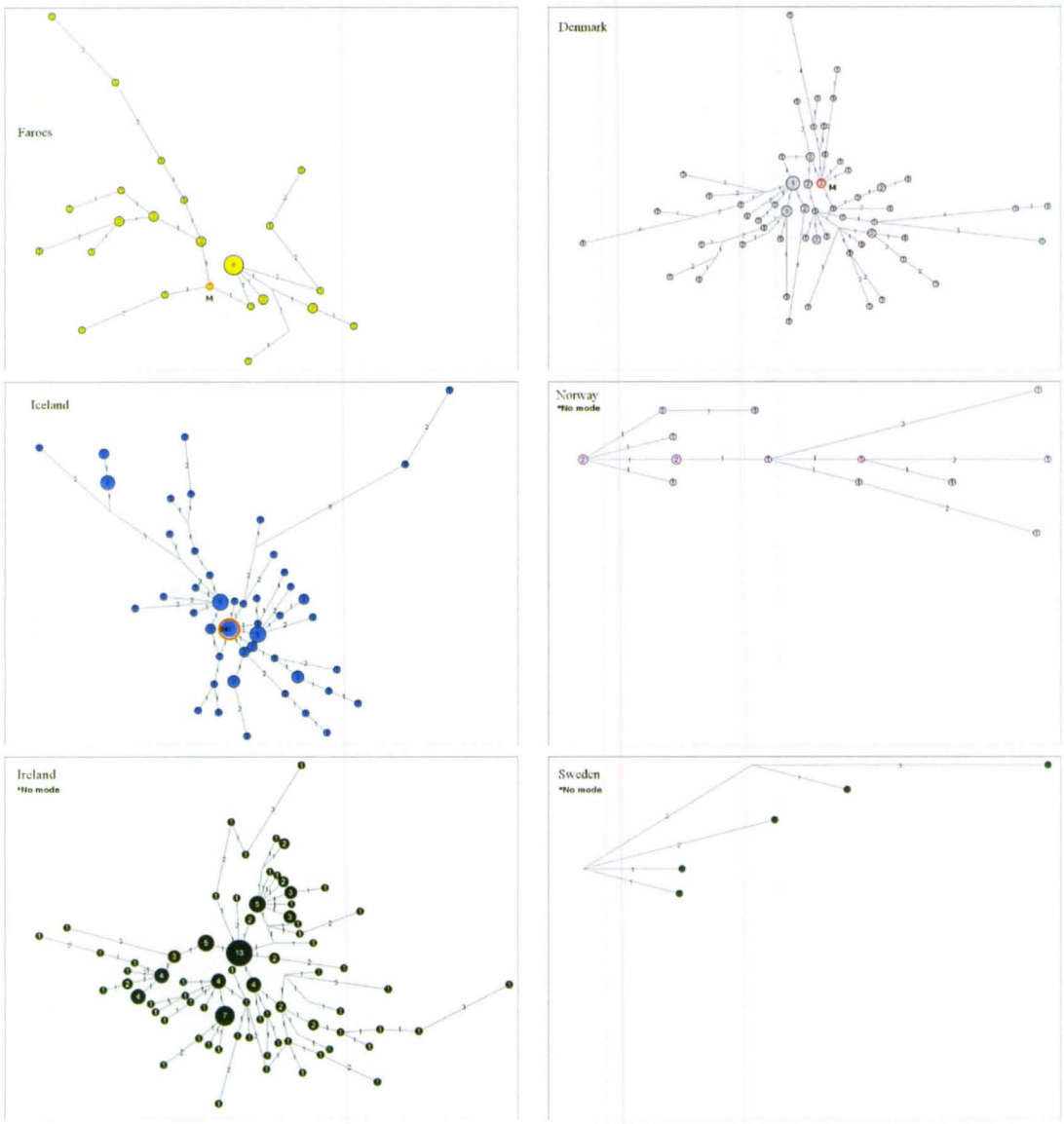

Figure 7. Median-joining network of R1b individuals. Faroes(yellow), Denmark(gray), Iceland(blue), Norway(pink), Ireland(black) and Sweden(green). Size of node is relative to number of individuals, length of branch between nodes is relative to number of mutational steps away. Modal node ("M") is circled in red.

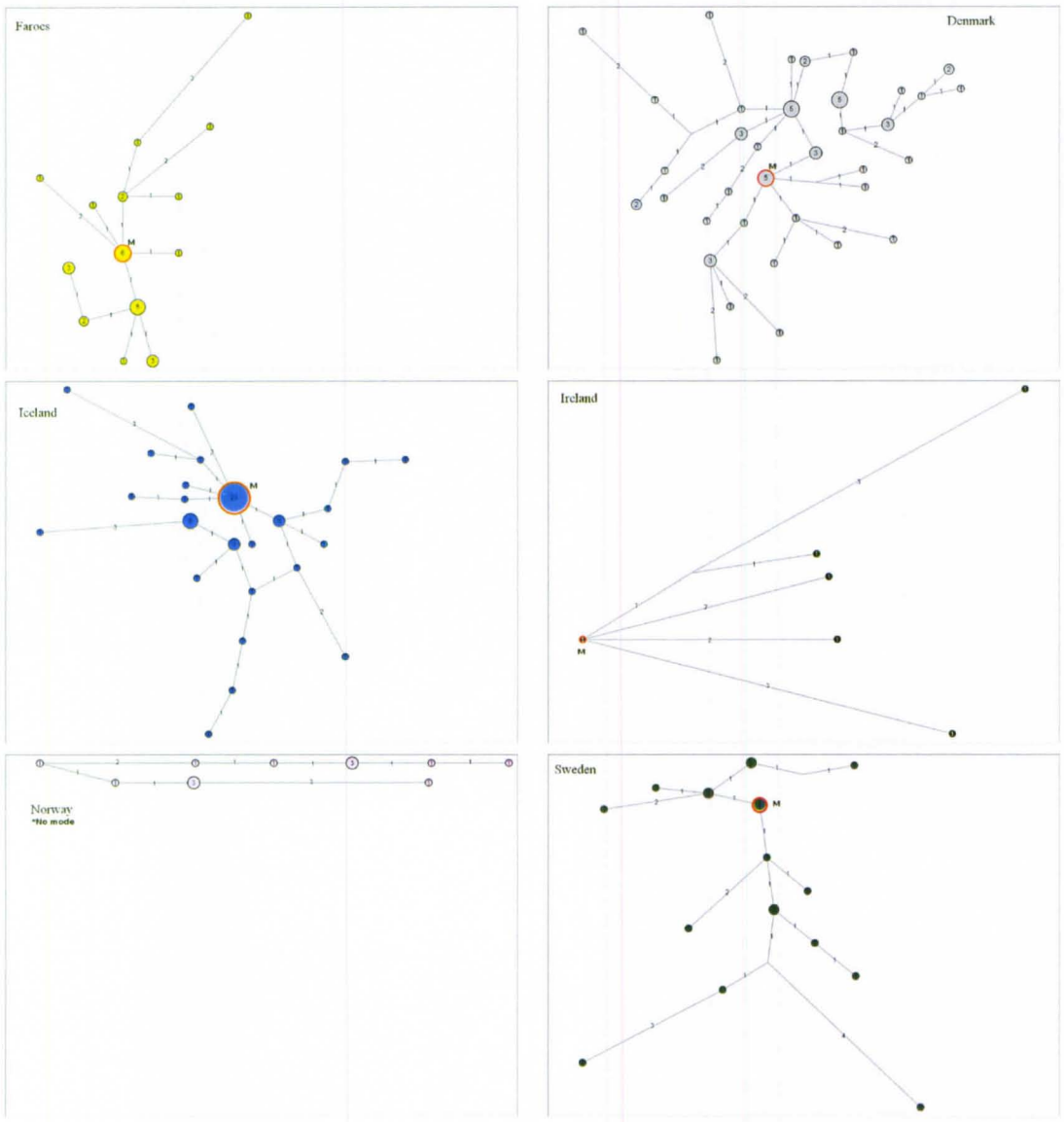

Figure 8. Median-joining network of I1 individuals. Faroes(yellow), Denmark(gray), Iceland(blue), Norway(pink), Sweden(green) and Ireland(black). Size of node is relative to number of individuals, length of branch between nodes is relative to number of mutational steps away. Modal node ("M") is circled in red.

Aggregate median-joining networks representing all samples by common haplogroups are found in Figure 9. Faroese samples are fairly dispersed in all three networks. Large Faroese nodes tend to be located proximally within the network while private haplotypes are located more distally throughout. Some clustering of source populations is evident, but less so than the Faroese or Icelandic samples. Faroese and Icelandic clusters are distinct. While Faroese haplotypes tend to be located closer to the aggregate mode for each haplogroup, Icelandic haplotypes consistently cluster on a cohesive separate branch. The pattern established by Icelandic haplotypes in these aggregate networks is unique to that population.

In all three aggregate networks, specific Faroese nodes are shared by one or more source populations. R1a and R1b have shared nodes with Norway, Denmark and Ireland, while I1 has shared nodes with Denmark, Norway, Ireland and Sweden.

Rare haplogroups in the Faroese and source populations have too few samples for construction of robust networks and were therefore excluded.

Neighbor haplotype counts: Histograms representing stepwise mutational distances from calculated modes are found in Figures 10, 11, and 12. Different distribution patterns around modal haplotypes are indicative of various demographic processes including population size, age, and the success of individual lineages. In the case of recent founder effect we might expect the majority of mutational differences from the modal haplotype to be less than or equal to two (Table 7). The impact of time since founding, population age, or admixture will be to broaden the distribution. Interpretation of histograms of mutational distances must consider the shape and nature of the distribution of samples on the networks.

