Finding the "Scot" in the Scottish-American: An Investigation of Scottish Identity through Mitochondrial DNA and Y-chromosome Markers

By

Kristine G. Beaty

Submitted to the Graduate Degree Program in Biological Anthropology and the Faculty of the Graduate School of the University of Kansas In partial fulfillment of the requirements for the degree of
Master's of Arts

Dr. Michael Crawford, Chairperson

Dr. James H. Mielke, Committee members

Dr. Bartholomew Dean

Date Defended: 27 October 2009

The Thesis Committee for Kristine G. Beaty certifies
That this is the approved Version of the following thesis:

**Finding the "Scot" in the Scottish-American:
An Investigation of Scottish Identity through Mitochondrial DNA
and Y-chromosome Markers**

Committee:

Dr. Michael H. Crawford, Chairperson

_____

_____

Date Approved: 17 December 2009

Abstract

Individuals from Scotland began migrations to the Americas in the 1600s and, despite admixture, continue to celebrate their Scottish roots. Molecular markers were used to answer the following: 1) do individuals who claim a Scottish identity have maternal or paternal markers found in Scotland; 2) which of these markers are more successful in determining a Scottish ancestry? Of the study's participants, 60 percent shared HVS-I haplotypes with Scotland. However, analyses based on Tajima and Nei's distances indicated that these haplotypes are found throughout Europe due to historical migrations. Seventy-seven percent of males shared a haplotype with Scotland based on five Y-STR loci. Slatkin's R_{ST} analyses, as well as clan and surname distributions, showed a stronger Scottish paternal genetic input in this study, in agreement with historical records of higher rates of male *versus* female migration. These results indicate that ethnic identity can be detected through genetic markers.

This thesis is dedicated to my family: to my mom for telling me I could be anything I wanted to be, to my dad who taught me to ask questions and to look at an issue from the other side, to Jeff for always believing and making me laugh when I needed to most, and to Mikah and Miles who inspire me to keep at it.

Acknowledgements

I would like to thank all of the enthusiastic participants who volunteered for this project. Without their enthusiasm and patience, this study would not have been possible. I would also like to thank Marion L. Mealy, who recruited volunteers, collected demographic information, and helped perform some of the lab work. Without her enthusiasm, this project would have not have begun.

I would also like to thank my committee members: Dr. Michael Crawford, Dr. James Mielke, and Dr. Bartholomew Dean; Dr. Mike Grose of the University of Kansas Sequencing Lab; Dr. Alan Redd, Norberto Baldi, Geetha Chittoor, Anne Justice, Dr. Philip Melton, Dr. Kristen Young, Dr. Mark Zlojutro, and my other friends and fellow students in the Department of Anthropology and the Laboratory of Biological Antrhopology; the Department of Anthropology office staff: Carol Archinal, Judy Ross, and Kathleen Womack; and my husband, Jeff, and children, Mikah and Miles. Without the support of these individuals, this project would not have been possible.

This research was funded in part by a Clan Donald, USA Academic Scholarship and a Carroll D. Clark award from the University of Kansas, Department of Anthropology.

Table of Contents

List of Figures

List of Tables

Chapter One: Introduction

> "*Somewhere within the depth of all of us who are of Scottish blood, there is a knowledge that despite our dispersion throughout the continents and our constantly increasing assimilation into other nations, we are still somehow one people, held together by fragments of a common culture and genes inherited of ancient kings…Cuimhnih air na daoine o'n d'thaing thu-Remember where you came from…*" (Bruce, 1996)

Scots, or any individual born in Scotland, began to migrate to the New World roughly four hundred years ago, and have continued to contribute to the American gene pool (Fogleman, 1992). The United States Census Bureau reported that Scottish ancestry was claimed by 5 million (1.7 percent) Americans in the 2000 Census. Another 4.3 million (1.5 percent of the U.S. population) claimed a Scotch-Irish ancestry, a term usually indicating descent from the estimated 100,000 Scots that left the Scottish lowlands for the Irish Province of Ulster during the 1600s, and began to migrate to North America some five generations later (Houston, 1996). There are an estimated 24 million (estimated from D.A. Bruce, 1997) or so individuals around the world who claim Scottish ancestry but yet have never seen Scotland itself. Despite the genetic admixture that has occurred with the spread of "Scottish" genes throughout the globe, many individuals still identify strongly with their Scottish heritage.

The concept of ethnicity has long been a subject of study, particularly research in self-identity within societies made up of multiple ethnic groups (Eriksen & Nielsen, 2001). This point is underscored by Ned Landsman (1999), who writes,

> that (the) concept of ethnicity [is] in fact derived not from the experiences of groups arriving in the Americas during the seventeenth and eighteenth centuries but rather their much later followers. It emerged principally to highlight those traits that groups managed to retain during the seemingly inevitable process of adapting themselves to a new American nationality.

Levi and Dean (2003) discuss this further, stating "ethnic groups are not given naturally…but rather are products of structure and agency, forged in the crucibles of culture and history." In other words, ethnicity or an individual's ethnic group, is a complex construct which contains a sum of cultural, political, economic and biological factors in its usage. In the 1980 census, the U.S. began to ask individuals of their ancestry, instead of the place of their parents' birth. As ethnic boundaries have blurred over generations of admixture, this question has become more and more complicated. In the United States 2000 census, individuals were allowed to write in one or two ancestral or ethnic origins. The U.S. Census Bureau defines ancestry as a "person's ethnic origin, heritage, descent, or 'root,' which may reflect their place of birth, place of birth of parents or ancestors, and ethnic identities that have evolved within the United States" (Brittingham & de la Cruz, 2004). While this is an improvement in previous attempts in understanding the cultural identities of the American population, it still does not allow an individual to express what may be a more varied and complex ethnic background. And, as an individual can identify only

two of what may be a more complex ancestral background, there is an underestimate of individuals that would claim a Scottish identity. With only two choices available, how does an individual chose their identity?

While this thesis cannot examine the question of why an individual identifies with their Scottish heritage, it can ask how this identity relates to an individual's genetic makeup. Two major trends can be detected in the historical records. The first is that Highlander migration was dominated by single males, although families and single females participated in this movement as well. The second trend is that a majority of the New World immigrants came from the Western Islands of Scotland and from the mainland region of Argyll on the West coast of Scotland or the Central Highlands of Inverness. If we assume that an individual will chose an identity that is handed down from a maternal or paternal line, then the individual should have "Scottish" maternal or paternal markers. This work sets out to examine identity through mitochondrial DNA (mtDNA) and Y chromosome markers, and will address the question: do individuals who claim Scottish identity have maternal or paternal genes associated with the Scotland or are they claiming Scottish identity based on other ancestral or cultural associations?

The goals of this study are 1) to characterize the mitochondrial and Y-chromosomal genetic variation in a group of individuals that identify themselves as Scottish; 2) to compare these individuals, and the group as a whole, to populations within Scotland; 3) to compare these individuals to other European populations that

have contributed significantly to the current gene pool of the United States. These goals will be met to answer the following questions:

1. *Do individuals who claim Scottish ancestry have maternal or paternal markers associated with Scotland?*
2. *Do the Scottish-American individuals more closely resemble other European populations that occupied the New World*?
3. *Which markers, mitochondrial DNA or Y-chromosome, are better indicators of a Scottish ancestry?*

This thesis is divided into five additional chapters. Chapter two describes the movement of Scottish people to the New World, and gives brief history of Scotland and its genetic variation. Chapter three provides the sampling methods used in this study, the populations chosen for comparison, and the laboratory and statistical methods. The forth chapter describes the results of the laboratory and statistical methods used. A discussion on the interpretation of these findings is provided in chapter five, and chapter six forms the conclusion of this study.

Chapter Two: Literature Review

This chapter is divided into two parts. The first section describes the major movement of Scots into the American colonies and the later United States of America. The second part defines what it is to be a Scot based on the mitochondrial and Y-chromosome markers used in this study. These markers should also be found in the Kansas City Games group if they have a maternal or paternal Scottish ancestry.

Scots in America:

Immigration to the Colonies during the 17th and 18th Centuries

Early immigration numbers and rates from Scotland to the United States have been difficult to accurately assess. Standardized records of early immigration were not kept, but some records are located in various writings and dispersed in both published and unpublished reports. It is known that Scottish immigration to the New World began in the 1600s (Fogleman 1992; Basu, 2007). During this pre-industrial era, immigration occurred in small waves, with only a few hundred Scots settling in the New World between 1629 and 1632 (Withers, 1988; Landsman, 1999). In 1696, an attempt at systematic settlement was organized by the *Company of Scotland* as a Scottish bid at major colonization of the New World. The attempt, aimed at Spanish territories in Central America, was a disaster. Over a quarter of Scotland's liquid assets were lost and hundreds of Scots died from disease, accidents at sea, or

interactions with the native people and Spaniards (Magnusson, 2000; MacLean, 2009). However, a small number of migrants, some 6,000 individuals, made their way from Scotland to North America by 1700 (Houston, 1996).

A shift in the economic structure of Scotland led to a breakdown in the traditional clan system, where a group of people were historically tied to a chief, that had defined the Scottish political system for centuries (Scottish Government, 2004; Magnusson, 2000). The word clan comes from the Gaelic term c*lann*, meaning children or descendants, and the clan system provided protection for the people who pledged their loyalties to a particular chief (Scottish Government, 2004). A shift away from this system led to a slow increase in emigration. For example, as opportunities on Scottish soil dissipated, the Union of the Parliaments in 1707 provided Scots more opportunities for migration to the New World (Magnusson, 2000). Between 1735 and 1754, small groups of Scottish Highlanders were being recruited as laborers for settlement in the new American colonies. An example of this movement is the recruitment of 170 men, women, and children from the central Scottish Highlands who were brought to Florida and Georgia in 1735. Other groups from the central Highlands would follow in 1737 and 1741 (Murdoch, 1998). Another organized attempt at settlement brought individuals from the Scottish island of Islay to work for Captain Lachlan Campbell on lands in New York. After Captain Campbell failed to obtain permission to work on the land he had thought to acquire, the Scottish migrants dispersed to find land and work of their own (Murdoch, 1998). In 1739, North Carolina would receive another influx of immigrants from the Scottish

Highlands of Argyll (Murdoch, 1998). Fleeing Scotland as refugees following the Jacobite uprising of 1745-1746, more Highlanders travelled directly to North Carolina (McDonald & McDonald, 1980). However, this short burst of immigration from the Scottish Highlands was halted with the English threat of war with the French and the Spanish (Murdoch, 1998).

Small groups from the Lowlands of Scotland were leaving for the American colonies up until the 1760s. While families did travel to the Americas, the emigrants were largely made up of young men who tended to move directly to commercial centers after arriving in Philadelphia (Landsman, 1999; McDonald & McDonald, 1980). Between 1763 and 1776, with the threat of war abated, an increase of Scottish families from both the Lowlands and the Highlands began to immigrate (Landsman, 1999; Murdoch, 1998). Records from this time period show that Scottish highlanders

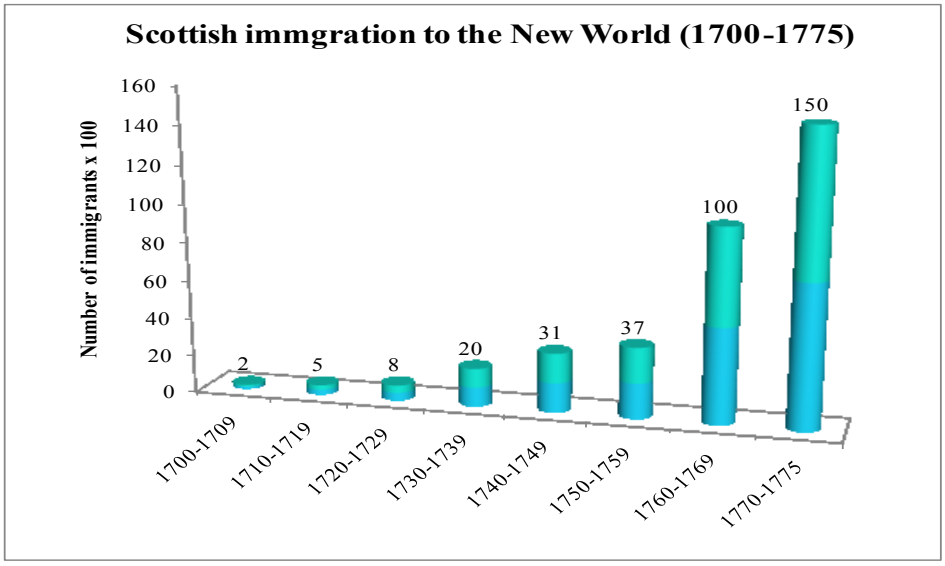


Figure 1. Estimated Scottish immigration to the New World between 1700-1775, by Fogelman (1992). Immigration numbers are in the hundreds.

began to form settlements in the state of New York, and continued to occupy the interior portions of the North Carolina colony (Landsman, 1999; Murdoch, 1998). The movement into North Carolina was recorded by the Governor in a report to London that states that over 1,600 immigrants had arrived from Scotland in 1770. Not only was this movement large by early settlement standards (see Figure 1), but it was also from a very specific region of Scotland. Nearly all of the 1600 immigrants had come from the islands of Argyll. By 1771, individuals from other regions of the Highlands, including the Isle of Skye, began to head to the American colonies as a part of organized efforts (Murdoch, 1998).

Scots were settling abroad in increasing numbers by the 1700s, but they were not approaching the numbers of Scotch-Irish individuals that had made their way to the New World. The term Scotch-Irish usually refers to individuals of Scottish decent that had moved to the Ulster Plantations of Northern Ireland during the 1600s. After 1715, these Scotch-Irish immigrants began to arrive in the American colonies in significant numbers (McDonald & McDonald, 1980). The Scots of Ulster were the descendants of Scottish individuals, particularly of clans on the Scottish-English border, who had settled in Ireland in the 1600s. Estimates of the number of Scotch-Irish individuals in the colonies before 1776 are put at 250,000 (McDonald & McDonald, 1980). Other estimates of Scottish *versus* Scotch-Irish immigration based on surname frequencies place the numbers of immigrants from Ulster at 114,000 individuals compared to the 62,500 immigrants from Scotland (Purvis, 1984). However, emigration from all parts of Great Britain would halt in 1776 with the

American Revolution, and a British prohibition of emigration to the colonies (Murdoch, 1998).