Two basic patterns of distributions of individuals from the modal haplotype are apparent, substantial clusters and widely dispersed individuals. Putative source populations — Denmark, Ireland, Norway, and Sweden — tend to have different dispersal patterns than the Faroes and Iceland, the individuals of which are

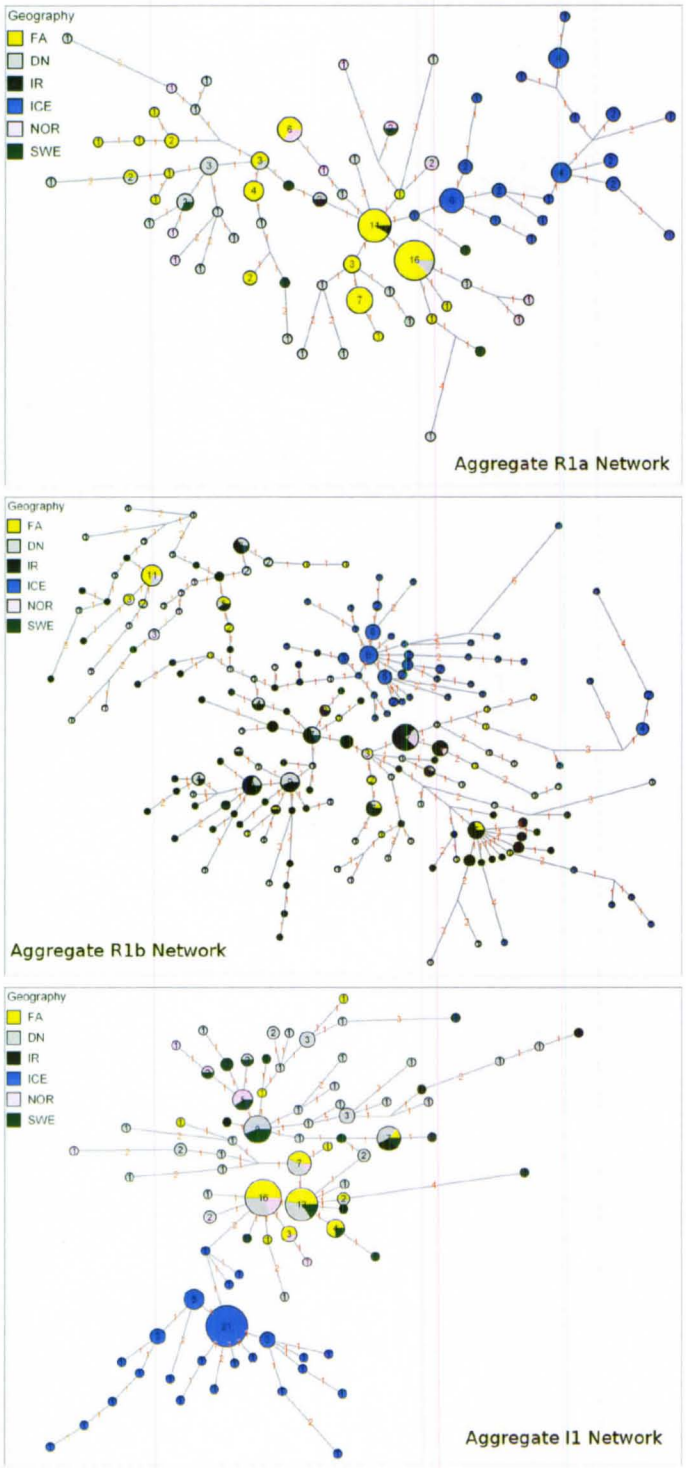

Figure 9. Aggregate haplogroup networks. From top to bottom: R1a, R1b, I1. Yellow:Faroes, Grey:Denmark, Black:Ireland, Blue:Iceland, Pink:Norway, Green:Sweden.

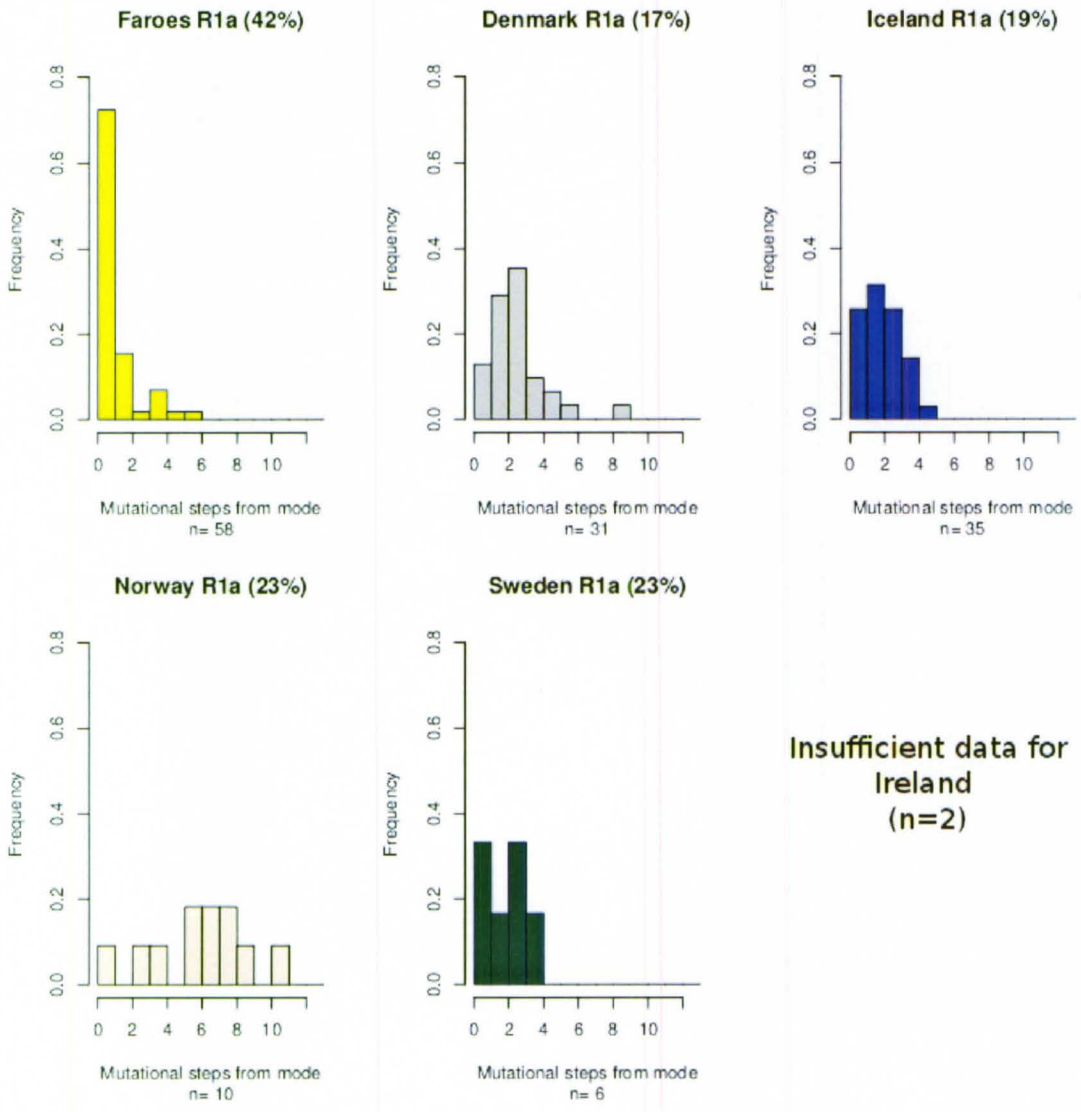

Figure 10. Number of mutational steps away from modal haplotypes in each population.