When the Revolutionary War ended, emigration from Scotland and Ireland resumed. From 1785-1791, at least five ships full of immigrants arrived in North Carolina, leaving Scotland and Ireland in response to poor harvests and increasing rents (Murdoch, 1998). By 1790, individuals of Scottish ancestry were making up a significant portion of the American population with about 24 percent of all European immigrants claiming Scottish ancestry (See Figure 2) (Fogleman, 1992). The census of 1790 shows that there were particular areas where Scots tended to settle. In New England, where approximately one-eighth of the populations were of Scottish ancestry, Scots settled preferred the upper portions of the region over lower New England. From Pennsylvania southward, some estimates indicate that up to a forth of

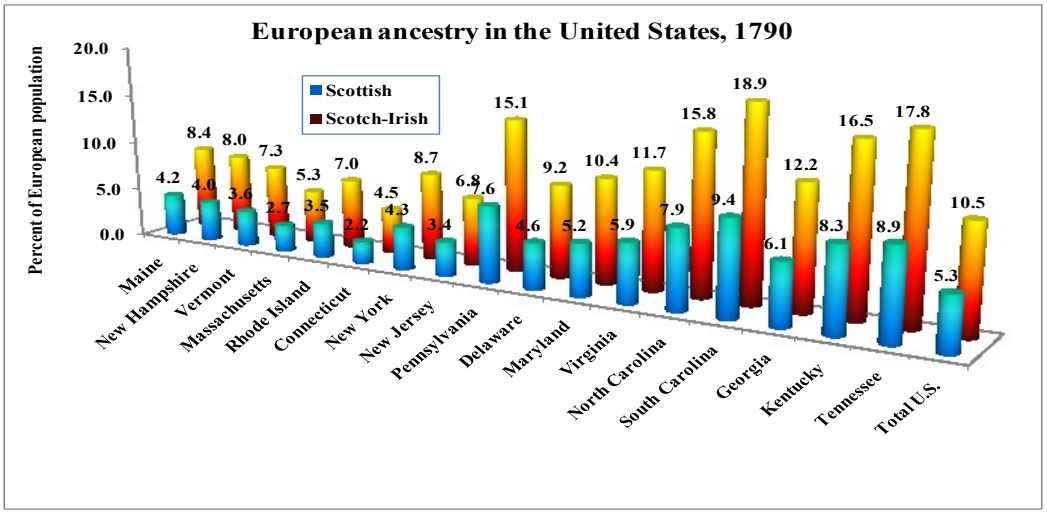


Figure 2. Estimated percent of European individuals of Scottish and Scotch-Irish ancestry in the United States, 1790, based on surname analyses and historical records (Purvis, 1984).

the population was made of Scots (McDonald & McDonald, 1980). By 1801, nearly 9700 individuals from the Isle of Skye had left Scotland for the United State. Steady immigration continued until the 1803 British Passenger Act Legislation made it more difficult for individuals to afford passage to the U.S. (Murdoch, 1998).

A New World: Potatoes, Capitalism, and the Highland Clearances

The Highland Clearances occurred between 1790 and 1855, and were the result of a shift in the political and economic structure of Scotland. The nobility of the eastern and southern portions of the Highlands were acting under the idea of improvement and were altering their lands from small tenant farms to large-scale sheep pastures. This transition meant the permanent removal, or forced seasonal labor movements of the families and individuals who had run these small farms for generations (Landsman, 1999; Newby, 2007; Symonds, 1999; Withers, 1988). Seasonal labor allowed many Scots a way to provide for themselves and their families. By the 1800s, many Highlanders were moving to obtain agricultural employment, work in fisheries, or work in the industrial towns of the Lowlands (Devine, 1979).

The transition to large-scale sheep pastures was also accompanied by an increased reliance on staple crops. By the early 1800s, the potato had become a major component of the Scottish diet (Withers, 1988). This dependence on the potato as a dietary staple was particularly the case in the heavily populated northwestern

portions of Scotland, where over half of the land was used for potato cultivation, and in some locations contributing up 80 percent of all nourishment (Gray, 1955; Withers, 1988). The reliance on potatoes also led to an increase in population size, despite earlier depopulation through permanent out-migration, persisted. This was particularly common in areas where land had not been turned over to sheep pastures and crofting tenants, small scale farms for subsistence, were maintained (Gray, 1955; Withers, 1988; Devine 1979). Between 1801 and 1841, the total population of Scotland had increased by 7 percent, with some regions increasing in population size by 53 percent (Gray, 1955).

Economic conditions worsened with increasing population size (Young, 1996). Two episodes of crop failures in 1836 and 1837 led the government and charitable groups to organize relief efforts to help the hard hit Scottish Highlands.

Figure 3. Scottish Highland and Lowland boundary with regions mentioned in this study labeled. Modified from Wikimedia.org.

Over 100,000 people, mainly small holding farmers, were at risk of starvation during this crisis despite migration efforts to control population size (Devine, 1988). The situation worsened in 1845, when potato blight destroyed Scotland's staple crop and continued to have an impact on the harvest until 1857. In 1846, 76 percent of crofting districts and 59 percent of the farming districts reported total crop failures (Devine, 1988). The potato blight would only accelerate the clearance process that had already been underway since the 1790s (Devine, 1979).

While many areas of Scotland were experiencing this emigration, the Scottish Highlands were particularly affected, where up to a third of its entire population left for opportunities overseas or in industrial towns in the Scottish Lowlands or England (Basu, 2007). The United States was the most popular destination and received over half of these emigrating individuals, with many Scots settling in New York (Landsman, 1999). However, it was not until 1873 that British ships began to record

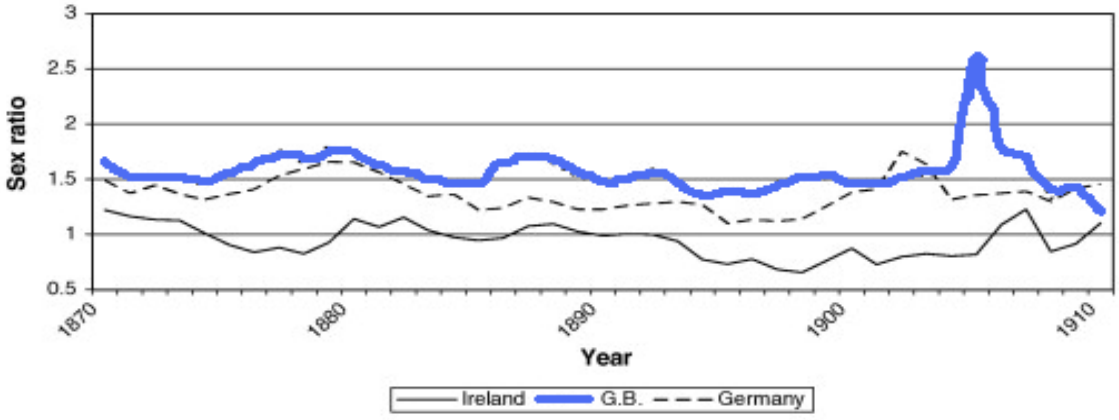


Figure 4. Male to female sex ratio of immigrants to the U.S. from 1870-1910. G.B. represents immigration from England, Scotland and Wales. Image modified from Greenwood, 2008.

the origin of individuals travelling from a British port to the United States (Cohn, 1995). A U.S. law passed in 1819 required that all ships arriving in a U.S. port had to present information on a passenger's name, age, gender, occupation and place of national origin. These U.S. Passenger Lists have been the primary source of information for researchers trying to understand North Atlantic immigration to the U.S. from 1819 to the 1870s. Records from U.S. Passenger Lists from 1836-1853 showed that young, single males made up a large component, some 55-70 percent, of all males leaving English ports, although the records do not differentiate between Welsh, English or Scottish passengers (Cohn, 1995). There is also an observed difference in the sex ratio of males *versus* female immigrants from England at this time, with men outnumbering female immigrants. Figure 4 shows the differences in the sex ratios of immigrants to the United States between 1870 and 1910. Over this time period, males outnumbered females at an approximate 3:2 ratio (Greenwood, 2008). These and other historical documents indicate that roughly 2.33 million Scots left the British Isles between 1825 and 1938 (Szasz, 2000; Symonds, 1999).

The tremendous decline in the Scottish population can still be seen today, as the highlands remain one of the least densely populated areas in Europe (Basu, 2007). Out-migration from Scotland, particularly from the Scottish Isles, continued to be larger than in-migration up until the 1990s (See Figure 5). This recent out-migration, while still including migrants to the United States, is largely made of movements to other parts of the United Kingdom (General Register Office for Scotland, 2005).

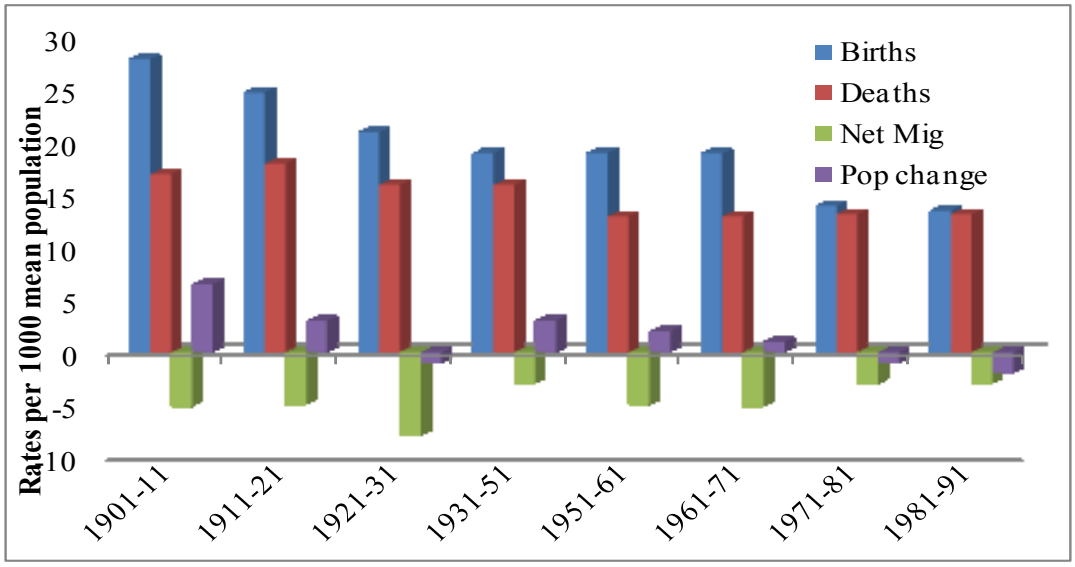


Figure 5. Birth, deaths, net migration, and total population change rates adapted from Anderson (1996).

What Makes a "Scot"?: A Review of Scotland's History and mtDNA and Y-chromosme Markers

The Peopling of Scotland:

During the Last Glacial Maximum (LGM), approximately 20,000 years ago, the British Isles were still connected to the European continent (Adams & Faure H, 1997). Present-day Scotland was covered by ice and, unlike the rest of Europe, no evidence of anatomically modern humans have been found that date before this time (Stringer, 2006). As temperatures rose and the ice sheets began to recede, the first anatomically modern humans started to move out of the Iberian *refugia*, and eventually occupied regions along the Scottish coast. Located in the southern regions of Britain, Gough's Cave contains the oldest evidence of modern humans, with materials dating to around 15,000 years ago (Stringer, 2006; Sykes, 2006).

The earliest archaeological evidence of human occupation in Scotland is found on the Northwest coast and dates to about 10,000 B.P, (Sykes, 2006). Other human activity in Scotland, as well as in Ireland, can be found in coastal areas as well, with stable isotope analyses indicating an exploitation of marine resources (Waddington, et al., 2003; Schultting & Richards, 2002). Food resources would have been abundant in this environment, but foraging groups rely on a greater amount of land to sustain a population. Estimates of population size of the British Isles reconstructed a population of about 2,750-5,500 inhabitants (Hunter & Ralston, 1999).

The Introduction of Farming to Scotland:

The Neolithic revolution brought agriculture to Europe around nine thousand years ago and spread along the same routes that the first anatomically modern humans used during the Paleolithic expansion. It would, however, take another three thousand years (around 6,000 B.P.) for agriculture to reach Scotland (Weale et al., 2002). On the southwest coast of Scotland, the transition to agriculture can be seen throughout the archaeological record from signs of woodland clearances and an increase in the frequency of cereal-type pollens between 5,000 and 4,700 years ago. This experimental phase of agriculture ended around the same time that the region experienced a major drying phase that provided a more stable farming environment (Macklin, et al., 2000; Bonsall, et al., 2002). There has been much debate on the origins of farming in Europe, and whether it was a strictly cultural spread of ideas or

whether it was accompanied by the spread of people. In short, the genetic and archaeological record shows that people did move with agriculture, but this influence decreases the further west people had to travel from the Fertile Crescent. For Scotland, this transition is thought to have been a gradual one, with the majority of the descendants of the orginal foraging populations and the addition of small groups of individuals whose ancestors originated in the Near East (Oppenheimer, 2006). Estimates of a shift from hunting and gathering to agriculture can lead to a 5-50 fold growth in population density within the first thousand years, and the European population would expand sharply after this cultural transition occured (Novelletto, 2007).

The Picts and the Scotti:

The coming of agriculture meant a more stable food source for populations of the North Atlantic, allowing for population growth and the appearance of more complex societies with complex technology. The archaeological record gives evidence for burials and ritual monuments, tools made of bronze, and changes in pottery. By 2500 B.C.E., the first of the region's henges, or ancient monuments, had been built out of timber (Hunter & Ralston, 1999). Most of the changes during this time period seem to be of a cultural nature, with little evidence of large movements of people into Scotland or the rest of the Britain Isles. These earlier groups would give rise to the human groups known as the Picts and the Scotti (Magnusson, 2000; Oppenheimer, 2006).

Debate on the origin and disappearance of the Picts has been continuing for decades. There are no written Pict texts, little surviving mythology, and the original Pictish language is unknown (Sykes, 2006). The name 'Pict" originates from the word *Picti*, meaning 'the Painted People,' an ethnonym the Romans often gave to local groups who decorated their bodies with tattoos or body adornment (Sykes, 2006) . This group labeled Picts is now thought to represent a mixture of foraging and farming peoples during the Neolithic, the native population of Scotland that had gradually assimilated to the various cultural influences to which they have been exposed (Sykes, 2006; Magnusson, 2000).

Several waves of immigrations from Ireland also influenced and shaped the genetic topography of Scotland . The name of Scotland can be attributed to the Irish migrations to region, since the *Scotti* were Romans references to Irish immigrants (Sykes, 2006). In the beginning of the first millennium, the Irish king, known as Dal Riata, set up three colonies in present day Scotland. Between 500 and 1000A.D., Dunadd was a capital of the Dalriada kingdom and a major trading hub which facilitated trade from Ireland to the Mediterranean (Magnusson, 2000). The Picts regained control of the region in the 6th century, but the balance of power would go back and forth between the Picts and the Gaels for hundreds of years (Sykes, 2006). No written records exist about the genetic interaction and possible fusion of the original inhabitants with the incoming Irish settlers (Altson, 1999). However, as both groups date back to the initial mesolithic, with later neolithic influences, they would have had similar genetic backgrounds (Sykes, 2006).

Figure 6. The Dal Riada kingdom (Wikimedia.org).