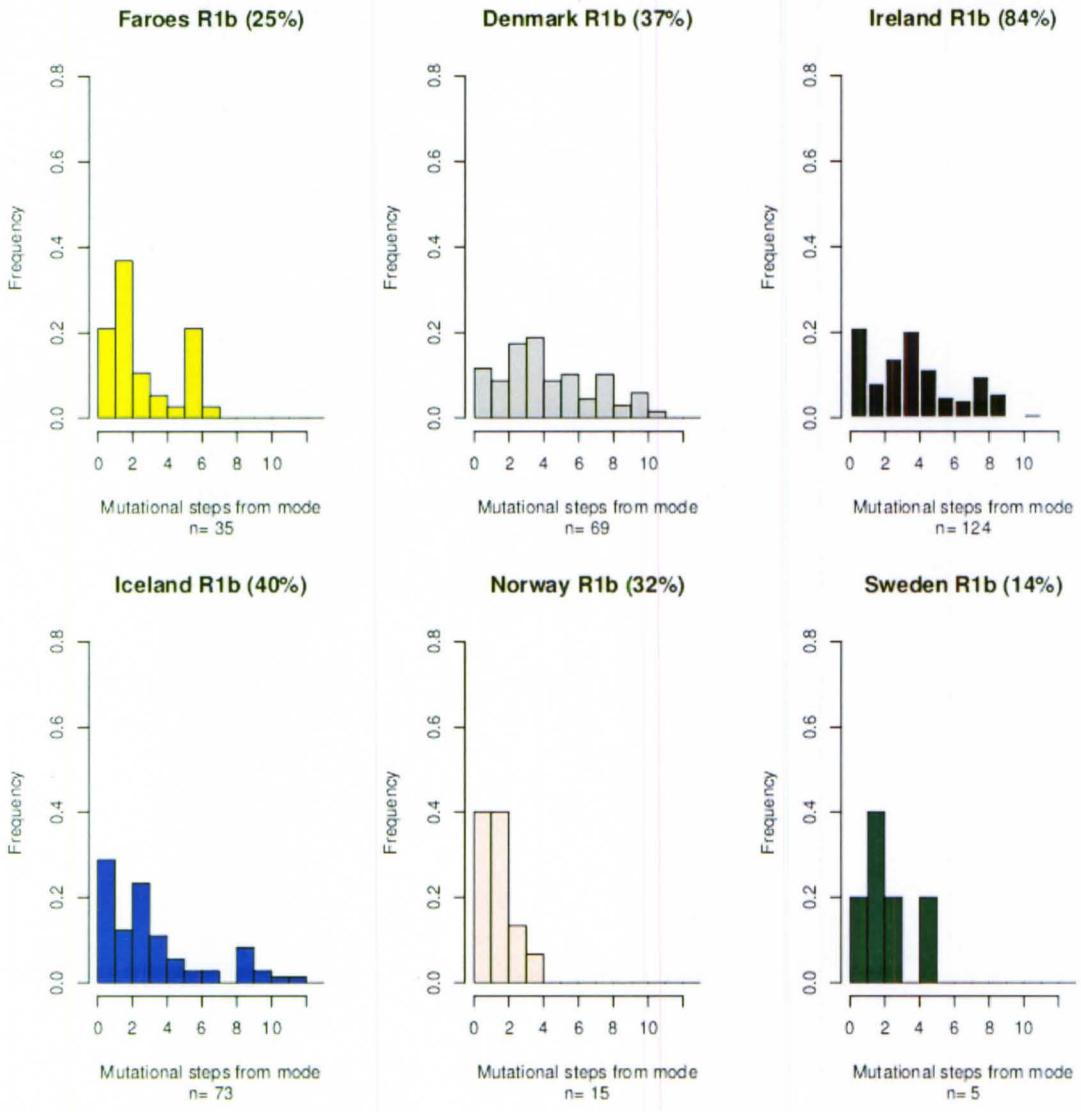

Figure 11. Number of mutational steps away from modal haplotypes in each population.

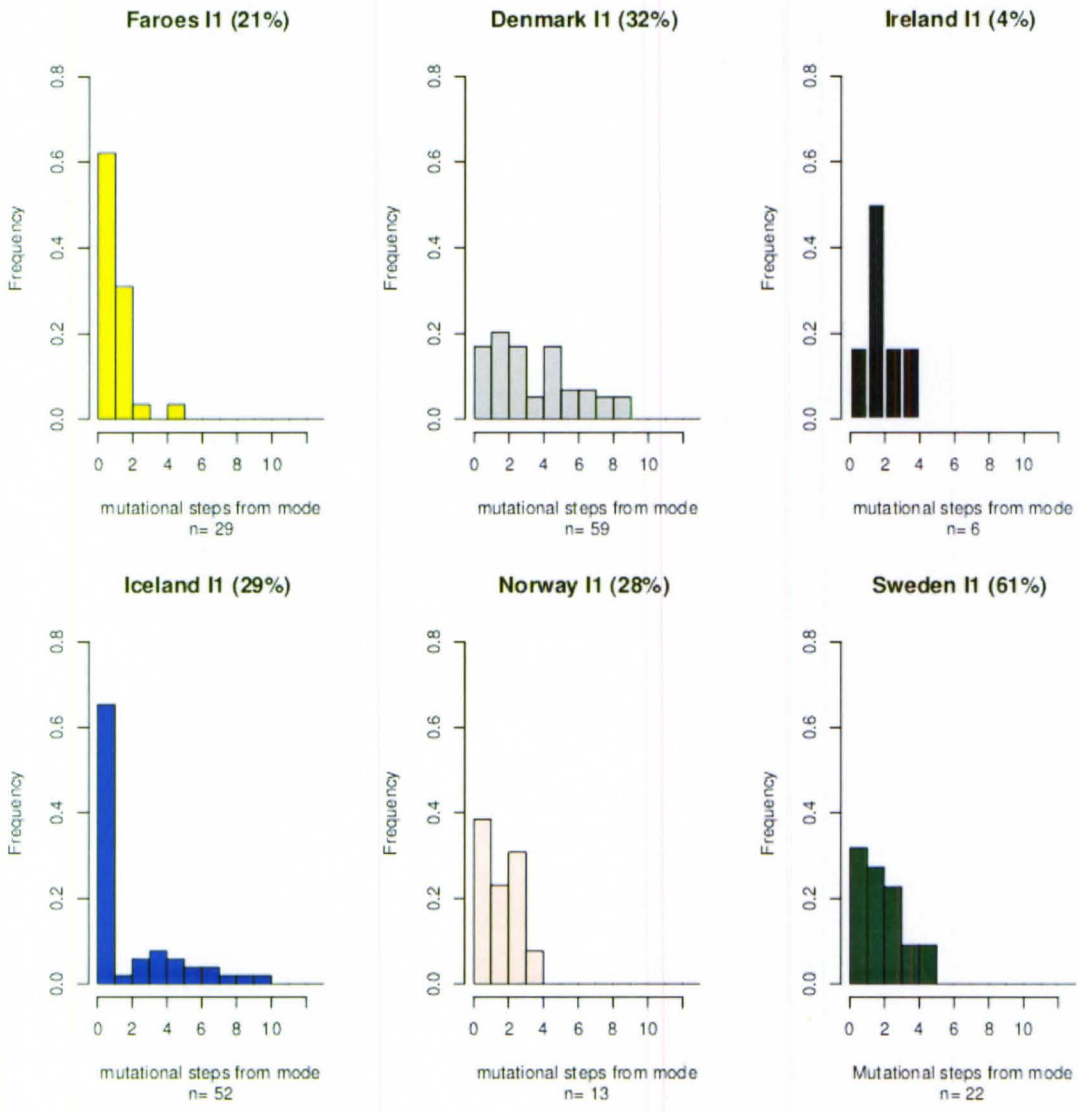

Figure 12. Number of mutational steps away from modal haplotypes in each population.

TABLE 7

Percent of individuals located $\leq$2 steps from the modal haplotype

Pop	R1a	R1b	I1
Faroe	87.93	62.86	93.10
Denmark	41.94	20.29	37.29
Ireland	-	29.03	66.67
Iceland	57.14	41.10	67.31
Norway	10.00	80.00	61.54
Sweden	50.00	60.00	59.09

clustered. Individuals in the Faroese and Icelandic samples are tight to the modal haplotype while histograms of source populations reveal the accumulation of divergent haplotypes. In general, source populations have broader distributions, notwithstanding sample size and the effect of latent population structure.

Time to most recent common ancestor: General patterns of TMRCA reflect the distribution of haplotypes in individual and aggregate networks with more broadly dispersed haplogroups scoring older TMRCA estimates and those with tight distributions estimated to be much younger. TMRCA estimates are listed in Table 8. TMRCA of haplogroups R1a and I1 in the Faroes are highly similar. Both have coalescent time estimates at 2,792 years $\pm$ 1,023 and 1,225 respectively. Estimates of TMRCA for R1b haplotypes in the Faroese are significantly older at 6,575 years $\pm$ 1,922.

Locus data: Variance among loci within the most common Faroese haplogroups are listed in Table 9. Among the quickly mutating loci ($\mu \geq 20\%$) some variation is expected because of their tendency towards rapid evolution. Slower mutating loci, on the other hand, should be more stable from one generation to the next. Variation among these loci reveals information on population substructure, age and admixture rates.

TABLE 8

Time to most recent common ancestor in each of the most common haplogroups (*Network 4.6.1.0. User Guide* 2011)

Haplogroup R1a	Faroes	Denmark	Iceland	Ireland	Norway	Sweden
Age in mutations	1.31	2.94	2.26	-	5.82	-
Age in years	2792.34	6255.52	4809.97	-	12398.55	-
σ	0.57	0.65	0.72	-	1.85	-
σ in years	1225.20	1386.82	1528.22	-	3941.77	-
Haplogroup R1b						
Age in mutations	3.09	4.55	2.78	4.10	2.07	-
Age in years	6575.66	9697.59	5925.93	8730.23	4404.07	-
σ	0.90	1.08	0.50	0.87	0.71	-
σ in years	1922.49	2301.02	1069.39	1859.37	1510.19	-
Haplogroup I1						
Age in mutations	1.31	3.80	1.79	2.17	-	2.29
Age in years	2792.34	8090.58	3811.21	4617.17	-	4870.86
σ	0.48	1.12	0.54	0.65	-	0.71
σ in years	1023.50	2386.57	1148.92	1375.55	-	1508.55

TABLE 9

Variance by loci within main Faroes haplogroups

Loci	μ	R1a	R1b	I1
DYS389b	0.35	0.35	0.35	0.00
DYS391	0.28	0.15	0.29	0.06
DYS19	0.25	0.10	0.08	0.16
DYS390	0.25	0.26	0.38	0.21
DYS389I	0.24	0.10	0.13	0.03
DYS385a	0.21	0.08	0.00	0.13
DYS385b	0.21	0.07	0.36	0.25
DYS388	0.10	0.00	0.00	0.13
DYS393	0.08	0.01	0.02	0.03
DYS392	0.07	0.01	0.46	0.00
DYS426	0.07	0.01	0.05	0.00
DYS438	0.07	0.00	0.00	0.09

TABLE 10

Gene Diversity ($1-\sum_{i=1}^{k} p_i^2$) by locus in each of the main Faroese haplogroups

Locus	R1a	R1b	I1
DYS19	0.099	0.157	0.328
DYS385A	0.158	0	0.245
DYS385B	0.131	0.552	0.485
DYS388	0	0	0.666
DYS389I	0.188	0.251	0.666
DYS390	0.432	0.545	0.380
DYS391	0.307	0.513	0.128
DYS392	0.034	0.340	0
DYS393	0.034	0.056	0.666
DYS426	0.034	0.108	0
DYS438	0	0	0.185
DYS389B	0.549	0.460	0

Noticeable differences between the three common haplogroups exist in loci DYS385a, DYS385b, and DYS391 among the Faroese. In particular, variance of the DYS392 locus is inconsistent across the three most common haplogroups. DYS392 has a relatively slow mutation rate (0.07%) and has high variability in the R1b haplogroup (0.46) while there is no or very little variability in R1a and I1. It is important to note that while the Icelandic and Irish sample had zero variability by haplogroup among certain loci, none of the mainland Scandinavian groups did (data not shown).