Invasion and Occupation: Romans, Vikings, Normans and Anglo-Saxons

Other groups have influenced the cultural and genetic makeup of Scotland and the British Isles over the next few thousand years. Roman influence in the British Isles was felt long before the occupation of Britain had occurred. In the lowlands of the southeast, Roman coins are found in the archaeological record, an indication of economic ties to Rome (Kearney, 2006). In 43 A.D. a massive invasion of Britain began. By 79 A.D., southern Britain had fallen under Roman control, leaving the Lowlands and Highlands of Scotland as the only unoccupied territories (Magnusson, 2000). In an attempt to rule the whole of Britain, Julius Agricola, the Roman Governor of the Province of Britannia, launched a full scale invasion of Scotland in 83 A.D. (Magnusson, 2000). This attempt failed, and after more than a century of unsuccessful attempts, and with military expenses greater than any other province

within the empire, the Romans finally withdrew in 214 AD from Scotland (Kearney, 2006).

Angles from northern Germany began to arrive in southeastern England during the 4th century to assist the Romans in their attempts to conquer the island (Magnusson, 2000). By the 5th century, Angles were in England for their own conquests, and would eventually conquer most of eastern Scotland. Once again, there was not a replacement of the population by the invaders, but a battle of dominance which would go back and forth for a few hundred years (Oppenheimer, 2006). The Angles controlled Northumbria, and between 653 and 685 A.D. they began to occupy portions of southern Scotland (Magnusson, 2000). At the end of the seventh century A.D., the Angles invaded Scotland even further, but were cast off by the Picts (Magnusson, 2000). While Anglo-Saxon invasion can be seen through cultural, economic, and political influences, there were no large-scale movements of people into Scotland (Weale, et al., 2002).

In the 8th century A.D., improved seafaring technology, a growing population, and political changes led Scandinavian Vikings to initiate their search for additional lands and resources (Helgason, et al., 2000). The Viking age in Scotland lasted for 400 years and had different effects on the mainland and the various islands through which the Vikings made contact with. The first islands to be colonized were Shetland and Orkney around 780 A.D.. Signs of this colonization can be seen in the Norse place names still in use today (Magnusson, 2000). Norm, the Norse dialect, survived in Shetland and Orkney until the 19th century (Sykes, 2006). Genetic, archaeological

and linguistic evidence all point toward the strongest Viking influence in Orkney (Helgason, et al., 2001). Eventually, Orkney, productive for agriculture, would become an a political and strategic center for managing Viking activities throughout the Atlantic (McGovern, 1990). Archaeological excavations in Orkney suggest that this take over was a peaceful affair (McGovern, 1990). The Western Isles were also influenced by the Vikings. Here the two groups lived together, spoke both Norm and the indigenous tongue, and shared cultural traditions (Sykes, 2006). However, most Viking interactions were not peaceful, and evidence from the Hebrides points to a violent conquest followed by population replacement (McGovern, 1990). In the Scottish mainland, constant threats of Viking invasion led to the unification of various Scottish tribes (Sykes, 2006). Viking influence over the Scottish Isles would continue until the mid-13th century (Kearney, 2006).

In sum, the initial settlement of foragers some 10,000 years ago, and the possible movement of people with agriculture around 6,000 years later, make up the foundation of the Scottish population. Several other groups have also brought new people to the region in the last few thousand years, including the Romans, Irish settlers, and Anglo-Saxon and Viking conquests. A timeline of these influences is shown in Figure 7.

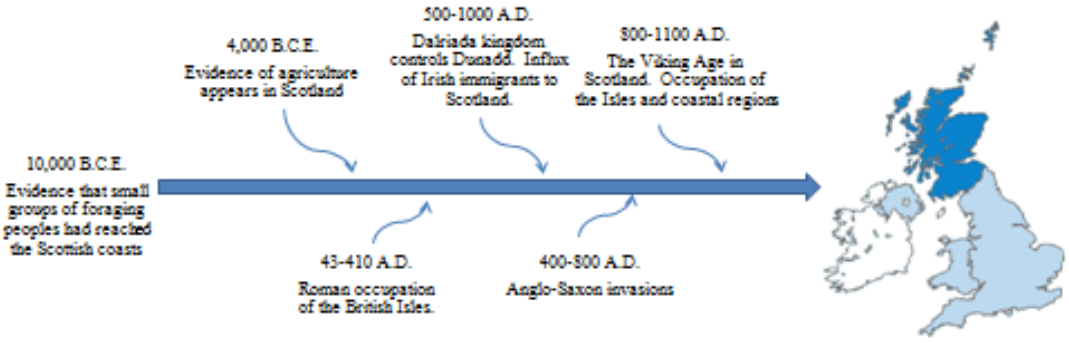


Figure 7. Timeline of Scottish settlement and invasion.

Migration, War, and Disease:

The movements of people into Scotland would largely be seen following the LGM up until around the 1800s when Irish migrants came following the Potato Famine. These movements would bring new genes to the region, increasing the diversity of genetic markers, while and the massive out-migration of the last few centuries decreased the genetic variability that may have been present in the Scottish population. Not only did Scots move to other regions of the United Kingdom and the United States, but they would also immigrate to Canada, Australia, New Zealand, South Africa, and other regions of Europe (Basu, 2007). Migration has undoubtedly shaped the genetic landscape of Scotland, but there are other events that have also played a role.

Disease and famine has long played a role in shaping the cultural and genetic structure of human populations (Hatcher, 1996). For the purpose of this study, these

two factors played a role in decreasing the genetic diversity of the Scottish population both before massive out-migration began, and before the sampling of individuals that represent the Scottish comparative population were obtained. Before the large scale migrations to the New World began in the 1700s several waves of disease would affect the genetic diversity of Scotland. The 'Black Death' of 1348-9 was one of the most devastating epidemics in recorded history. While records of mortality during this time period was scarce, a rough estimate of 30-45 percent of the English (including Scottish) population fell to the first wave of plague epidemics. Several other waves would appear in 1361-2, 1369, and 1375 (Hatcher, 1996). The last serious outbreak of plague was recorded in 1645-9, where it was particularly felt in Scotland's urban population (Houston, 1996). Smallpox and typhus was also present in the 1600 and 1700s, and would affect the genetic diversity of the Scottish population (Houston, 1996).

Disease, famine, and war following the 1700s included epidemics of smallpox, dysentery, typhus, measles, influenza, and two World Wars (Anderson, 1996; Winter, 1977). Famines in the 1690s may have led to the deaths of 10 percent of Scots, and the Potato Famine of the 1840s played a role in increasing rates of Scot emigration to the New World and Irish immigration to areas of Scotland (Anderson, 1996). World War I would also decrease the genetic diversity of Scotland, particularly that found in the Y-chromosome. Some 690,000 Scots, 41 percent of Scottish men between the ages of 15-49, participated in the war. At least 12 percent of these men died in service (Winter, 1977). The influenza pandemic of 1918-1919,

had the highest mortality of any infectious disease in Scotland's history, with over 17,500 deaths attributed to influenza or complications arising because of influenza (General Register Office for Scotland, 2005). About 265,000 men died in service in the Second World War (Anderson, 1996). These are just a few examples of historical events that have shaped the genetic landscape of Scotland.

Mitochondrial DNA and Y-chromosome Variation in Scotland

Two portions of the human genome have been used extensively to understand the movement of humans throughout history. Mitochondrial DNA (mtDNA) is passed on through a maternal mode of inheritance so that every daughter will carry an exact copy of her mother's genome, and in turn will pass it on to her offspring (Giles, al., 1980). In the same fashion, a large portion of the Y-chromosome does not participate in recombination and is inherited in a paternal mode (Novelletto, 2007). Both of these markers have allowed population geneticists to develop a male and female human history. Comparative data have been compliled for mtDNA and Y-chromosome markers, making them useful tools, for identifying the region of the world where an individual's ancestors originated.

Mitochondrial DNA (mtDNA) is a circular molecule of 16,569 nucleotide bases that does not recombine in humans, making it identical by descent (Cann R. , 1988). The molecule lies outside of the nucleus within the mitochondria, and hundreds to thousands of copies of mtDNA are found in each cell (Rubicz, et al., 2006). Most of mtDNA is coding DNA, meaning that it provides instructions to build

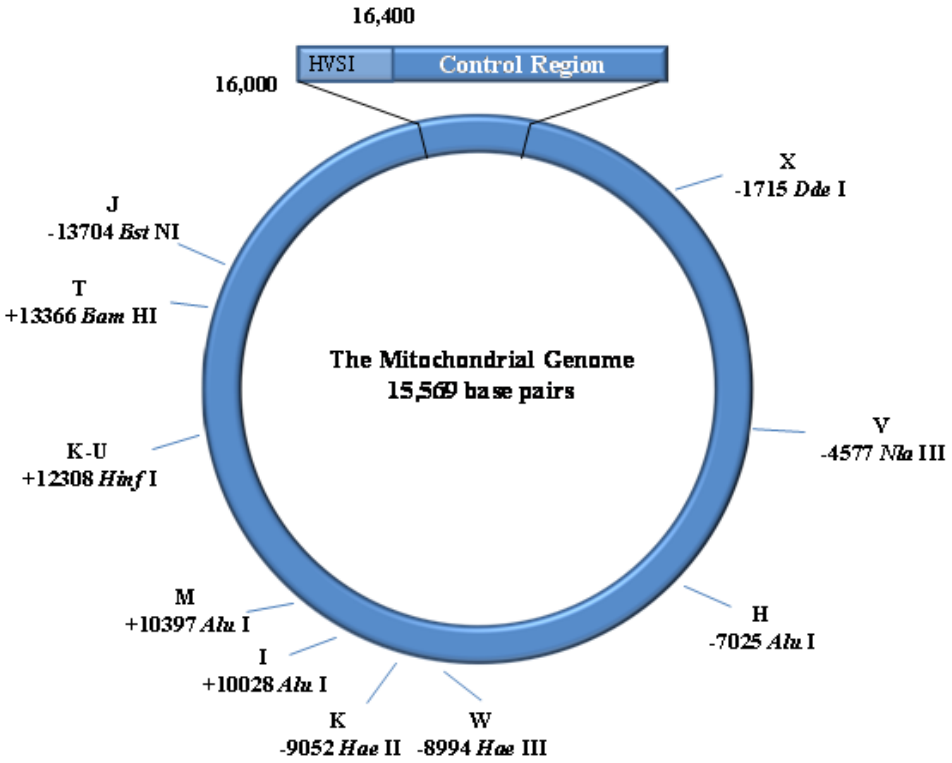


Figure 8. The human mitochondrial genome. Letters along the outside indicate haplogroups with defining enzyme cut sites following (Rubicz, et al., 2007).

a particular protein (see Figure 8). The coding DNA of the mitochondria has a rapid mutation rate of 0.017 x 10^{-6} substitutions per site per year, exceeding that of nuclear DNA by 5-10 times (Pakendorf & Stoneking, 2005). These mutations are usually single base substitutions, where repair mechanisms fail to catch the error, and are shown in Figure 8 (Cann R. , 1988). These types of mutations have been used to categorize the worldwide mtDNA diversity into haplogroups, or groups of sequences that have similar evolving mutations due to a common ancestor (Jobling, et al., 2004).

Further investigation of this genome has provided researchers with sites on the chromosome that mutate at an even faster rate, giving geneticists information on

historical events in human history. For this, there is a non-coding fragment ~1100 base pairs long that codes for regulatory functions in the mtDNA genome called the control region (CR), or d-loop in Figure 8 (Pakendorf & Stoneking, 2005). There are two commonly sequenced hypervariable segments (HVS) in the CR which mutate at a relatively faster rate than the rest of the genome, and have been termed 'mutational hot spots' (Stoneking, 2000). Estimates of mutation rates at the HVS currently vary based on phylogenetic comparisons (0.075-0.165 x 10^{-6} subs/sites/yr) or pedigree analysis (average 0.47 x 10^{-6} subs/site/yr) (Pakendorf & Stoneking, 2005). Meyers et al. (1999) surveyed patterns of nucleotide substitutions in hypervariable regions, and using a maximum likelihood approach, found that HVS I has a mutation rate twice as high as HVS II. This information has been used to further classify the mtDNA genome into subclades, some of which are specific to a particular region of the world.

In Europe, mitochondrial studies have identified ten haplogroups using the slowly mutating regions within the coding region of the chromosome. These haplogroups characterize nearly all European genomes and include H, U, T, K, J, X, I, V, W and M (Finnila, et al., 2000). These haplogroups are although dispersed throughout the European continent. There are, however, differences in the frequencies of haplogroups and their subclades throughout Europe (Achilli, et al., 2004; Helgason, et al., 2001; Torroni, et al., 1998; Torroni, et al., 2001; Torroni, et al., 2000). As a European populations, Scotland's mtDNA is represented by these same, haplogroups with 99 percent of the lineages belonging to H, I, J, K, T, U, V, W, and X (Sykes, 2006).

The diversity in mtDNA decreases with latitude in Europe, and the farther northwest the population, the higher the frequency of haplogroups associated with the expansion out of the Iberian refuge following the LGM (Gonzalez, et al., 2003). This expansion brought some the oldest and most common haplogroups to Scotland (Oppenheimer, 2006). MtDNA lineages that are associated with post-glacial expansion are haplogroups V and H and U5 (Torroni, et al., 1998; Gonzalez, et al., 2003). Haplogroup U appears to have been brought in by Europe's first settlers. This haplogroup is found mostly in Europe at a frequency of about 7 percent, with an occasional appearance in sub-Saharan Africa (Finnila, et al., 2000). A subhaplogroup, U5, dates to around 50kya in Europe, and is present in Scotland with a frequency of around 8 percent (Richards, et al., 2000; Gonzalez, et al., 2003; Sykes, 2006). The most common haplogroup in Europe, haplogroup H characterizes between 40-50% of all European population's mtDNA (Achilli, et al., 2004). Haplogroup H arose in the Middle East around 30-25 kya, and probably arrived in Europe during the Paleolithic as well (Richards, et al., 2000; Achilli, et al., 2004). Haplogroup H makes up about 45 percent of all mtDNA haplotypes in Scotland and probably arrived there in several waves (Helgason, et al., 2001; Sykes, 2006). Several H haplotypes probably arrived in initial movements into Scotland, including haplotypes belonging to subhaplogroups H1 and H3 which have been shown to have spread out of the Iberian refugia (Achilli, et al., 2004). Haplogroup V dates to around 16 kya and was dispersed from the Iberian Peninsula throughout western Europe following the LGM (Torroni, et al., 2001)

Neolithic contributions can be seen in increasing frequency the farther south in the continent with high levels of diversity seen in the Iberian Peninsula (Gonzalez, et al., 2003). Genetic and archeaological evidence point to two different routes of agricultural diffusion to Scotland. Some J (including J1 and J1a), T (T1 and T2), U3, and U5a1 and U5a1a haplotypes traveled to Scotland through a continental route . Other haplotypes, such as J2, traveled to Scotland through a Mediterranean coastal route and is more commonly found in Scotland then in the rest of Britain (Richards, et al., 2000; Gonzalez, et al., 2003). Haplogroups J and T both originated in the Near East and each have a frequency of around 8 percent in Europe (Mogentale-Profizi, et al., 2001). Several studies (Richards, et al. 2000; Barbujani & Bertorelle, 2001; Chikhi, et al., 2002; Currat & Excoffier, 2005; Dupanloup, et al., 2004) have attempted to define the Neolithic to Paleolithic genetic contribution to the whole of Europe (See Figure 9). These studies indicate a Neolithic contribution of between 25-65 percent, with decreasing genetic input the further west the individuals had to travel (Novelletto, 2007). The native people of Scotland, known as the Picts, are some combination of these Paleolithic and Neolithic influences (Sykes, 2006). Maternal contributions from Ireland are harder to determine as Ireland and Scotland share many of their mtDNA lineages due the shared ancestry of initial settlers, as well as later movements from Ireland to Scotland (Helgason, Siguroardottir, Gulcher, Ward, & Stefansson, 2000). Some estimates indicate that by the early 1700s, nearly a forth of Ireland's inhabitants were 'of Scottish and English blood' (Houston, 1996). Ireland shares over 70 percent of its lineages with other populations throughout the

North Atlantic and on the European mainland (Helgason, et al., 2000). Viking ancestry is easier to examine, and in 2000, Helgason et al. (2001) determined the amount of maternal ancestry in the Scottish Isles that is shared by the Irish and the Scots verses the maternal influences of Scandenavian Vikings. They found that even in the island of Orkney, where the Vikings had their greatest influence, there was a maternal Scandinavian imput of around 38 percent with the rest of the lineages belonging to Scottish or Irish origin (Helgason, et al., 2001). This influence weakens on the coast where Scandinavian mtDNA input drops to around 14 percent (Helgason, et al., 2001). Maternal haplotypes belonging to haplogroups H and K were a part of this Scandinavian contribution (Helgason, et al., 2001). A summary of maternal contributions can be seen in Figure 10.

The Y chromosome found in males has a large non-recombining region (NRY), which is inherited in a paternal fashion. The Y chromosome consists of about

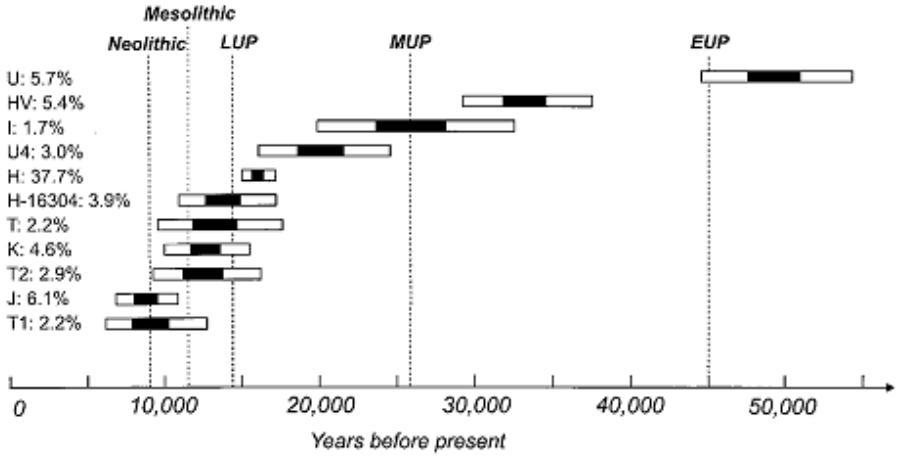


Figure 9. Major European founding lineages, frequencies, and estimated ages from Richards et al. (2000).

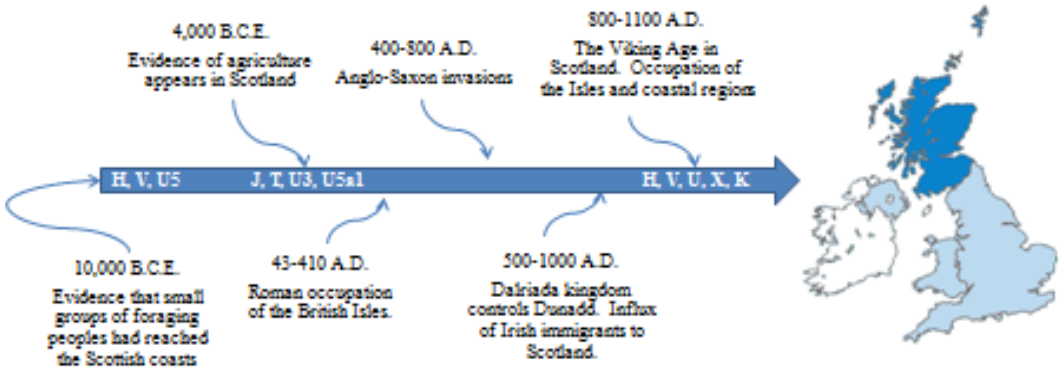


Figure 10. The arrival of the most common mtDNA haplogroups to Scotland (shown in dark blue). Haplogroups are listed in white along the arrow.

60 million bases of DNA, most of which do not code for protein (Novelletto, 2007). A combination of slow mutating single nucleotide polymorphisms (SNPs) and faster mutating segments of DNA, called short tandem repeats (STRs) are found in the NRY, and have been used to classify the worldwide Y chromosomal variation into major haplogroups and subclades (YCC, 2002; Karafet et al., 2008). SNPs are changes at a single base that mutate at the relatively slow rate of about 10^{-8} per base pair per generation (Novelletto, 2007). Because of this, SNPs are better for tracking prehistoric changes, while STRs can give information on more recent events (Roewer, et al., 2005). Mutations in STRs consist of additions or deletions of a repeated segment of DNA that occur at rates 10^{5} to 10^{6} times faster than mutations in other parts of the Y chromosome (Roewer, et al., 2005).

In Europe, differerentiation of the Y-chromosome has been found to be strongly correlated with geography (Rosser, et al., 2000; Roewer, et al., 2005).

Roewer et al. (2005) found that Y-STR diversity in Europe differs more with longitude than with latitude. Due to a low effective poppulation size of one fourth of the total population, the Y-chromosome is more subject to the effects of genetic drift (Rosser, et al., 2000). Another factor that reduces the diversity in a population's Y chromosomes is called the Genghis Kahn effect, where a dominating group of males or an individual male makes a significant contribution to the gene pool. An example of this is suggested from the political influence of Genghis Kahn's regin, where a particular Y chromosome lineage carried by 8 percent of all males in Asia (some 0.5 percent of the world's men) is believed to be the result of male dominance of Genghis Kahn and his descendants during the Mongolian Empire (Jobling, et al., 2004). Geographic differentiation combined with the Genghis Kahn effect could account for the lower levels of diversity within European populations (Novelletto, 2007).

Populations examined through both of these markers have shown that there are often different histories for males and females due to differences in male and female migration within and between populations. For example, diversity in the Y chromosome has been found to be lower within a population while being higher among populations. This differs from mtDNA and many autosomal markers that have a higher within group diversity than among group diversity (Jorde, et al., 2000; Wells, et al., 2001; Pereira, et al., 2001). In other words, many mtDNA haplogroups

and haplotypes are shared by many different populations within the European continent, while Y chromosome haplotypes tend to be geographically specific (Mielke & Fix, 2006). These differences in diversity between the two markers are often due to differences in the genetic contributions of an individual male *versus* an individual female or differences in patrilocal *versus* matrilocal patterns of residence (Novelletto, 2007).

The European Y-chromosome is largely described by five major haplogroups: R , I, J, G and E (Jobling, et al., 2004). Figure 11 shows the movement of these haplogroups into Europe and the timing of these events. In Scotland, these

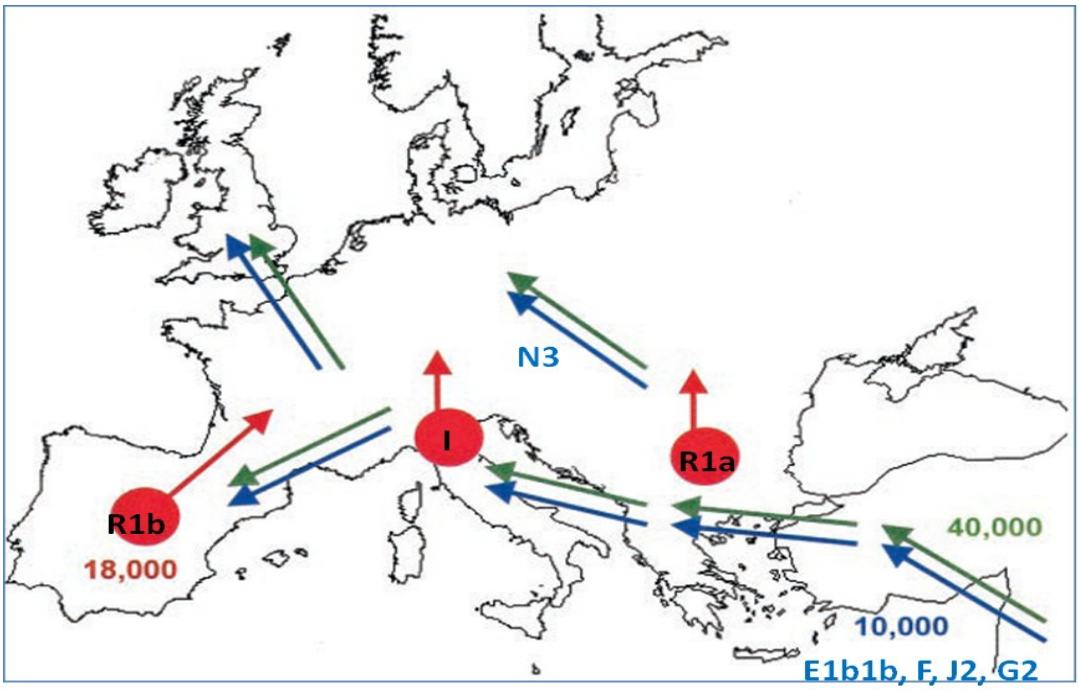


Figure 11. Movement of prehistoric Y haplogroups in Europe. Green lines indicate a Paleolithic expansion by 40,000 years BP. Red circles indicate refugia during the LGM, and the haplogroups associated with dispersal following the retreat of ice. The blue lines indicate expansion routes of Neolithic settlers that began to occupy Europe around 8,000 years ago. Image adapted from Barbujani and Bertorelle, 2001.

haplogroups are represented by R1b (73%), R1a (9%), E1b1b (2%), I (15%) and J (1%). These estimates are based on over 2,400 hundred participants to the Oxford Ancestry Project (Sykes, 2006).

Several studies of Y-chromosomal variation have found that members of the R1b haplogroups, especially those with the Atlantic Modal Haplotype (AMH), are also associated with the expansion out of the Iberian Peninsula (Capelli, et al., 2003; Roewer, et al., 2005; Semino et al., 1996). Haplogroup R1b probably arrived in these first expansions into Scotland, supported by a study by Rosser et al. (2000) that usedprincple component analysis to show that Basque are more similar to Scottish and Irish populations than to populations closer to them geographically. This haplogroup characterizes 77 percent of the Scottish population (Oppenheimer, 2006), with subhaplogroup R1b1c being the most common. Individuals with the AMH, or a one step mutation divergence from AMH (AMH + 1), make up about 47 percent of Scottish lineages (Capelli, et al., 2003).

Often described as the first peoples of Scotland, the Picts actually appear to be the result of admixture between the original Paleolithic settlers and the incoming Neolithic farmer. Because of their early roots, they are identified by some of the oldest haplogroups in Europe: the Y-haplogroup of R1b, and the mitochondrial haplogroups of H, U5, J, T, and V (Torroni, et al., 1998; Richards, et al., 2000). Sykes, in his 2006 book, *Blood of the Isles*, identifies an R1b haplotype, which he labels OGAP4, that he believes is the Pict haplotype (Campbell, 2007). Similarly,

several genetic genealogy groups now offer testing for the Pict haplotype, although what that haplotype is, they fail to disclose.

The genetic influences of Viking interaction in Scotland are evident in multiple genetic systems. For example, Poser (1994) suggests that a form of multiple sclerosis may have spread through the British Isles due to Viking influence. Similarly, variants in the phenylanlanine hydroxylase gene are believed to have arisen in Denmark and brought to England by Danish Vikings (Tyfiled, et al., 1997). Scandinavia shares several Y-chromosome haplotypes with Scotland. Most of the interactions between the indigenous populations and Vikings were in the form of Viking men mating with local women, and evidence of this is seen in Y markers (Helgason, et al., 2000; Sykes, 2006). The most prevalent Viking Y-haplotypes are those which belong to haplogroups I1a and R1a, although haplotypes belonging to

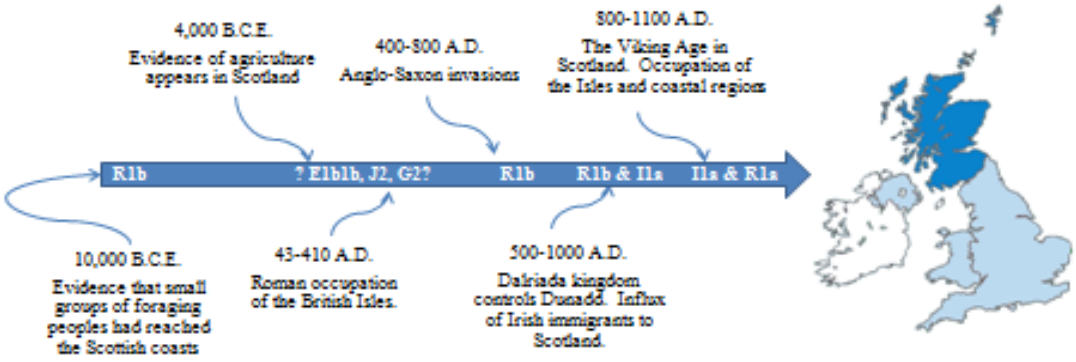


Figure 12. Arrival of the major Y-chromosome haplgroups to Scotland. Haplogroups are listed in white along the arrow, with question marks around haplogroups that entered during either during the Neolithic or were brought to Britain during Roman occupation or later movements.

R1b are represented as well. which point to a Viking component of 42% in Shetland and 37% in Orkney

Other groups may have contributed to the Scottish gene pool as well. While Roman occupation in Scotland never held, some genetic influence has been felt in the region. There are a few Y chromosome haplotypes that may have spread into the English population through the Roman army, including haplotypes from haplogroups J, G, and possibly E. These haplotypes probably made their way into Scottish populations through later migrations from Britain (Bird, 2007). All of these haplogroups are present in low frequencies in Scottish populations. Signatures of Anglo-Saxon genetic contribution, particular in Y haplogroups R1b (in particular R1b1c9) and I1a are seen in Scotland, but due to genetic similarities and shared haplotypes between the two areas, the amount of influence is difficult to determine (Faux, 2008).

All of these movements of people, from the first Mesolithic settlers to the historical invasions and occupations, brought different mitochondrial and Y-chromosome haplogroups to Great Britain and Ireland. However, the distribution of these haplogroups differ in their frequencies and haplotypes found in each region due to gene flow, during prehistoric and historic movements of people, and genetic drift in small founding populations or as a result of disease and war. Figure 13 illustrates the variation in the frequencies of major haplogroup clusters in Britain and Ireland.