Gene diversity (D) scores can be found in Table 10. Loci with high gene diversity scores ($D \geq 0.5$) are also variable by haplogroup. Of particular interest are loci DYS393 and DYS388 in haplogroup I1, both of which have relatively slow mutation rates, high D, and score D=0 in the other two groups.

CHAPTER VI

DISCUSSION

Previous genetic studies of the Faroe Islands have found the population to be highly homogeneous compared to mainland Europeans (Als et al., 2006, p. 503)(Jorgensen et al., 2002, p. 385). This a a function of strong genetic drift. Because of the effects of drift on the Faroese Y-chromosome, it is difficult to do direct haplotype frequency comparisons between the Faroese and their potential parental populations (Jorgensen et al., 2004, p. 19). The state of haplogroups, haplotypes and individual loci on the NRY of a particular population predominantly depends on their demographic history and the time since that population split from the larger group. This study uses haplotype data from the Faroe Islands, Iceland, Norway, Ireland, Sweden and Denmark to distinguish between Y-chromosomal signatures of ancient and recent population history events. Based on historical data, three main hypotheses are proposed to explain the haplogroup and haplotype diversity found in the Faroes: 1) the majority of the Faroese samples is highly homogeneous as a function of strong genetic drift; 2) admixture from other regional European populations is evident; and, 3) there is evidence of admixture from populations outside of the region.

Homogeneity: Haplotype diversity and percentages of unique haplotypes within mainland European populations–Norway, Sweden, and Denmark–have considerably higher percentages of private haplotypes than those found in the island populations–Faroese, Icelandic, and Irish. The most constrained genetic diversity is consistently found in the Faroese. Only 30% of the Faroese sample consisted of

private haplotypes while approximately 70% of individuals shared their full haplotype with one or more person (Table 5). This is in stark contrast to the mainland Scandinavian samples of which private haplotypes constituted more than 60% of the population. A high number of private haplotypes is a signature of a heterogeneous, and often large, population with a higher likelihood of gene flow. Differences in the proportion of private haplotypes in different populations can be indicative of: (1) the amount of migration or isolation in a population, (2) missing data about populations that may have contributed to the genetic diversity of the study population; and, (3) a more thorough sampling of one population versus another (Helgason et al., 2001, p. 729). The relatively low number of private haplotypes in the Faroes is most likely due to genetic drifting and allele fixation within the Faroese population's more common haplogroups. Genetic drift in isolated populations with small founder sizes decreases heterozygosity within the population and increases the differentiation between populations (Chiaroni, Underhill, and Cavalli-Sforza, 2009, p. 20174). Therefore in the Faroese sample a loss of heterozygosity would be expected. Gene diversity scores for loci in the R1a haplogroup are low (Table 10). This is consistent with the signal of an early and small founder event of this haplotype in the Faroes followed by genetic drift.

Haplogroup results reveal an interesting trend in the Faroese sample. Despite geographic barriers between islands there appears to be a fair amount of movement between each of the three locations sampled in this study (Table 4). Each region is characterized by a high frequency of R1a, R1b and I1 and the inclusion of rarer groups. Although the majority of the Faroese population belongs to one of three highly frequent haplogroups, there are some rare haplogroups present in the population that are infrequent even among Europeans as a whole. Since haplogroup association is characterized by a single nucleotide polymorphism, haplogroup data give insight into older demographic processes. It is likely that haplogroup R1a is representative of the original founder population. In subsequent generations it was

strongly influenced by genetic drift leading it to have the least heterogeneity. Similar to previous genetic studies that examined Y-chromosomal haplotypes, R1a is found to have the lowest amount of inner diversity in this study despite its high frequency among the Faroese (Jorgensen et al., 2004, p. 23). The homogeneity of this group is exemplified by the fact that the majority of individuals (72.4%) are either part of or neighbor haplotypes to the modal node (Figure 10). Despite the effects of drift on individual lineages in the R1a group, dispersal patterns of haplotypes suggest that sufficient time has elapsed for this group to accumulate new mutations since its introduction to the islands. The higher frequency of R1a among Western and Central modern Norwegian populations (32%) (Wiik, 2008, p. 55) is consistent with the premise that the original founder population of the Faroe Islands came from Western Norway. In fact, this region has much higher frequencies of R1a when compared to the larger Scandinavian world (Dupuy et al., 2006). It is especially interesting that the frequency of R1a is relatively high among modern residents of Bergen (Dupuy et al., 2006) because it is likely that this port city would have had historical contact with the Faroe Islands through trading and other economic activities.

Haplogroup I is the third most common group found in the Faroese sample. The most common subhaplogroup of I1 in Northern Europe, I1a, is found at frequencies ranging from 7-54% and is at its highest frequencies in Norway and Sweden (Wiik, 2008, p. 56). The I1 haplogroup network in the Faroe Islands exhibits a remarkably similar dispersal pattern to that found in the R1a haplogroup. It is highly homogeneous with little mutational divergence from the modal haplotype. TMRCA point estimates of the I1 and R1a haplogroups among Faroese haplotypes themselves are very similar. Both haplogroups have an age estimate of 2,792 years with a σ of 1023.5 in the I1 haplogroup and 1225.2 in R1a. Interestingly, these dates are in range of the proposed time of colonization in 825 CE. However, estimating the age of recent haplogroups that have been subjected to the effects of

genetic drifting of individual lineages may lead to an underestimation of the age of the haplogroup (deKnijff, 2000, p. 1059). Inner-haplogroup allele variances in the I1 and R1a groups also show a remarkably low level of heterogeneity (Table 9). Gene diversity scores among individual loci in the R1a haplogroup also exhibit low levels of variance with only one locus, DYS389b, displaying a moderately high diversity score (0.54) (Table 10). After population size reductions, including the original founding event and subsequent population declines due to epidemics or other disasters, the number of alleles declines rapidly in a population (Maruyama and Fuerst, 1985, p. 678). Heterozygosity has been shown to be useful for the measurement of both autosomal DNA and Y-chromosome sequences (Chiaroni, Underhill, and Cavalli-Sforza, 2009, p. 20178). Gene diversity among loci in the I1 haplogroup, however, shows surprisingly high levels of genetic diversity among three loci, two of which have relatively slow mutation rates. This could be evidence for subsequent migration of I1 individuals in the Faroese population or could be a reflection of the original level of diversity of this haplogroup in the founding I1 group. The totality of data presented here, however, suggests that the I1 haplogroup, like R1a, is representative of the original founder group and has remained highly homogeneous due to small founder size and genetic drift. The overall signal from this Faroese sample is one of a founder effect from haplogroups R1a and I1 with high levels of homogeneity due to the effects of genetic drift.

It should be mentioned that the geographic, social, cultural and political environment of the Faroes may have contributed to its modern genetic variability. Along with evolutionary pressures like genetic drift and a small founder size, specific historical events may have conferred further genetic homogeneity to the Faroe Islanders. This is especially true of the Y-chromosome. It has been noted that specific cultural traits like mating practices can reduce effective population size of the Y-chromosome and in turn increase the divergence between groups (Jorgensen et al., 2004, p. 20). After the reformation in the 16th century in the Faroe Islands,

the civil government (*Storidomur*) was replaced by laws formed by the Danish crown. One of these laws was designed to control population size. It significantly limited the ability of a Faroese man to marry and have children (Wylie, 1987, p. 14). Harsh punishments were in store for anyone who bore or fathered a child out of wedlock (death was the punishment if the couple were cousins), and the poor were effectively barred from marriage since only those with a certain amount of land could wed (Wylie, 1987, p. 14). The effectiveness of these rules are unknown, but theoretically, could have had a profound effect on the Y-chromosome of the Faroese.