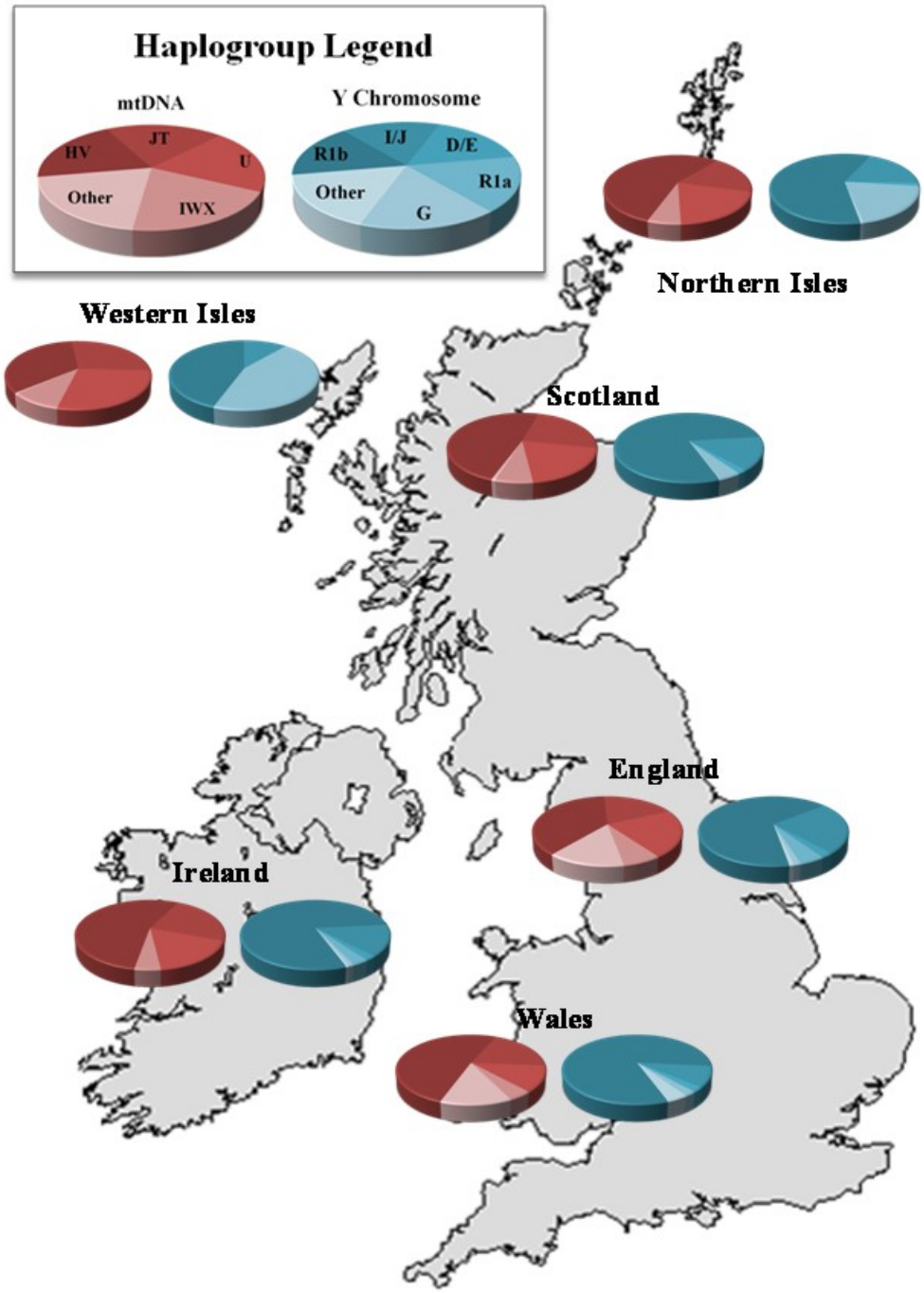


Figure 13. Mitochondrial DNA and Y-chromosome superhaplogroup frequencies for the United Kingdom. Map was created at http://www.aquarius.geomar.de/cgi-bin/map-cgi.pl. Mitochondrial haplogroup frequency data for Ireland (Helgason, et al., 2001), England and Wales (Simoni, et al., 2000), Northern and Western Isles, and the Mainland (Scotland) from (Sykes, 2006). Y-chromosome haplogroup frequencies for England and Wales (Campbell, 2007), the Northern Isles, Western Isles and Scottish Mainland (Goodacre, et al., 2005), and Ireland from Europedia (http://www.eupedia.com/europe/european_y-dna_haplogroups.shtml).

Chapter Summary

Immigration to the New World colonies began in 1600s from the Lowlands and the Highlands of Scotland as well as Scots from the Irish plantations of Ulster. In the Highlands, some areas were impacted more than others, with significant movements of people to the New World in groups originating from the Western Isles, including the Isle of Skye and the islands of Argyll, and groups from the Northwest Coast of Scotland. Highlanders and Lowlanders contributed significant numbers to the population of the colonies, and the later population of the United States. However, the majority of individuals of Scottish decent are the Scotch-Irish from Ulster. There are also significant differences in the numbers of single male *versus* single female immigrants to the U.S. Males outnumbered female immigrants from Scotland throughout the 17th, 18th, and 19th centuries, as the demand for female labor in textiles, agriculture, and domestic service kept many females in Scotland. However, the movement of single Irish females, including the so-called Scotch-Irish, was often more prevalent than single male Irish immigrants, and must be taken into consideration in this study. The people of Scotland have been genetically influenced by several migrations, from the first arrivals during the Mesolithic, to the later invasions of Romans and Vikings. All of these movements have had an effect on the markers used in this study. If the individuals of this study are of Scottish maternal descent, they should carry mtDNA markers belonging to haplogroups H, I, J, K, T, U, V, W and X. If their ancestry is of a paternal Scottish origin, the individual should carry Y-markers associated with haplogroups R1b, R1a, E1b1b, I and J.

Chapter Three: Materials and Methods

Sampling

Participants for this study were recruited from the Kansas City Highland Games in June of 2006. Marion L. Mealey, a graduate student in the Department of Anthropology at the University of Kansas, collected demographic information and samples of mouth rinses for this study. Each participant gave his or her signed consent that met the University of Kansas Internal Review Board's requirements. Participants also completed a brief questionnaire which included their name, date of birth, sex, birth location, first language, clan affiliation (although septs were also identified), as well as genealogical information consisting of both parents and maternal and paternal grandparents. Listed clan affiliations were grouped into the major clans recognized by the Court of Lord Lyon (http://www.lyon-court.com/lordlyon/CCC_FirstPage.jsp), the official Scottish clan registry. A distribution of these clans for individuals included in mtDNA and Y chromosome analyses can be viewed in Appendix A. Some participants also included addresses and email addresses in order to obtain information on the results of this study.

For DNA analysis, saliva was collected from each individual. Participants were given 10 mL of distilled water in a small cup and asked to swish the water in their mouths for 10 seconds before expelling the contents back into the cup. The samples were then transferred to a 15 mL collection tube for transport to the Laboratory of Biological Anthropology at the University of Kansas, where the

samples were frozen until analysis. A total of 58 males and 35 females volunteered for this study. All participants were born in the United States except for one individual who was born in Scotland. For the purpose of this study, this individual was removed from the analyses. Of the individuals sampled, 93 percent (n=86) reported an affiliation with a clan of Scotland. Two databases were used to find surnames that have a historical presence in Scotland: 1) The Scotland's People, a government run genealogical database that carries statutory registers from 1855-2006, parish records dating from the early 1500s to the 1850s, and old census records (http://www.scotlandspeople.gov.uk/) and 2) The National Trust Names website, a University College London project that examines the distribution of surnames throughout Britain in 1881 and 1998 (http://www.nationaltrustnames. org.uk/).

Comparative Populations

Comparative European populations which contained data for mtDNA HVS-I sequences and for Y-STRs were compiled from the literature. Populations were chosen to represent those groups which contributed the most reported European ethnicities to the United States of America (See Table 1). For mtDNA analysis, Scottish populations were obtained from Helgason et al. (2001) and were compared between nucleotide positions (np) 16070-16400 [GenBank

Table 1. European ethnicity as reported in the 2000 US Census. Population estimated at 281 million (US Census Bureau).

Ethnicity	Percent
German	15.2
Irish	10.9
English	8.7
Italian	5.6
French	3.9
Polish	3.2
Dutch	1.6
Norwegian	1.6
Swedish	1.4

accession numbers AY024369-AY026032]. Sequence data for European populations (np 16090-16367) were compiled by McEvoy et al. (2004) (http://www.gen.tcd.ie/molpopgen/data.htm). Y-STR data from Scotland were obtained from Goodacre et al. (2005) Capelli et al. (2003), and contained information on the loci DYS19, DYS390, DYS391, DYS392, and DYS393. Comparative European populations were obtained from Roewer et al. (2005) in the Y Chromosome Haplotype Database (http://www. YHRD.org). The following markers were used for comparison with European populations: DYS19, DYS389I/II, DYS390, DYS391, DYS392, and DYS393. A list of all comparative populations and their sources are listed in the Table 2.

Table 2. Comparative populations and number of samples used in this study

		mtDNA HVS-I sequences		Y-STR data	
Ethnicity	**Population**	**n**	**Source(s)**	**n**	**Source(s)**
Scottish	Scotland	895		495	Capelli et al. 2003 & Goodacre et al. 2005
	Northwest Coast			154	Capelli et al. 2003 & Goodacre et al. 2005
	Western Isles & Skye		Helgason et al. 2001	160	Capelli et al. 2003 & Goodacre et al. 2005
	Orkney			121	Capelli et al. 2003
	Shetland			256	Capelli et al. 2003 & Goodacre et al. 2005
German	Germany	582	McEvoy et al. 2004	3442	Roewer et al. 2005
Irish	Ireland	300	McEvoy et al. 2004	107	Roewer et al. 2005
English	England	242	McEvoy et al. 2004	247	Roewer et al. 2005
	Cornwall	92			
	Wales	92			
Italian	Italian	248	McEvoy et al. 2004	1340	Roewer et al. 2005
French	France	379	McEvoy et al. 2004	208	Roewer et al. 2005
Polish	Poland	473	McEvoy et al. 2004	1313	Roewer et al. 2005
Dutch	Netherlands	n/a		275	Roewer et al. 2005
	Belgium	33	McEvoy et al. 2004	125	Roewer et al. 2005
Norwegian	Norway	629	McEvoy et al. 2004	300	Roewer et al. 2005
Swedish	Sweden	32	McEvoy et al. 2004	708	Roewer et al. 2005
Danish	Denmark	38	McEvoy et al. 2004	63	Roewer et al. 2005

Laboratory Methods

DNA Extraction

DNA was extracted from the samples using a Chelex extraction method. Samples were thawed and kept still at room temperature for 20-25 minutes to allow cells to settle to the bottom of the tube. The cells were transferred to two 2.0 mL microcentrifuge tubes and centrifuged at 8,000 rpm for five minutes. The aqueous overflow was discarded leaving only the cells in the bottom of the tube. A solution of 100 µL 10% Chelex (BioRad, Hercules, CA) was added to each sample. The samples were resuspended through vortexing and then placed on a heating block at 100°C for 10 minutes. This step causes the cells to lyse and release the DNA and proteins. The next step, cooling the samples by placing them on ice for 3 minutes, allows the Chelex to bind to all of the cellular material except the DNA. The samples were centrifuged for 5 minutes at 10,000 rpms, allowing the Chelex bound material to form a pellet in the bottom of the tube. The Chelx bound material was discarded, and the aqueous overflow with DNA was transferred into a new 0.5 mL tube and stored at 4°C. Problematic samples were extracted using QIAamp Minikit (Qiagen, Valencia, California) following the manufacturer's protocol.

Mitochondrial DNA: RFLP and HVS-I sequencing

Population studies utilizing mtDNA began in the 1980s with the analysis of restriction fragment length polymorphisms (RFLPs), which are variations in the length of DNA segments detected by cuts in the molecule by specific restriction

enzymes (Schanfield, 2006). One of the earliest of these RFLP studies used to examine the origin of modern humans was performed by Cann et al. (1987) which found 195 polymorphic sites in the mtDNA of 145 individuals representing different geographic regions of the world. Further research has identified several RFLPs that are characteristic of populations in Europe, and define the major haplogroups found there (Torroni, et al., 1994; Torroni, et al., 1996).

Recently, scientists have begun to routinely sequence the entire mtDNA genome. While this information may provide better resolution in phylogenetic trees, it has not been show to further our understanding of human population history in a way in which makes it more cost effective than the original studies on RFLP and HVS I sequences (Pakendorf & Stoneking, 2005). HVS sequences alone can be misleading since similar mutations are found in different haplogroups (Torroni, et al., 2000; Kivislid, et al., 2002). However, RFLP analysis alone can only show which haplogroup the mtDNA belongs to, but it cannot further describe their branching orders (Kivislid, et al., 2002). Because of this, and to control cost, this study used both RFLP analyses and sequencing information from the hypervariable segment I to characterize the population.

Sequencing of the non-coding portion of mtDNA, HVS-I was amplified between positions 15976 and 16401. This region was chosen because of a fast mutation rate, and the availability of comparative sequences in several European populations. Polymerase chain reaction (PCR) of the HVS-I was performed using 5

Table 3. MtDNA primers used in the analysis. All primers were synthesized by Integrated DNA Technologies (Coralville, IA). Restriction enzymes are from New England BioLabs (Beverly, MA).

mtDNA loci & enzymes	Primers	Sequence 5'-- 3'	Annealing temperature
HVS I	L15976 For	CCA CCA TTA GCA CCC AAA GCT AAG	57°C
	L16401 Rev	TGATTTCACGGAGGATGGTG	
H	L6958 For	CCT GAC TGG CAT TGT ATT	58°C
Alu I 7025	H7049 Rev	TGTAAAACGACGGCCAGTTGATAGGACATAGTGGAAGT	
K	8931 For	ACC CCT TAT CCC CAT ACT AGT TA	51°C
Hae II 9052	9102 Rev	TTA CTA GAA GTG TGA AAA CGT AGG	
U/K	L12216 For	CAC AAG AAC TGC TAA CTC ATG C	55°C
Hinf I 12308	H12338 Rev	ATT ACT TTT ATT TGG AGT TGC ACC AAG ATT	
V	4519 For	CACTCATCACAGCGCTAAGC	55°C
Nla III 4577	4620 Rev	TGGCAGCTTCTGTGGAAC	

µL of 5X Flexi buffer, 4.3 µL of MgCl2, 0.5 µL of dNTP, 0.2 µL of Taq polymerase, 5.0 µL ddH2O, 2.5 µL each of forward and reverse primers (See Table 3), and 5µL of DNA dilution for a total reaction volume of 25 µL. All of the above reagents, except for the primers, were purchased from Promega (Madison, WI). The thermal profile of the PCR reaction consisted of a 3 minute denaturation period at 94°C, and 35 cycles of a 30 second 94°C denaturation period, a 30 second 57°C annealing period and a 30 second 72°C extension period. The samples where then put through a final 3 minute extension period of 72°C. The fragments were visualized through electrophoresis on a 1.5% Seakem Agarose gel, and photo documented under a UV fluorescent light. Following amplification, the samples were purified using Qiagen's QIAquick PCR purification kit following the manufacturer's protocol (Qiagen, Valencia, California). Purified samples were sent to the University of Kansas DNA Sequencing Laboratory, where they were sequenced by Dr. Mike Grose using an ABI 3700 sequencer.