European admixture: R1b, the second most common haplogroup in the Faroese sample, has considerably higher levels of haplotype diversity (40%) when compared to R1a or I1. Variance at the haplotype level within a haplogroup present in a specific population largely depends on the time since its introduction and on the results of immigration (deKnijff, 2000, p. 1059). R1b has high levels of allele variance as well as gene diversity scores among specific loci. In particular, variance of DYS392 is very high (0.46) especially when compared to the variability of the same locus in the other two haplogroups (Table 9). The mutation rate for this locus is estimated to be approximately 0.07%. Therefore, it is unlikely that the high levels of variance exhibited in this particular locus are the result of stochastic mutational processes. TMRCA estimates of the R1b haplogroup also support the notion that this is representative of a more diverse group. R1b within the Faroese has a much older age estimation than the actual historical event of founding (Table 8). Admixture from disparate R1b lineages could increase the estimated TMRCA since an influx of private haplotypes would lead to an overestimation of the haplogroup's age. TMRCA estimations, while they do not always accurately estimate the time of haplogroup introduction into a particular population, are of interest because they reflect demographic aspects of the size and structure of the ancestral populations themselves (Pritchard et al., 1999, p. 1791). These results suggest that the R1b

haplogroup in the Faroes is more diverse than the other two common haplogroups and may be a signature of a latent SNP in the population. The R1b haplogroup is present at high frequencies in much of Europe and along the Iberian Peninsula in modern populations (Wiik, 2008, p. 37). In Ireland, R1b frequencies can reach as high as 90% (Wiik, 2008, p. 80). The haplogroup is also very common among other North Atlantic groups. In the Shetland Islands, R1b constitutes 66% of the population, in Orkney 64%, and the Isle of Man 70% (Wiik, 2008, p. 79).

In addition, networks illustrating population-specific and aggregate haplotypes reveal a higher level of diversity among the three mainland European populations used in this study (Norwegian, Swedish, and Danish) when compared to island populations (Faroese, Icelandic, Irish). There is, however, a notable increase of diversity among R1b haplogroups in all populations. This pattern of diversity in the R1b haplogroup among the Faroese may be representative of a composite R1b founder group in the Faroes. Reportedly, the Faroese have a strong association with the Celtic world. The islands were likely founded by Scandinavians living in the larger Celto-Norse world for some time before the islands were settled. The diversity of the R1b haplogroup may be due to an influx of independent Celtic male settlers or the incorporation of Celtic male slaves that likely accompanied the original founding group. The genetic association between the Faroes and the British Isles is most visible in the mtDNA of women living on the islands. An interesting observation of island populations in the North Atlantic is the discrepancy between male and female origins in many Viking founded populations. Usually, males are of Scandinavian origin while the females are either indigenous to the settled location or are from elsewhere. This indicates that the vast majority of Viking migrations were undertaken by groups of men and not by families (there are some well researched exceptions to this, i.e.: the Orkney and Shetland Islands (Goodacre et al., 2005)). Genetic studies have found that many North Atlantic populations, including Greenland, follow this trend (Jorgensen et al., 2004, p. 27). This pattern also holds

true in the Faroe Islands. A study of the mtDNA of the Faroese found that only 17% of the females living in the Faroes were of Scandinavian descent and 83% had their origins in the British Isles (Als et al., 2006, p. 501). Faroese Y-chromosomes show the opposite, where only 13% of males had their likely origins in the British Isles but 87% were of Scandinavian origin (Als et al., 2006, p. 501).

It is also possible that the diversity found in the R1b haplogroup was influenced by subsequent admixture of divergent European groups after the time of settlement. As was discussed, historical contact with European traders and modern contact due to the fishing industry could be represented in Y-chromosomal haplotype frequencies in the Faroes. It may be that the diversity of R1b is due to a composite founding group. Alternatively, this diversity may be secondary to more recent admixture. Since the shape of a genealogical tree often denotes that population's particular demographic history (Donnelly, 1999, p. 188) it can be assumed with either scenario that the demographic history of the R1b haplogroup in the Faroe Islands is very different from that in the R1a or I1 group.

The pattern of haplotype dispersal in the Norwegian and Swedish R1b samples are interesting in that they suggest a lower level of diversity among individuals in this haplogroup. Since the samples from these two populations were collected from geographically diverse regions within Norway and Sweden, it is unlikely that this is due to sampling errors. Haplogroup R1b is thought to have moved westward across the European continent as a part of the movement of people and culture during the Neolithic (Wilson et al., 2001, p. 5078). The method and pace of this movement is an often debated topic but no one theory is absolute. It has been suggested that the reality of the affinities between individuals in the R1b haplogroup is obscured by the lack of information concerning the various R1b subhaplogroups (Myres et al., 2011, p. 95). Specific R1b subhaplogroups dominate in geographically defined populations. These patterns of locality have been found in various Central and Western European populations (Myres et al., 2011, p. 99). It is

possible that specificity of a particular R1b subhaplogroup may explain the lack of variation seen in Norway and Sweden. However, this is outside the scope of this discussion.

Non-European admixture: To help clarify the origins of the rare haplotypes within the Faroese population, a comparison was made to the YHRD's database of Y-chromosomal haplotypes (*Y-Chromosome Haplotype Reference Database* 2012). Of the J1 individuals, four haplotypes matched a haplotype found in Kuwait that is associated with the Afro-Asiatic-Semitic ethnic group. The other two J1 individuals had no direct match but were neighbors (i.e., only one mutational step away) of the previously-mentioned haplotype. Three haplotypes in the Q group were matched to haplotypes found in Iceland and among African Americans. The remaining Q individual had a neighbor haplotype. In modern populations this haplogroup is most commonly associated with specific Siberian populations, but is also at very high frequencies among Native American groups. Among E1b1b individuals in the Faroe data set, matching haplotypes were found in multiple populations across Europe. One individual had direct matches with Europeans in Australia, in Alpujarra de la Sierra in Spain, and a match with Mestizos in Costa Rica. E1b1b occurs at relatively low frequencies in Scandinavia, but is common among other European groups (Wiik, 2008, p. 54). In modern populations, E1b1b is found at relatively high frequencies in the Mediterranean region, especially in Greece, and represents one of the earliest migrations of farmers from the Middle East into Europe (Wiik, 2008, p. 37). Because of the age and wide dispersal of the E1b1b haplogroup it is impossible to parse relation by descent or by equivalent mutations in this group. Haplogroup I2b1 had no matches or neighbor haplotypes in the YHRD. It may therefore be a unique haplotype in the Faroes, or may be one not yet represented in the database. I2a(xI2a1) had many matches across Europe. I2b(xI2b1) had no direct matches but had a neighbor haplotype match in European Australians.

Haplogroup L had direct haplotype matches in multiple places in South Korea and in Gunma, Japan. In modern populations this haplogroup is most commonly associated with the Indian subcontinent where it is at its highest frequencies. It is also found along the Mediterranean coast line (Karafet et al., 2008, p. 835). Haplogroup N had multiple matches in Northern Europe including Russia, Sweden, Belarus and Germany, and also had a match in South Korea. This haplogroup, like haplogroup Q, is most strongly associated with specific Siberian groups (Karafet et al., 2008, p. 835). These haplotype matches do not necessarily imply relation by descent, but may instead be analogous due to parallel mutations in the disparate populations. In fact, one Faroese E1b1b haplotype matched haplotypes found in the Danish sample (n=2) while all Faroese Q individuals matched with Q haplotypes found in the Norwegian sample. Since haplogroup Q is associated with modern populations far from Europe, it is likely that these haplogroups reached the Faroes very recently by way of Scandinavia or through the fishing industry.

Of all rare haplogroups found in this Faroese sample, haplogroup J1 provides the best evidence for non-European admixture in the Faroe Islands. Haplogroup J originated in the Middle East and is associated with the origins and spread of early farming groups (Wiik, 2008, p. 37). Unlike the European J-M172 subhaplogroup, however, the subhaplogroup found in this Faroese sample is most closely associated with Middle Eastern, Northern African and Ethiopian populations (Karafet et al., 2008, p. 834). J1 is characterized by the SNP M267 and is found at high frequencies in modern populations living in the Arabian Peninsula and along the Northern coast of Africa (Semino et al., 2004, p. 1026). J1 is not found in any of the parental populations represented in this study and is therefore highly unlikely to have originated from the founding population. Perhaps most intriguing is that there is historical evidence that may account for this haplogroup's presence in the Faroes. The Barbary corsair attack on the islands in the 1600s could explain J1's presence in modern Faroese Y-chromosomes (West, 1980). Modern populations living in

regions where Barbary corsair ports once existed still possess high frequencies of the J1 haplogroup. Interestingly, the J haplogroup is also found at higher than expected frequencies in other Atlantic island populations with complex colonization histories. An investigation into Y-chromosome diversity in the Azores revealed that the haplogroup frequencies were twice those in mainland Portugal (Pacheco et al., 2005, p. 150). Since the haplogroup is common among Jewish, Turkish and Arab populations, the authors speculate that this could be a signature of influence from early Jewish settlers on the Azoreans (Pacheco et al., 2005, p. 153). Unfortunately, there are very few J1 individuals in the Faroese sample. A larger data set with more J1 individuals could help clarify the age and nature of this haplogroup among the Faroese.

Other findings: An interesting observation in this study is the apparent high similarity of the Faroese to the Danish population. In previous studies of the Faroese Y-chromosome, it was found that Danish admixture on the islands was minimal even though they were under Danish rule since the 1380s (Jorgensen et al., 2004, p. 26). After the Faroes were appropriated into the Danish kingdom, the islands were settled by the odd Danish priest or government official (Jorgensen et al., 2004, p. 20) but there is little evidence that the newcomers thrived in the Faroes. This discrepancy could be an error due to sampling, the statistical lack of distinction between Danish haplogroup frequencies from other Scandinavian haplogroup frequencies (Jorgensen et al., 2004, p. 22), a product of sampling size error across the parental populations (for example, the Danish data set used here is much larger than the Norwegian data set), a representation of more recent movements between the Faroes and Denmark, or it could be due to the effects of drift. In particular, this could be evidence of Faroese migrations to Denmark seeking higher education and economic opportunities. The University of the Faroe Islands was not founded until the 1960s. Prior to that time those Faroese who wanted to seek a higher eduction or

economic opportunities would often emigrate to Denmark (Coull, 1966, p. 164).

Perhaps most surprising is the lack of relationship between the Faroes and Iceland in the current study. Previous studies of Faroese Y-chromosomes have found a high degree of similarity between the Faroese and Icelandic populations (Jorgensen et al., 2004, p. 23). It is likely that this discrepancy is due to the number of STR loci used in the (Jorgensen et al., 2004) study (five) and that of the current study (twelve). Divergence can be more easily detected if more loci data are available to compare populations. In particular, four relatively fast mutating loci were used for this study and not in previous investigations. Changes in these loci could easily affect the detected amount of divergence between populations. If Iceland and the Faroes are indeed genetically divergent as is suggested in this study, two scenarios may explain their differentiation: (1) the populations were settled by different groups; or, (2) there could have been sufficient isolation between the two groups to result in today's divergence. From the data presented here, however, it is apparent that the Icelandic and Faroese populations had distinguishably different founding fathers. Median-joining networks created in this study consistently place the Icelandic group apart from all other populations while the Faroese tend to be well dispersed. Data presented here also suggest that the Faroese male population was founded by a more diverse group and may have been subject to more admixture from divergent groups than their Icelandic neighbors. In any case, this study demonstrates a clear difference in parental population and diversity of founding groups between the two island populations. It also suggests that there has been no post-founder admixture between the two. Moreover, the Icelandic population may not be as homogeneous as is often proposed. Arnason (2003) has suggested that the distribution of pairwise differences among Icelandic mtDNA is comparable to those found in European populations, and that the Icelandic population is among the most genetically heterogeneous European groups (Arnason, 2003, p. 5). TMRCA calculated in this study of the Icelandic branch of the R1a aggregate network

(Figure 9) estimated it to be approximately 5,784 years old ± 1,802 years — decidedly older than expected of a homogeneous and isolated population. The estimated coalescent age of a population will be older if divergent haplotypes are included in the sample. Lineages that have had time to accumulate novel mutations in isolation will increase the divergence time between haplotypes when analyzed in the same sample. The severity of this divergence depends on the length of the isolation time. For example, if two isolated populations converge and then break off into a founder group after a moderate amount of time (meaning that enough mutations have occurred to differentiate between the two), then the TMRCA tends to be older than the actual founding event. If, however, the time of isolation has been so great that a high number of parallel mutations have occurred in the two populations, it may be the case that the TMRCA will not be as severely affected in the same scenario (deKnijff, 2000). Since, however, the populations used in this study are part of a closer regional population, one can assume that this is not the case in the current study . Network distributions and neighbor haplotype counts of Icelandic individuals in this study, especially those haplogroups that do not present a smooth distribution of mutational steps away from the modal haplotype, may be the result of latent SNPs present in these groups.

Limitations: As with any study, there are limitations in how the data can be manipulated and interpreted. This study assumes the stepwise mutation model. However, the Y-chromosome evolves in very complex ways and haplotype history may not be fully representative of the actual population history (Zhivotovsky et al., 2004, p. 51). In fact, there is recent evidence that Y-microsatellites undergo mutations that do not follow the stepwise mutation model and can undergo multistep mutations (Zhivotovsky et al., 2004, p. 51). This can have dramatic effects on the estimation of mutation rate for a specific locus, and it can negatively affect appraisal of the age of population events (TMRCA) (Zhivotovsky et al., 2004,

p. 51). If the mutation rate is not properly calculated it could have the effect of producing a 10x deviation from the actual age of the population event examined (Zhivotovsky et al., 2004, p. 51) While the networking software used here is an effective tool for visualizing phylogenetic relationships, it has some notable limitations in the realm of genetic analysis. Among these is a reliance on assumptions made for STR evolution and mutation rates. To help mediate error due to these concerns, the same networks presented here were given different weighting schemes. Little movement was exhibited by larger nodes. Therefore, the networks themselves can be considered relatively robust. Moreover, the accuracy of TMRCA estimates is dependent on the demographic model that is assumed for the population under study (Pritchard et al., 1999, p. 1791).

Haplogroup analysis is also a challenge when studying the Y-chromosome. Due to the fact that additional methods to test for haplogroup association have been developed to help curtail the price of SNP typing samples, there are a variety of programs that can estimate haplogroup from haplotype data. The method used here is rather effective in determining the actual SNP haplogroup but also presents some limitations. Employing this method, necessarily infers a less precise haplogroup analysis because of the low number of loci used. Also, because a haplotype can be assigned to only one of ten common groups, the method only provides a superficial haplogroup inference.

The unavailability of complete data sets also poses a problem when comparing populations. Of the 12 STRs used in this data set, only eight could be compared across populations. This is due to the lack of data on three loci from the other published haplotype sets. As it becomes easier and more cost effective to score and analyze Y-chromosomal microsatellites, more complete data sets will become available and more precise analyses can be achieved using more loci. Another potential problem is the differences between total population size in each group's data set as well as the potential for other parental groups not included in this study.

In particular, the current investigation did not use Y-chromosome haplotype data from Scotland which has been shown to have genetic ties to the Faroese (Jorgensen et al., 2004).

Both measures of genetic distance used in this study had conflicting results likely due to the prior assumptions of each model. Both R_{st} and D_c have beneficial and limiting aspects and therefore were used in this study to compare one method to the other using the same data set. Since there are multiple measures of genetic distance, none of which have been proven conclusively to be more effective than the other, it may be the case that the distances (R_{st}, D_c) reported here are equally biased and therefore should not be considered reliable estimates of the true genetic distances between populations in this study.

Future research: It would be interesting to see what rare haplotypes exist in the female population in the Faroe Islands. If geographic mtDNA analogues of the non-Scandinavian haplotypes from this study are found in the females of the Faroe Islands, then the hypothesis that the admixture is from a largely male trading and fishing population would have to be reanalyzed.

CHAPTER VII

CONCLUSION

Despite the geographic isolation of the Faroese population, there is evidence that the population has been in contact with multiple groups over the course of its history. Like other North Atlantic Island groups, the Faroese are thought to be uniquely suited for gene mapping studies which rely on the assumption that the population in question is truly isolated and inbred over multiple generations. Because of this, it is important to determine the full demographic history of the population, including potential sources for admixture from other regions. Demographic events such as admixture are also increasingly being used to explain and detect high levels of congenital disease prevalence among recently admixed groups (Reich and Patterson, 2005). Many genetic studies have been done on the islands of the North Atlantic but minor admixture from traders and other migrants is rarely addressed or only briefly mentioned. Some studies like the (Pacheco et al., 2005) study of the Azores have investigated the signatures of these brief contacts between disparate populations and island inhabitants, but this is the first study to directly examine these rare haplotypes within the Faroe Islands. While these results are intriguing, a larger sample would be needed to determine the estimated age of rarer haplogroups in the Faroese. In conclusion, the overall pattern of Y-chromosomal markers demonstrated here is one of a very strong founder effect from Scandinavian countries (especially in the case of haplogroups R1a and I1), and minor instances of contact with disparate populations within and beyond Europe.

REFERENCES

Als, Thomas D. et al. (2006). "Highly discrepant proportions of female and male Scandinavian and British Isles ancestry within the isolated population of the Faore Islands". In: *European Journal of Human Genetics* 14, pp. 497–504.

Arge, Simun V. (1991). "The Landnam in the Faroes". In: *Arctic Anthropology* 28(2), pp. 101–120.

Arge, Simun V. et al. (2005). "Viking and Medieval Settlement in the Faroes: People, Place and Environment". In: *Human Ecology* 33, pp. 597–620.

Arnason, E. (2003). "Genetic heterogeneity of Icelanders". In: *Annals of Human Genetics* 67, pp. 5–19.

Athey, T. Whit (2005). "Haplogroup Prediction from Y-STR Values Using an Allele-Frequency Approach". In: *Journal of Genetic Genealogy* 1, pp. 1–7.

Ballard, D.J. et al. (2006). "Y-chromosome STR haplotype data for an Irish population". In: *Forensic Science International* 161, pp. 64–68.

Bandelt, H-J, P. Forster, and A. Rohl (1999). "Median-joining networks for inferring intraspecific phylogenies". In: *Molecular Biological Evolution* 16, pp. 37–48.

Bronner, Hedin (1972). *Faroese Short Stories.* Twayne Publishers: New York.

Buschiazzo, Emmanuel and Neil J. Gemmell (2006). "The rise, fall and renaissance of microsatellites in eukaryotic genomes". In: *BioEssays* 28, pp. 1040–1050.

Cavalli-Sforza, L.L. and A.W.F. Edwards (1967). "Phylogenetic Analysis Models and Estimation Procedures". In: *American Journal of Human Genetics* 19(3), pp. 233–257.

Chiaroni, Jacques, Peter A. Underhill, and Luca L. Cavalli-Sforza (2009). "Y chromosome diversity, human expansion, drift and cultural evolution". In: *PNAS* 106(48), pp. 20174–20179.

CIA (2011). *The World Factbook*. English. CIA. URL: https://www.cia.gov/library/publications/the-world-factbook/.

Consortium, Y Chromosome (2002). "A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups". In: *Genome Research* 12, pp. 339–348.