RFLP is a type of analysis which uses restriction enzymes to target a particular part of the mtDNA genome. A positive result is indicated by the cutting of the DNA fragment, whereas a negative result leaves the fragment intact (Rubicz, et al., 2006). This method was used for haplogroup assignments by way of a hierarchical approach where the most common European haplogroup, haplogroups H, was examined first, and so on down the list until most of the haplotypes were categorized (following Finnila, et al., 2001). PCR amplification for mtDNA haplogroups were prepared using 7.8 µL of sterile water, 2.5 µL 5X Flexi buffer, 4.0 µL of MgCl2, 2.5 µL of primers forward and reverse (see Table 2 for primer list), and 0.2 µL of Taq polymerase. Each multiplex was run under the following thermalcycle profile: an initial 3 minute denaturation period at 94°C, and 35 cycles of a 40 second 94°C denaturation period, a 30 second primer specific temperature for annealing (see Table 2), and a 45 second 72°C extension period. A final 5 minute extension period at 72°C finalized the PCR. The samples were stored at 4°C for further analysis. The presence of DNA was verified by running the samples in a 1.5% Seakem LE Agarose gel and visualized under a UV illuminator. Restriction digest was performed using 9.0 µL of sterile water, 2.0 µL of enzyme specific buffer, 1.0 µL of BSA, and 0.5 µL of the specified enzymes (see Table 3 for list of enzymes). Samples were incubated at 37°C for 12 to 18 hours. The reactions were halted by adding 5 µL of blue loading dye (Promega). Electrophoresis was then used to visualize digested fragments in a 3% NuSieve 3:1 agarose gel with an ethidium bromide stain. Each run was photo documented using a UV fluorescent light.

Sequence data of mtDNA HVSI were aligned to the revised Cambridge Reference Sequence (Andrews, et al., 1999) using the program CLUSTALW (Thompson, et al. 1994) in MEGA version 4 (http://www. megasoftware.net/) (Tamura, et al., 2007). The sequences included up to 354 base pairs at positions (16045-16399), and mutational positions were identified. Mutational sites along the HVSI were used in combination with RFLP data to confirm haplogroup assignments. Samples in which RFLPs were not informative or failed to amplify were assigned to haplogroups through HVS-I haplogroup defining mutations, and were cross checked using the haplogroup assignment program at the National Geographic Genome Projects website (http://nnhgtool.nationalgeographic.com/ classify/).

Y Chromosome: short tandem repeats

The Y-chromosome has been used extensively in forensic and population studies. The extensive use of Y-STRs in various types of investigations has led to the recommended use of eleven core loci for Y chromosome studies (SWGDAM, 2004). The first nine of these loci (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, and DYS393) were recommended in 1997 as the core European loci (Kayser, et al., 1997). Two other loci, DYS438 and DYS439, were added to this core loci profile (SWGDAM, 2004). These markers were chosen for the following study because of the SWGDAM recommendations and the extensive use of these markers in the literature.

PCR was performed on all male samples to amplify regions of the Y chromosome. These reactions were performed in three multiplexes following Redd et al. (1997) and Zlojutro (2008). Multiplex I (DYS385a/b, DYS390, DYS391, and DYS393) consisted of 4.4 µL of 5X buffer, 3.8 µL of $MgCl_2$, 1.0 µL of dNTP mix, 0.3 µL of Taq polymerase, 0.4 µL of BSA, 5.7 µL ddH_2O, 2.0 µL Primer Mix, and 4.4 µL of diluted DNA for a total reaction volume of 22 µL. Multiplex II (DYS19, DYS392, DYS438, and DYS439) consisted of 3.6 µL of 5X buffer, 3.1 µL of $MgCl_2$, 1.0 µL of dNTP mix, 0.2 µL of Taq polymerase, 0.3 µL of BSA, 4.1µL ddH_2O, 2.0 µL Primer Mix (See Table 4), and 3.6 µL of diluted DNA for a total reaction volume of 18µL. PCR reaction mix for DYS389 I/II consisted of 3.6 µL of 5X buffer, 3.1 µL of $MgCl_2$, 0.4 µL of dNTP mix, 0.2 µL of Taq polymerase, 0.3 µL of BSA, 5.2 µL ddH_2O, 0.8 µL each of DYS389I/II For and Rev primers, and 3.6 µL of diluted DNA for a total reaction volume of 18µL. Each multiplex was run under the following thermalcycle profile: an initial 3 minute denaturation period at 94°C, and 35 cycles of a 30 second 94°C denaturation period, a 30 second primer specific temperature for annealing (see Table 4), and a 30 second 72°C extension period. Samples were visualized under a UV illuminator in a 1.5% Seakem LE Agarose gel to ensure successful amplification.

To determine allele size, approximately 1.0 µL of Multiplex I, II and DYS389I/II PCR products were combined and then diluted with ddH_2O at a ratio of roughly 1:100. The samples were sent to the DNA Sequencing Lab (University of

Table 4. List of Y-STR loci and primers for each multiplex used in this study. All primers were synthesized by Integrated DNA Technologies (Coralville, IA). See Roewer & Epplen (1992), de Knijff et al. (1997), Kayser et al. (1997), and Ayub et al. (2000).

	Y-STR loci	Repeat Motif	Primer	Sequence 5' -- 3'	Annealing Temp
Multiplex I	**DYS385a/b**	$[GAAA]_n$	For	AGCATGGGTGACAGAGCTA	
			Rev	TGGGATGCTAGGTAAAGCTG	
	DYS390	$[TCTG]_n[TCTA]_n[TCTG]_n[TCTA]_n$	For	TATATTTTACACATTTTTGGGCC	
			Rev	TGACAGTAAAATGAACACATTGC	57°C
	DYS391	$[TCTA]_n$	For	CTATTCATTCAATCATACACCCA	
			Rev	GATTCTTTGTGGTGGGTCTG	
	DYS393	$[AGAT]_n$	For	GTGGTCTTCTACTTGTGTCAATAC	
			Rev	AACTCAAGTCCAAAAAATGAGG	
Multiplex II	**DYS19**	$[TAGA]_{3tagg}[TAGA]_n$	For	CTACTGAGTTTCTGTTATAGT	
			Rev	ATGGCCATGTAGTGAGGACA	
	DYS392	$[TAT]_n$	For	TCATTAATCTAGCTTTTAAAAACAA	
			Rev	AGACCCAGTTGATGCAATGT	56°C
	DYS438	$[TTTTC]_n$	For	TGGGGAATAGTTGAACGGTAA	
			Rev	GTGGCAGACGCCTATAATCC	
	DYS439	$[GATA]_n$	For	TCCTGAATGGTACTTCCTAGGTTT	
			Rev	GCCTGGCTTGGAATTCTTTT	
DYS389I/II	**DYS389I**	$[TCTG]_3[TCTA]_n$	For	CCAACTCTCATCTGTATTATCTAT	
			Rev	TCTTATCTCCACCCAGA	56°C
	DYS389II	$[TCTG]_n[TCTA]_nN_{28}[TCTG]_3[TCTA]_n$	For	CCAACTCTCATCTGTATTATCTAT	
			Rev	TCTTATCTCCACCCAGA	

Kansas) for analyses by Dr. Mike Grose, where they were loaded onto the ABI®3130xI. The ABI®3130 Data Collection ver. 3.0 program was used to determine fragment lengths at each loci. The alleles were scored and probable haplogroup assignments were inferred using Whit Athey's Haplogroup Predictor (http://www.hprg.com/hapest5/hapest5a/hapest5.htm?order=num), a program which uses a Bayseian method to give the probability of a haplogroup assignment. This program is particularly well suited for populations of Northwestern Europe, in the chosen prior area of origin option, as it has been tested against European populations in the YCC database (Athey, 2006).

Statistical Methods

Diversity measures of mtDNA HVS I sequences and Y-STRs were computed in Arlequin version 3.1 (http://cmpg.unibe.ch/software/ arlequin3/) (Excoffier & Schneider, 2005). The measurements included in this study are Nei's gene diversity statistic (H) (Nei, 1987), the mean number of pairwise differences between all pairs of haplotypes within a sample (π) (Tajima F. , 1993), and two estimates of the expected diversity in a population (θ). Gene diversity (H) is a statistic similar to heterozygosity with diploid data, and measures the probability that two randomly selected haplotypes from a sample will be different (Nei M. , 1987). Gene diversity is defined by the equation:

$$H = \frac{n}{n-1}(1 - \sum_{i=1}^{k} p^2) \quad (1)$$

where n is the number of samples, k is the number of haplotypes in the sample, and p_i is the frequency of the i^{th} haplotypes within the sample (Nei M. , 1987). The mean number of pairwise differences between haplotypes is defined by the equation:

$$\pi = \frac{n}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{k} p_i p_j d_{ij} \quad (2)$$

where n equals the size of the sample, d_{ij} estimates the number of mutations that differentiate haplotypes i and j, k equals the number of haplotypes in a population sample, p_i equals the frequency of haplotypes i (Tajima F. , 1993). For a haploid

locus the expected diversity (θ) equals $2N_e\mu$, where N_e is the effective population size and μ is the mutation rate of the marker studied. Estimations of θ used in this study include $\theta_{k,}$ and θ_S for mtDNA sequence diversity. An estimator of θ based on the expected number of alleles (k) in a population and the sample size in relation to the expected diversity found in the population is computed as θ_k, which equals:

$$\theta_k = \theta \sum_{i=0}^{n-1} \frac{1}{\theta + i} \quad (3)$$

(Ewens, 1972). Arlequin also produces a 95 percent confidence interval around θ_k by assessing the probability that k will equal 0 (lower bound) and 1 (upper bound), and are the sums of the probability of observing k' alleles when k'=0,…k (Excoffier & Schneider, 2005). Theta S is an estimate of the expected diversity within a population based on the number of segregating sites in a sample, and is defined by the equation:

$$\theta_S = S / \sum_{i=1}^{n-1} \frac{1}{i} \quad (4)$$

where S is the number of segregating sites and a_1 (Watterson, 1975). Measurements of mtDNA sequence diversity are based on the infinite sites model, proposed by Kimura (Kimura, 1969), assumes that because the mutation rate at each site is small and the number of sites in a DNA segment is large the probability that a mutation will occur more than once at the same site is negligible. For Y-STR data, the stepwise mutation model was assumed following Ohta and Kimura (1973) which suggests that

mutations increases or decreases an allele's length (or a single repeat) with equal probability (Jobling, et al., 2004).

Summary statistics for the Kansas City Games Group also included an examination of each of the eleven Y-STR loci giving the number of gene copies where information on the loci was present, the number of alleles at each locus, the expected heterozygosity at each locus, the allelic range of each locus, and a Garza-Williamson index (G-W). Gene diversity (*H*) at each locus calculated by Arlequin as:

$$H = \frac{n}{n-1}(1 - \sum_{i=1}^{k} p_i^2) \tag{5}$$

where n is the number of gene copies in the sample, k is the number of alleles at each locus, and p_i is the frequency of the i^{th} allele at each locus (Excoffier & Schneider, 2005). The G-W statistic is a statistic that examines the relationship between the number of alleles at a locus and the range of lengths at each locus. This measure has been shown to be sensitive to bottleneck events and will yield a low G-W statistic. A static population should yield a number close to one (Excoffier & Schneider, 2005). The G-W statistic is calculated by taking the number of alleles at a locus and subtracting the loci's allelic range plus one (Garza & Williamson, 2001).

Diversity measures for Y-STR data were calculated for the Kansas City Group and comparative populations in three combinations. The first combination included the available five loci (DYS19, DYS390, DYS391, DYS392, and DYS393) for Scottish populations found in the literature and included Goodacre et al. (2005) and

Capelli et al. (2003). The next combination compared European population, including a combined set of all Scottish regions labeled as Scotland and the KC Games group, using the same 5 loci listed above. The Scottish populations were then removed from the computations to examine diversity measures of 7 loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393) that are found in other European populations that made contributions to the American gene pool.

Two tests of neutrality were also computed on mtDNA HVS-I sequence data, Tajima's *D* and Fu's F_S. Both methods are based on the infinite-sites model for DNA sequences and RFLP markers that do not recombine. Tajima's *D* compares θ_π, which is equal to π as defined by Equation 2, and θ_S by randomly generating samples under the hypothesis of selective neutrality and a population that is in equilibrium. The *D* is defined by the equation:

$$D = \frac{\theta_\pi - \theta_S}{\sqrt{Var(\theta_\pi - \theta_S)}} \tag{6}$$

(Tajima F. , 1993). A significant *D* value indicates a population which is under selection, expansion, or contraction (Excoffier & Schneider, 2005). Fu's F_S statistic also tests neutrality by randomly generating samples, and is defined by the equation:

$$F_S = \ln(\frac{S}{1-S}) \tag{7}$$

where *S* equals the probability that randomly generated values of *k* (*K*) will be greater or equal to the observed values of *k*, so that θ equals θ_π. This statistic is particularly

sensitive to demographic expansions which give large, negative values that display a p-value of 0.02 and below (Fu, 1997; Excoffier & Schneider, 2005).

The amount of Y-STR and mtDNA sequence haplotype sharing between the Kansas City Highland Games sample and other European populations were also assessed using the shared haplotypes option in Arlequin. A percentage of Kansas City Games haplotypes that were shared with each population were calculated by taking the number of shared haplotypes divided by the number of the sample. Similarly, the percentage of European haplotypes within each population that matched with the Kansas City Games haplotypes were also computed.

Another way to represent the molecular diversity in a population is through a mismatch distribution. Mismatch distributions can be suggestive of demographic events in a population's history. For example, a unimodal distribution has been shown to represent a recent population expansion while a multimodal distribution indicates a population that has maintained its size over time (Rogers & Harpending, 1992). A raggedness index (*r*) asseses these mismatch distributions, as a higher index is found in a static population that usually produces a multimodal distribution (Harpending, et al., 1993). This index is defined by the following equation:

$$r = \sum_{i=1}^{d+1} (x_i - x_{i-1})^2 \tag{8}$$

where *d* is the maximum differences between haplotypes, and *x* is the relative frequency of the mismatch. To visualize the diversity of the Kansas City Games

sample and those from Scotland, mismatch analysis and *r* index were performed on mtDNA sequences data based on pairwise distances computed in Arlequin 3.1.

To examine the population structure of the Kansas City Games group when compared to other Scottish populations, an Analysis of Molecular Variance (AMOVA) was performed in Arlequin 3.1. Each population was placed into one of three groups: 1) Kansas City Games group, 2) Scottish mainland (including the Northwest Coast of Scotland), and 3) the Scottish Isles (including the Western Isles, the Isle of Skye, Orkney, and Shetland. This statistical approach is a way of estimating population differentiation based on the molecular data, and provides the percent of variation explained by variation among the three groups, the variation among populations within groups, and the variation found within the populations (Excoffier & Schneider, 2005).

A median-joining network (MJ) was computed for mtDNA sequences and RFLPs in the statistical program NETWORK 4.5.1 (Fluxus Technology Ltd) following Bandelt, et al., (1999). MJ networks represent relationships between haplotypes by inferring likely ancestral haplotypes. This method was employed to examine the Kansas City Games samples. HVS-I sequences were analyzed between nucleotide positions 16070-16400. Both actual RFLP data and inferred RFLPs were added to the MJ network to identify the major haplogroups represented in the sample and to better assess the relatedness of each individual. To account for variation in the rate of evolution at differing sites, the sites were classified by their rate of evolution following Roostalu et al. (2007). Fast evolving positions were weighted to 1 and

included 16093, 16129, 16189, 16311, and 16362. Sites that evolve at an intermediate rate include 16126, 16145, 16168, 16172, 16184, 16192, 16209, 16218, 16223, 16256, 16261, 16278, 16291, 16293, 16294, 16304, 16320, and 16325, and were given a weight of 2. RFLP sites were weighted at 10, and all other positions were considered to be slow evolving and weighted at 4.

Two other methods were employed to show the relationship of the Kansas City Highland Games sample and populations of Scotland and Europe: Neighbor-Joining (NJ) trees and Multidimensional Scaling (MDS). A Neighbor-Joining (NJ) tree is produced in an attempt to find a tree that describes the relationship between populations with the shortest sum of branch lengths (S_O) (Saitou & Nei, 1987). This method begins with a star tree which is produced with the assumption that the groups do not cluster. The NJ method then computes the sum of all branch lengths until pairs of true neighbors are identified to find the smallest S_O. The smallest sum of branch lengths can be defined by the following equation:

$$S_O = \sum_{i=1}^{m} \mathrm{L}_{iX} = \frac{1}{m-1} \sum_{i<j}^{m} d_{ij} = \frac{T}{m-1} \quad (9)$$

where m the number of neighbors, L_{iX} equals the branch length estimate between nodes i and X, and T equals the sum of distance estimates d_{ij} ((Nei & Kumar, 2000). Multidimensional Scaling (MDS) is a method that displays information from various distance matrices in a form which can be examined visually, usually in two or three dimensions. A traditional MDS analysis was performed on Tamura & Nei's

(1993) distances for mtDNA sequence data because it accounts for differences in transversion and transition rates, as well as transition rates between purines or pyrimidines (Excoffier & Schneider, 2005). Slatkin's (1995) corrected linearized R_{ST} distances were used for MDS plots of Y-STR data as described in Manly (1994). A test of goodness-of-fit of the data to the graph can be assessed by the stress value defined by Kruskal's formula as:

$$Stress = \sqrt{\sum \left(d_{ij} - d_{ij}^{f}\right)^{2} / \sum d_{ij}^{2}}, \tag{10}$$

where d_{ij} represents the distance values for each pair of objects and d_{ij}^{f} represents a scaled distance for each *ij* (Kruskal, 1964). The lower the stress value the better the plot represents the information from the original distance matrix. Sturrock & Rocha (2000) found a relationship between the number of samples or populations used to produce a plot and the stress value that should be influenced by these numbers. MDS plots for this study included the use of 5, 6, 12 and 13 populations with the suggested two dimensional maximum stress values for goodness-of-fit to the data as 0.000, 0.019, 0.183, and 0.199 respectively (Sturrock & Rocha, 2000).

All participants in the study were sequenced at the HVS-I and analyzed for RFLPs in order to determine mtDNA haplogroups. Male participants in this study also underwent Y-STR analyses which were used to infer haplogroups within the sample. Diversity measures including *H* and π, expected diversity measures including θ_k an θ_S, measures of neutrality including Tajima's *D* and Fu's *FS*, and a

median-joining network and mismatch distribution for mtDNA sequences, are used to describe individuals from the Kansas City Highland Games and the European populations that they are compared to. To examine population structure among the Scottish samples and the Kansas City games group, both mtDNA sequence data and Y-STR data were used for AMOVA. Multidimensional scaling and neighbor-joining trees were used to visualize the relationship between the KC Games group and European populations.

Chapter Four: Results

Mitochondrial DNA Analyses

A total of 80 individuals were sequenced for HVS-I and underwent RFLP analyses. In all, the sample revealed 61 unique haplotypes belonging to haplogroups H, V, U, K, J, T, and W (See Table 5). Three individuals matched the CRS for HVS-I haplotypes but did not belong to haplogroup H, U or V based on RFLP analysis. These three samples are classified as unknown. The most common haplogroup represented in the sample was haplogroup H, making up 51 percent of the sample with haplogroups U, J, V, T, K and W making up between 14 and 1 percent. Figure 14 gives the haplogroup frequencies in the Kansas City Games group compared to other Scottish populations.

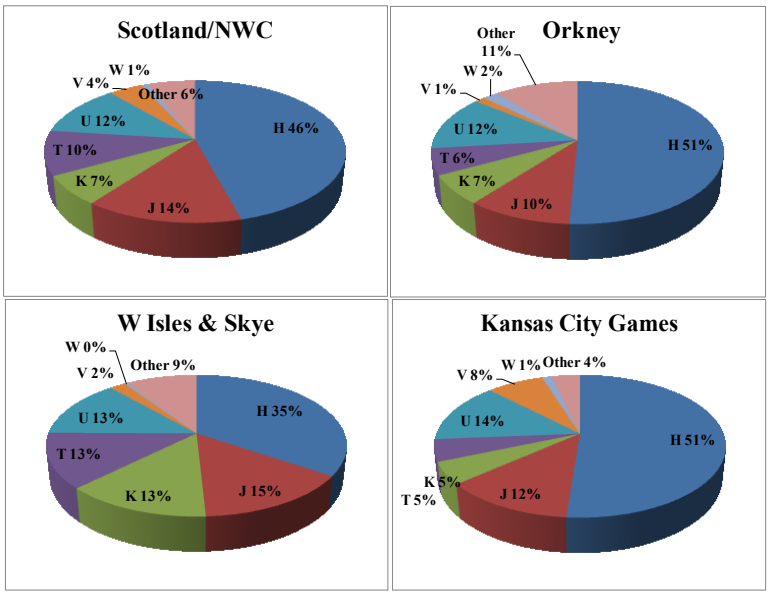


Figure 14. Mitochondrial haplogroup frequencies in Scotland, Orkney, Western Isles & Skye (Helgason, et al., 2001), and the Kansas City Highland Games group.

Table 5. Results from mtDNA HVS-I sequencing and RFLP analysis. A (*) denotes failed RFLP attempts.

	Coding region				Control region			Coding region				Control region	
Hg	7025	4577	12305	1034	HVS I haplotype (+16000)	n	Hg	7025	4577	12305	1034	HVS I haplotype (+16000)	n
H	-				93	1	V	+	-			298	3
H	-				93, 304	1	V	+	-			79, 298	1
H	-				75	1	V	+	-			261, 298, 311	1
H	-				60, 304	1	V	+	-			223, 240, 298	1
H	-				110, 162, 209	1	U	+	+	+	-	75, 162	1
H	*				129, 223, 391	1	U	+	+	+	-	126, 294, 296, 304	1
H	-				134, 356	1	U	+	+	+	-	75, 104, 189, 256, 270	1
H	-				140, 274, 356	1	U	+	+	+	-	93, 192, 256, 270, 291	1
H	-				147, 356	1	U	+	+	+	-	1014, 192, 256, 270, 294	1
H	-				148, 293	1	U	+	+	+	-	93, 256, 270, 291	1
H	-				162, 209	1	U	*				189, 270	1
H	-				176	2	U	+	*			189, 270, 319, 392	1
H	-				189, 356, 375	1	U	+	*			192, 256, 270	1
H	*				64, 184, 189, 356	1	U	+	-	*		51, 129, 189, 194, 197, 362, 377	1
H	-				189, 368,377	1	U	+	-	*		51, 129, 182, 183, 189, 233, 362	1
H	*				183, 189, 225, 278	1	K	+	+	+	+	224, 311	3
H	-				192, 239,311	1	K	+	+	+	+	93, 224, 311, 319	1
H	*				235, 291, 293	1	J	+	+	-		126	1
H	-				239	1	J	+	+	-		59, 126	1
H	-				266, 320	1	J	+	+	-		69, 126	2
H	-				287	1	J	+	+	-		69, 126, 311	1
H	-				289	1	J	+	+	-		75, 116, 126, 311	1
H	-				293	1	J	+	+	-		126, 294, 296, 304	2
H	-				293,362	1	J	+	+	-		60, 69, 126, 265, 319	1
H	-				30, 362	1	J	+	+	-		69, 126, 145, 172, 223, 261	1
H	-				304, 311	1	T	+	+	-		126, 294, 311	1
H	-				368	1	T	+	+	-		126, 163, 186, 189, 294	1
H	-				319	4	T	+	+	-		126, 153, 242, 294, 320	2
H	-				CRS	9	W	+	+	-		223, 278, 292	1
Unk	+	+	-		CRS	3							

Figure 15 displays a median-joining network of the haplotypes found in this study. The most common haplotype in this study belonged to nine individuals who were in haplogroup H and matched the HVS-I Cambridge Reference Sequence.

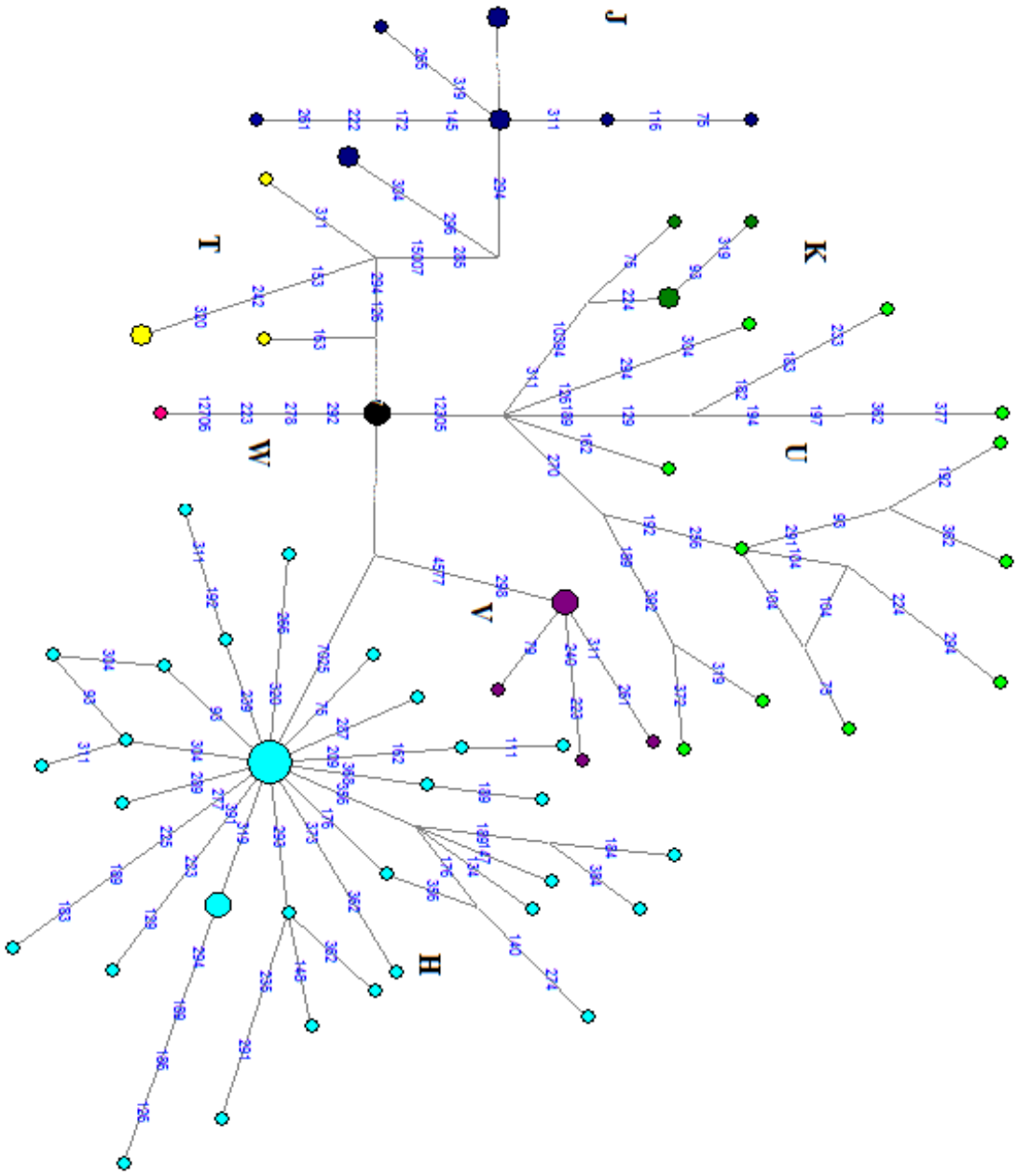


Figure 15. Median-joining network of eighty Kansas City Highland Game group mtDNA haplotypes. All mutation sites (shown in blue) are plus 16000 except for nucleotide positions 7025, 4577, 12305, and 10394. Nodes are proportional to haplotype frequency, with the smallest node representing one, and the largest node representing nine.

Table 6. Summary statistics for the KC Games group and four Scottish populations based on mtDNA HVS-I sequence data (np 16080-16380), including the number of samples (n), gene diversity (H), the number of pairwise differences between haplotypes (π), and the expected diversity based on k and S.

Sample	n	H	s.d (H)	π	s.d. (π)	θ_k	C.I. (θ_k)	θ_S	s.d (θ_S)	Tajiima's	Fu's FS
KC Games	80	0.9677	0.0125	4.1139	2.0704	71.7110	(45.4685, 114.3927)	12.5177	3.4850	-2.2402	-25.9087
Scotland	603	0.9669	0.0043	4.0658	2.0312	121.1191	(101.2758, 144.5571)	14.0433	2.8956	-2.0583	-25.0753
NW Coast	196	0.9683	0.0073	4.2943	2.1359	65.4162	(48.7873, 87.5209)	11.9601	2.9129	-1.9837	-25.5207
Wisles&Skye	159	0.9797	0.0040	4.4762	2.2168	58.1779	(42.1373, 80.1730)	10.8099	2.7461	-1.8379	-25.5680
Orkney	77	0.9498	0.0145	3.7708	1.9215	27.3594	(17.3423, 42.9931)	6.9183	2.0851	-1.4890	-26.0306
Mean	223	0.9665	0.0085	4.1442	2.0752	68.7567	(51.0021, 93.9274)	11.2499	2.8250	-1.9218	-25.6207

*Tajima's D are significant at p<0.05, Fu's FS are significatn at p<0.005

Another haplotype belonging to haplogroup H was found in four individuals with a transition at np 16319. Two haploptypes, one belonging to haplogroup V with a transition and np 16298, and a ht belonging to haplogroup K with transitions at nps 16224 and 16311, belonged to three individuals. All other haplotypes, were found in one individual in the study. The main purpose of a median-joining network is to show the relationship between the haplotypes and identify ancestral nodes. However, as this is a heterogeneous sample representing various European backgrounds, the MJ network presented here is intended to show the diversity observed in this sample.