Coull, J.R. (1966). "A Comparison of Demographic Trends in the Faroe and Shetland Islands". MA thesis. University of Aberdeen.

deKnijff, Peter (2000). "Messages through Bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome". In: *American Journal of Human Genetics* 67, pp. 1055–1061.

Donnelly, Peter (1999). "The coalescent and microsatellite variability. In: Microsatellites: Evolution and applications". In: ed. by D. Goldstein and C. Schlotterer. Oxford University Press. Chap. 9.

Dupuy, Berit Myhre et al. (2006). "Geographical heterogeneity of Y-chromosomal lineages in Norway". In: *Forensic Science International* 164, pp. 10–19.

Ellegren, H. (2000). "Heterogeneous mutation processes in human microsatellite DNA sequences". In: *Nature Genetics* 24, pp. 400–402.

Felsenstein, Joseph (1985). "Confidence Limits on Phylogenies: An approach using the bootstrap". In: *Evolution* 39(4), pp. 783–791.

Goldstein, David B. et al. (1995). "An Evaluation of Genetic Distances for Use with Microsatellite Loci". In: *Genetics* 139, pp. 463–471.

Goodacre, S. et al. (2005). "Genetic evidence for a family-based Scandinavian settlement of Shetland and Orkney during the Viking periods". In: *Heredity* 95(2), pp. 129–35.

Goodman, S.J. (1997). "Rst Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance". In: *Molecular Ecology* 6, pp. 881–885.

Hallenberg, C. et al. (2005). "Y-chromosome STR haplotypes in Danes". In: *Forensic Science International* 155, pp. 205–210.

Helgason, A. et al. (2000a). "Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland". In: *American Journal of Human Genetics* 67, pp. 697–717.

Helgason, A. et al. (2001). "mtDNA and the Islands of the North Atlantic: Estimating the proportions of Norse and Gaelic ancestry". In: *American Journal of Human Genetics* 68, pp. 723–737.

Helgason, Agnar et al. (2000b). "mtDNA and the Origin of the Icelanders: Deciphering signals of recent population history". In: *Am J Hum Genet* 66, pp. 999–1016.

Hjortshoj, T. Deuelund et al. (2009). "A novel founder BBS1 mutation explains a unique high prevalence of Bardet-Biedl syndrome in the Faroe Islands". In: *Br J Ophthalmol* 93, pp. 409–413.

Jackson, Anthony (1979). *North Atlantic Maritime Cultures: Anthropological essays on changing adaptations*. Ed. by Raoul Andersen. Mouton: New York.

Johnston, George (1975). *The Faroe Islanders' Saga*. Hazell Watson and Viney Ltd.

Jones, Gwyn (1984). *A History of the Vikings*. Oxford University Press.

Jorde, L.B. et al. (2000). "The Distribution of Human Genetic Diversity: A comparison of mitochondrial, autosomal, and Y-chromosome data". In: *Am J Hum Genet* 66, pp. 979–988.

Jorgensen, Tove H. et al. (2002). "Linkage disequilibrium and demographic history of the isolated population of the Faroe Islands". In: *European Journal of Human Genetics* 10, pp. 381–387.

Jorgensen, Tove H. et al. (2004). "The origin of the isolated population of the Faroe Islands investigated using Y chromosomal markers". In: *Human Genetics* 115, pp. 19–28.

Karafet, T. (1999). "Y chromosome microsatellite haplotypes and the history of Samoyed-speaking populations in north-west Siberia. In: Microsatellites: Evolution and applications". In: ed. by D. Goldstein and C. Schlotterer. Oxford University Press. Chap. 19.

Karafet, Tatiana M. et al. (2008). "New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree". In: *Genome Research* 18, pp. 830–838.

Lawson, Ian T. et al. (2005). "Historical Ecology on Sandoy, Faroe Islands: Palaeoenvironmental and Archaeological Perspectives". In: *Human Ecology* 33(5), pp. 651–684.

Lewis, Bernard (1973). "Corsairs in Iceland". In: *Revue de l'Occident musulman et de la Mediterranee* 15-16, pp. 139–144.

Linares, Andres Ruiz (1999). "Microsatellites and the reconstruction of the history of human populations. In: Microsatellites: Evolution and applications". In: ed. by D. Goldstein and C. Schlotterer. Oxford University Press. Chap. 9.

Magnusson, Magnus (2010). *The Vikings*. History Press Lmtd.

Maruyama, Takeo and Paul A. Fuerst (1985). "Population Bottlenecks and Nonequilbrium Models in Population Genetics. II. Number of Alleles in a Small Population that was Formed by a Recent Bottleneck". In: *Genetics* 111, pp. 675–689.

Mehler, Natascha (2009). "The Perception and Interpretation of Hanseatic Material Culture in the North Atlantic: Problems and Suggestions". In: *Journal of the North Atlantic* 1, pp. 89–108.

Minch E.; Ruiz-Linares, A.; Goldstein D.; Feldman M. & Cavalli-Sforza L. *MICROSAT Ver. 1.5.* URL: `http://hpgl.standord.edu/projects/microsat`.

Moore, Laoise T. et al. (2006). "A Y-Chromosome Signature of Hegemony in Gaelic Ireland". In: *The American Journal of Human Genetics* 78, pp. 334–338.

Myres, Natalie M et al. (2011). "A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe". In: *European Journal of Human Genetics* 19, pp. 95–101.

Nauerby, T. (1996). *No nation is an island: language, culture and national identity in the Faroe Islands*. SNAI–North Atlantic Publications.

Nei, Masatoshi (1973). "Analysis of Gene Diversity in Subdivided Populations". In: *PNAS* 70(12), pp. 3321–3323.

Network 4.6.1.0. User Guide (2011). English. Fluxus Technologies Ltd.

NIST Standard Reference Database (2012). English. NIST. URL: `www.cstl.nist.gov/strbase`.

Ohta, Tomoko and Motoo Kimura (1973). "A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population". In: *Genet. Res. Camb.* 22, pp. 201–204.

Pacheco, P.R. et al. (2005). "The Y-chromosomal Heritage of the Azores Islands Population". In: *Ann Hum Genet* 69(2), pp. 145–156.

Pritchard, Jonathan K. et al. (1999). "Population Growth of Human Y Chromosomes: A study of Y chromosome microsatellites". In: *Molecular Biological Evolution* 16(12), pp. 1791–1798.

Ravid-Amir, Osnat and Saharon Rosset (2010). "Maximum likelihood estimation of locus-specific mutation rates in Y chromosome short tandem repeats". In: *BioInformatics* 26, pp. 440–445.

Reich, David and Nick Patterson (2005). "Will admixture mapping work to find disease genes?" In: *Phil. Trans. R. Soc. B* 360, pp. 1605–1607.

Rienzo, A. Di et al. (1994). "Mutational processes of simple-sequence repeat loci in human populations". In: *Proc. Natl. Acad. Sci.* 91, pp. 3166–3170.

Roewer, Lutz et al. (1996). "Analysis of Molecular Variance (AMOVA) of Y-Chromosome-Specific Microsatellites in Two Closely Related Human Populations". In: *Human Molecular Genetics* 5(7), pp. 1029–1033.

Roewer, Lutz et al. (2005). "Signature of recent historical events in the European Y-chromosomal STR haplotype distribution". In: *Human Genetics* 116, pp. 279–291.

Ruzzante, Daniel E. (1998). "A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance". In: *Can. J. Fish. Aquat. Sci.* 55, pp. 1–14.

Semino, Ornella et al. (2004). "Origin, Diffusion, and Differentiation of Y-Chromosome Haplogroups E and J: Inferences on the Neolithization of Europe and Later Migratory Events in the Mediterranean Area". In: *American Journal of Human Genetics* 74, pp. 1023–1034.

Slatkin, Montgomery (1995). "A Measure of Population Subdivision Based on Microsatellite Allele Frequencies". In: *Genetics* 139, pp. 457–462.

West, J.F. (1980). *Faroese folk-tales and legends.* Shetland Publishing Company.

Wiik, Kalevi (2008). "Where Did European Men Come From?" In: *Journal of Genetic Genealogy* 4, pp. 35–85.

Wikimedia (2012). URL: `http://commons.wikimedia.org/`.

Wilson, James F. et al. (2001). "Genetic evidence for different male and female roles during cultural transitions in the British Isles". In: *PNAS* 98(9), pp. 5078–5083.

Wylie, J. (1987). *The Faroe Islands: Interpretations of History.* The University Press of Kentucky.

Xu, Xin et al. (2000). "The direction of microsatellite mutations is dependent upon allele length". In: *Nature Genetics* 24, pp. 396–399.

Y-Chromosome Haplotype Reference Database (2012). URL: `http://www.yhrd.org/`.

Young, G.V.C. (1979). *From the Vikings to the Reformation: A chronicle of the Faroe Islands up to 1538.* Bridson & Horrox, Lmtd.: The Isle of Man.

Zeller, Dirk and Jakup Reinert (2004). "Modelling spatial closures and fishing effort restrictions in the Faroe Islands marine ecosystem". In: *Ecological modelling* 172, pp. 403–420.

Zhivotovsky, L.A. et al. (2004). "The effective mutation rate at Y chromosome short tandem repeats, with application to human population divergence time". In: *American Journal of Human Genetics* 74(1), pp. 50–61.

APPENDIX A

R scripts

```
par(mfrow=c(2,2))
Rst <- read.table("Rst.csv", header=FALSE, row.names=1, sep=",")
d <- dist(Rst)
fit <- cmdscale(d,eig=TRUE,k=2)
fit
x <- fit$points[,1]
y <- fit$points[,2]
Rst_MDS <- plot(x,y,xlab="Coordinate 1", ylab="Coordinate 2", main="All haplotypes", type="n")
text(x,y, labels=row.names(Rst), cex=.7)
R1a_Rst <- read.table("R1a_Rst.csv", header=FALSE, row.names=1, sep=",")
d <- dist(R1a_Rst)
fit <- cmdscale(d,eig=TRUE,k=2)
fit
x <- fit$points[,1]
y <- fit$points[,2]
R1a_Rst_MDS <- plot(x,y,xlab="Coordinate 1", ylab="Coordinate 2", main="R1a only", type="n")
text(x,y, labels=row.names(R1a_Rst), cex=.7)
R1b_Rst <- read.table("R1b_Rst.csv", header=FALSE, row.names=1, sep=",")
d <- dist(R1b_Rst)
fit <- cmdscale(d,eig=TRUE,k=2)
fit
x <- fit$points[,1]
y <- fit$points[,2]
R1b_Rst_MDS <- plot(x,y,xlab="Coordinate 1", ylab="Coordinate 2", main="R1b only", type="n")
text(x,y, labels=row.names(R1b_Rst), cex=.7)
I1_Rst <- read.table("I1_Rst.csv", header=FALSE, row.names=1, sep=",")
```

```
d <- dist(I1_Rst)
fit <- cmdscale(d,eig=TRUE,k=2)
fit
x <- fit$points[,1]
y <- fit$points[,2]
I1_Rst_MDS <- plot(x,y,xlab="Coordinate 1", ylab="Coordinate 2", main="I1 only", type="n")
text(x,y, labels=row.names(I1_Rst), cex=.7)
pdf("/home/allison/Dropbox/Thesis Documents/Images/All_MDS_Charts.pdf")
par(mfrow=c(2,2))
Rst <- read.table("Rst.csv", header=FALSE, row.names=1, sep=",")
d <- dist(Rst)
fit <- cmdscale(d,eig=TRUE,k=2)
fit
x <- fit$points[,1]
y <- fit$points[,2]
Rst_MDS <- plot(x,y,xlab="Coordinate 1", ylab="Coordinate 2", main="All haplotypes", type="n")
text(x,y, labels=row.names(Rst), cex=.7)
R1a_Rst <- read.table("R1a_Rst.csv", header=FALSE, row.names=1, sep=",")
d <- dist(R1a_Rst)
fit <- cmdscale(d,eig=TRUE,k=2)
fit
x <- fit$points[,1]
y <- fit$points[,2]
R1a_Rst_MDS <- plot(x,y,xlab="Coordinate 1", ylab="Coordinate 2", main="R1a only", type="n")
text(x,y, labels=row.names(R1a_Rst), cex=.7)
R1b_Rst <- read.table("R1b_Rst.csv", header=FALSE, row.names=1, sep=",")
d <- dist(R1b_Rst)
fit <- cmdscale(d,eig=TRUE,k=2)
fit
x <- fit$points[,1]
y <- fit$points[,2]
R1b_Rst_MDS <- plot(x,y,xlab="Coordinate 1", ylab="Coordinate 2", main="R1b only", type="n")
text(x,y, labels=row.names(R1b_Rst), cex=.7)
I1_Rst <- read.table("I1_Rst.csv", header=FALSE, row.names=1, sep=",")
```

```
d <- dist(I1_Rst)
fit <- cmdscale(d,eig=TRUE,k=2)
fit
x <- fit$points[,1]
y <- fit$points[,2]
I1_Rst_MDS <- plot(x,y,xlab="Coordinate 1", ylab="Coordinate 2", main="I1 only", type="n")
text(x,y, labels=row.names(I1_Rst), cex=.7)
dev.off()
```

APPENDIX B

Confidence Intervals

CHD RST

Aggregate ChD: Aggregate Rst:

FADN

[1] 0.4262 0.2498 [1] -0.0386 -0.0974

FAICE

[1] 0.52024 0.34776 [1] 0.14452 -0.09852

DNICE

[1] 0.46596 0.26604 [1] 0.07964 -0.05364

FAIR

[1] 0.4286 0.2914 [1] 0.02252 -0.12252

IRDN

[1] 0.3568 0.2392 [1] -0.01984 -0.10216

IRICE

[1] 0.4428 0.2272 [1] 0.1106 -0.1246

FANO

[1] 0.42096 0.22104 [1] -0.04496 -0.13904

NODN

[1] 0.43676 0.21724 [1] -0.06648 -0.11352

NOICE

[1] 0.47168 0.24432 [1] -0.1324 0.1224

NOIR

[1] 0.43664 0.20536 [1] -0.07656 -0.13144

FASWE

[1] 0.48564 0.25436 [1] -0.04368 -0.11032

SWEDN

[1] 0.50032 0.33568 [1] -0.0744 -0.1136

```
SWEICE
[1] 0.51584 0.30416   [1] 0.12632 -0.13632
SWEIR
[1] 0.45104 0.25896   [1] -0.04096 -0.13504
SWENO
[1] 0.29844 0.04756   [1] -0.11752 -0.16848
R1a:
FADN
[1] 0.49424 0.22376   [1] -0.06584 -0.24616
FAICE
[1] 0.62224 0.25376   [1] 0.22196 -0.36996
DNICE
[1] 0.72472 0.20728   [1] 0.21192 -0.3839
FAIR
[1] 0.58948 0.44052   [1] -0.1338 -0.5062
IRDN
[1] 0.62672 0.50128   [1] -0.19232 -0.41968
IRICE
[1] 0.69256 0.25744   [1] 3415.642 -3612.588
FANO
[1] 0.49876 0.37724   [1] -0.08764 -0.24836
NODN
[1] 0.491 0.099   [1] -0.00592 -0.39008
NOICE
[1] 0.65748 0.11652   [1] 0.38604 -0.39404
NOIR
[1] 0.5712 0.3948   [1] -0.25752 -0.50448
FASWE
[1] 0.40416 0.12584 [1] -0.08648 -0.32952
SWEDN
[1] 0.50952 0.26648   [1] -0.17452 -0.22548
SWEICE
[1] 0.68464 0.25736   [1] 0.23768 -0.38168
SWEIR
```

[1] 0.53712 0.45088 [1] -0.346 -0.542

SWENO

[1] 0.42492 0.12308 [1] -0.06644 -0.30556

R1b:

FADN

[1] 0.56872 0.24728 [1] -0.07088 -0.25512

FAICE

[1] 0.65944 0.40856 [1] 0.24572 -0.27172

DNICE

[1] 0.61532 0.25468 [1] 0.30156 -0.13356

FAIR

[1] 0.51504 0.32296 [1] -0.03992 -0.22808

IRDN

[1] 0.35308 0.06692 [1] -0.1186 -0.1774

IRICE

[1] 0.54892 0.24708 [1] 0.21944 -0.12944

FANO

[1] 0.48772 0.16628 [1] -0.13372 -0.40028

NODN

[1] 0.54992 0.34608 [1] -0.09936 -0.23264

NOICE

[1] 0.63808 0.44992 [1] 0.14028 -0.32228

NOIR

[1] 0.56564 0.43236 [1] -0.1238 -0.2022

FASWE

[1] 0.4906 0.3534 [1] -0.1766 -0.3334

SWEDN

[1] 0.60092 0.49508 [1] -0.12924 -0.25076

SWEICE

[1] 0.686 0.49 [1] 0.34368 -0.27568

SWEIR

[1] 0.60768 0.47832 [1] -0.06948 -0.21452

SWENO

[1] 0.51744 0.26656 [1] -0.08792 -0.37408

```
I1:
FADN
[1] 0.53904 0.34696   [1] -0.04052 -0.28748
FAICE
[1] 0.52584 0.11816   [1] 0.5944 -0.4444
DNICE
[1] 0.65112 0.27088   [1] 0.42232 -0.23232
FAIR
[1] 0.58864 0.25936   [1] 0.02684 0.05714184
IRDN
[1] 0.51484 0.40116   [1] -0.16684 -0.24916
IRICE
[1] 0.58932 0.22868   [1] 0.52568 -0.28968
FANO
[1] 0.46004 0.26796   [1] -0.22232 -0.44968
NODN
[1] 0.56252 0.41748   [1] -0.12388 -0.30812
NOICE
[1] 0.58868 0.36132   [1] 0.43964 -0.37964
NOIR
[1] 0.5832 0.2108   [1] 0.19152 -0.34552
FASWE
[1] 0.53888 0.23312   [1] 0.04592 -0.35392
SWEDN
[1] 0.50364 0.27236   [1] -0.15612 -0.26588
SWEICE
[1] 0.58676 0.17124   [1] 0.55984 -0.33784
SWEIR
[1] 0.52872 0.30528   [1] -0.05068 -0.31332
SWENO
[1] 0.49684 0.18716   [1] -0.00444 -0.34156
```

CURRICULUM VITAE

NAME: Allison Elaine Mann

ADDRESS: Department of Anthropology
University of Louisville
Louisville, KY 40292

EDUCATION:

B.A. Anthropology
University of Louisville
2009

B.S. Sociology
University of Louisville
2009

PROFESSIONAL SOCIETIES:

American Association of Physical Anthropologists
American Association of Anthropological Genetics

POSTER PRESENTATIONS:

UofL Graduate Research Symposium (2012)
2012 AAPA Meeting