Table 6 provides the summary statistics for the Kansas City Highland Games group when compared to four regions of Scotland. Compared to Scottish HVS-I sequences, the Kansas City Games group displays a slightly higher level of genetic and nucleotide diversity than the average for all regions except, despite having a smaller sample size compared to every regions except Orkney. When levels of expected diversity are tested for neutrality, both Tajima's *D* and Fu's *FS* give significant values. These results are consistent with other studies of European

populations, and are usually indicative of population expansion or selection. Based on the historical data these numbers are most likely due to a population expansion.

When the Kansas City Games group was compared to Scottish populations, fifty-eight percent (n=46) of the group had a haplotype that is frequent in Scotland. Of these, fifty-six percent (n=45) matched with the Scottish mainland and other Scottish populations, one sample only matched with the Scotland mainland, one matched a ht found in the Northwest Coast of Scotland, and another sample matched with a haplotype found only in the Western Isles and the Isle of Skye. Figure 16 shows the percent of the Kansas City Games group that share a haplotype with each Scottish region and the percent of each region that shares a haplotype with the KC Highland Games sample. Maternal lines are more commonly found on the Scottish mainland and the Northwest of Coast of Scotland, whereas the Scottish Isles share the least maternal lines.

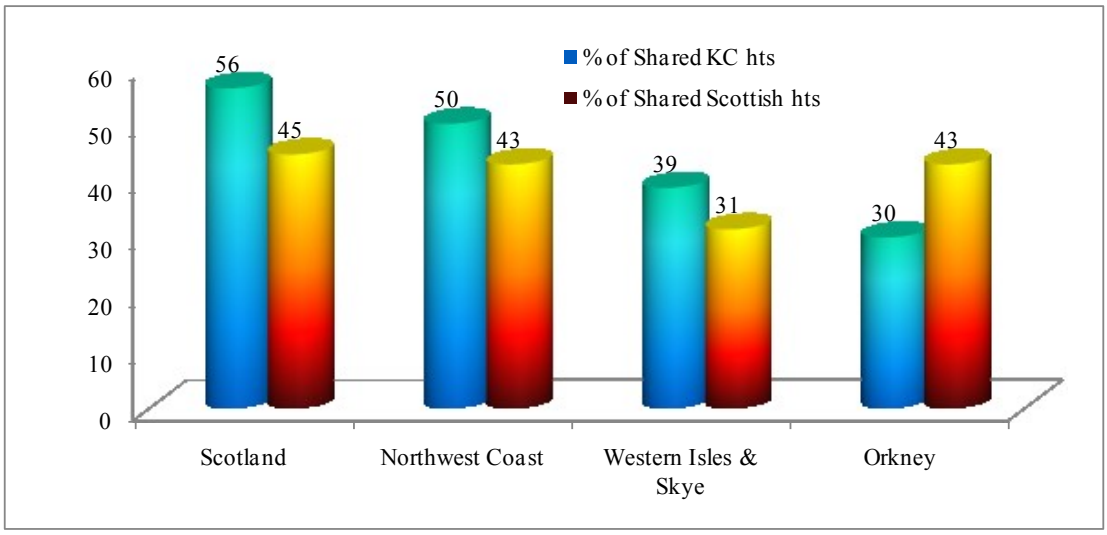


Figure 16. Percent of shared mtDNA HVS-I haplotypes (16080-16380 np)

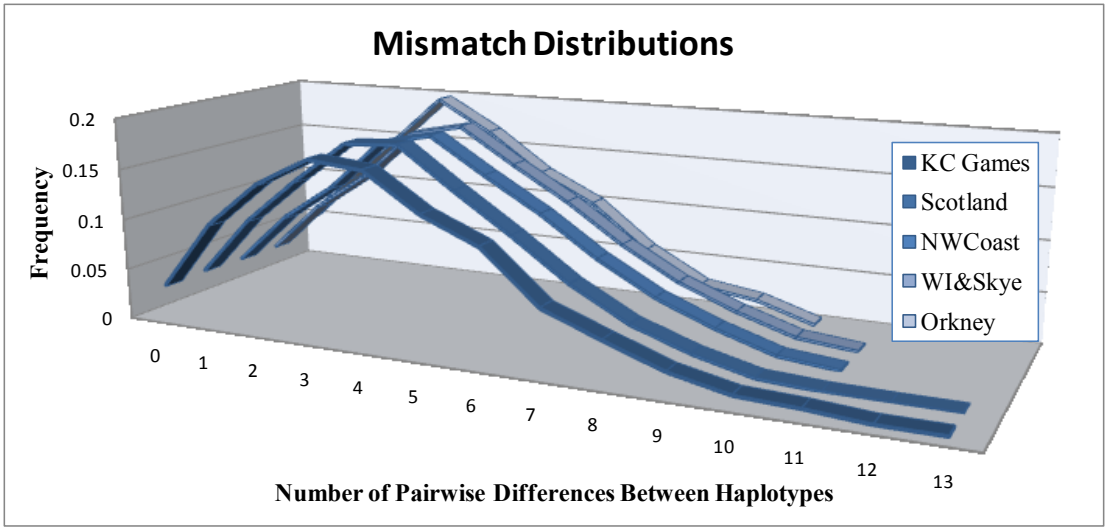


Figure 17. Mismatch distributions of the KC Games group and four Scottish populations.

Mismatch distributions of the Kansas City Games group and the four Scottish regions are provided in Figure 17. Overall, the distributions of the groups are similar. All five groups reach a peak in the number of pairwise differences between three and six, with frequency ranges between 17 and 20 percent. A raggedness index was calculated for each population, ranging between 0.0120 (Northwest Coast) and 0.0145 (Western Isles and Skye). None of the values were significant at p=0.05. These small raggedness values, coupled with a unimodal distribution, are usually associated with a population in demographic expansion (Excoffier & Schneider, 2005).

A neighbor joining tree showing the relationship between the Kansas City Games group and the four comparative Scottish regions can be viewed in Figure 18. In this plot, the Kansas City Games group forms a cluster with the Scottish Mainland, and the Northwest Coast of Scotland. Orkney, a population with a low frequency of

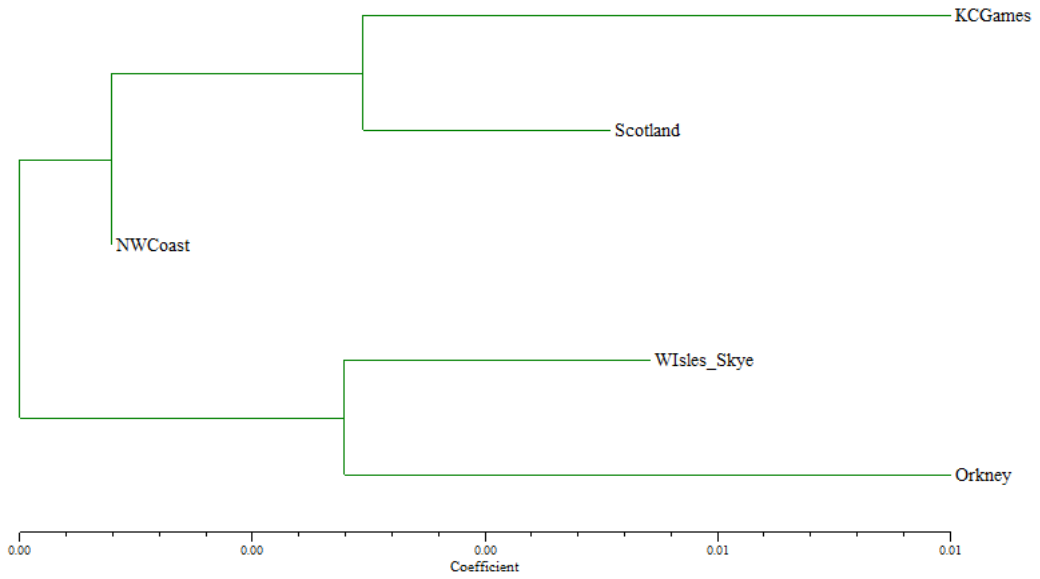


Figure 18. NJ tree of Tamura and Nei's distance between the Kansas City Games group and four Scottish populations.

haplogroup H compared to the other regions of Scotland, is the furthest away from the Kansas City Group and clusters with the combined samples from the Western Isles and the Isle of Skye. Historically, these two clusters are in agreement with studies on maternal Scandinavian contributions which are strongest in Orkney and the Isles than on the Scottish mainland (Helgason, et al., 2001). They may also be the result of greater female migration to the United States by women from the Scottish mainland *versus* the Scottish Isles (Gordon, 2002). A two-way Mantel test, testing the correlation between the original distance matrix and the cophenetic distance matrix represented in the tree, gave a low *r* of 0.27835. This suggests that the tree is a poor fit to the data, which may be better explained in a method such as Multidimensional Scaling.

A MDS plot was also constructed from Tamura and Nei's (1993) distances, and can be viewed in Figure 19. Orkney and the Western Isles and Skye are once again separated from the other populations, and pulled to the upper half of the plot by their high frequency of haplotypes not found in the Kansas City Game group, and a relatively low frequency of haplotypes belonging to haplogroup V. Similar to the NJ-tree, this plot shows the influence of Scandinavian maternal lines. Orkney is pulled to the right of the plot due to a high frequency of haplogroup X which reaches 7-8 percent in this population. This high frequency of a relatively rare haplogroup has been explained as the result of genetic drift following either post-glacial expansion or movement of the haplogroup into the area following Scandinavian settlement (Helgason, et al., 2001; Reidla, et al., 2003). Haplogroup X is absent in the Kansas City group and present at small frequencies in the other Scottish populations. Haplogroup V is found at a low frequency of 4 percent in Orkney *versus* 8 percent in the Kansas City Games group. The presence or absence of haplotypes belonging to these two haplogroups is the major influences of the first axis of the plot. The second axis of the plot is largely influenced by the presence or absence of haplotypes belonging to haplogroups K and U. Haplogroup K has the highest frequency in the Western Isles at 13 percent and the lowest frequencies are found in the Scottish mainland at 6.2 percent and the Kansas City Games group at five percent. Two subclades, haplogroup K2b found at 2.4 percent and haplogroup U1 found at 2 percent, are found in the Western Isles and the Isles of Skye and are present at less than 0.6 percent in all other Scottish populations (Helgason, et al., 2001).

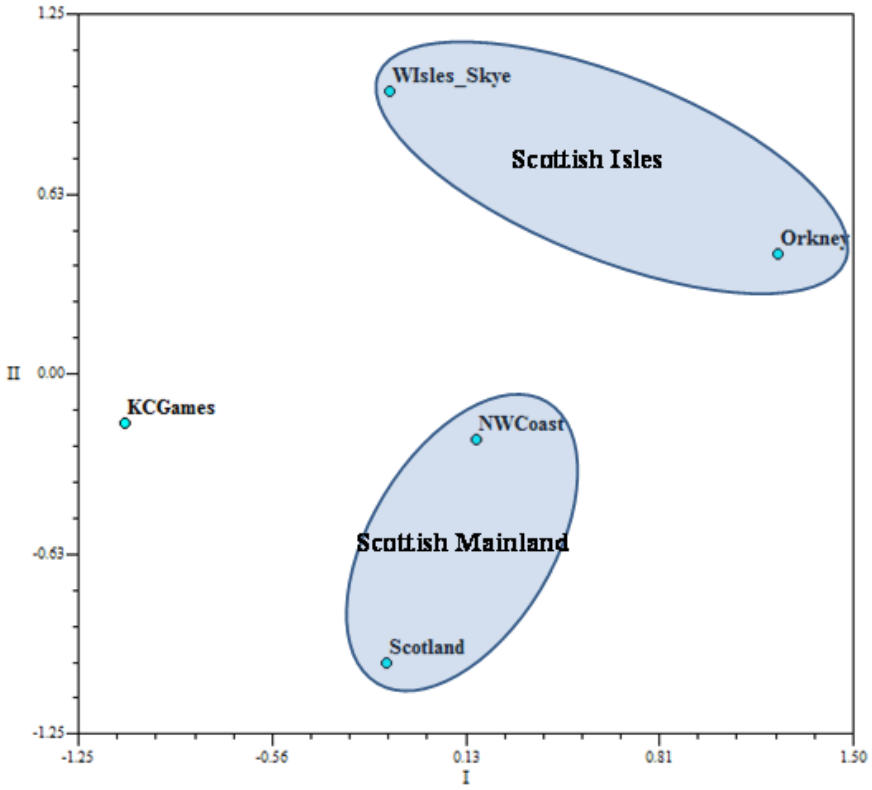


Figure 19. MDS Plot of Kansas City Games Population Compared to Four Scottish Populations based on Tamura and Nei's distances of mtDNA HVS-I sequences. Final stress = 0.00001.

Based on historical information, the populations were separated into three groups to assess the differences in genetic structure through AMOVA. The Kansas City Games individuals make up the first group as they are a heterogeneous sample. The second group is the Scottish mainland, including the Northwest Coast, and the third group consists of the Scottish Isles (Western Isles, Isle of Skye, and Orkney). The results of this analysis can be seen in Table 7. No significant differences were observed among the groups, which only explain 0.51 percent of the total variation (See Table 7). There were significant differences among the populations within each

Source of Variation	d.f.	Sum of squares	Variance Components	Percent of Variation	*F*-Statistic
Among groups	2	11.525	0.01106	0.51	$\Phi_{CT} = 0.00509$
Within Groups	2	6.563	0.0056	0.26	$\Phi_{SC} = 0.00259$*
Within Populations	1118	2411.458	2.15694	99.23	$\Phi_{ST} = 0.00766$**

*p <0.05, **p < 0.001

Table 7. AMOVA results for comparison of three groups (KCGames, Scottish Mainland (including the Northwest Coast), and the Scottish Isles (Orkney and the Western Isles and Isle of Skye)

group, but these differences only accounted for 0.26 percent of the total variation. The most significant result from this analysis was the variation within populations, accounting for 99.23 percent of the total variation. These results can be explained by the sharing of haplotypes between groups, and the diversity of haplotypes within each population.

Summary statistics for thirteen European populations and the KC Games group's mtDNA HVS-I sequences (np 16090-16397) are listed in Table 8. Despite the small sample size, the Kansas City Games group displays Nei's genetic diversity ($H = 0.9677$) and number of pairwise differences between haplotypes ($\pi = 3.7943$) figures similar to the rest of Europe. Compared to the rest of Europe, the Kansas City games group has lower expected levels of diversity when compared to the mean. These measures are sensitive to sample size, so they must be taken with some caution. When compared to populations in the North Atlantic (Scotland, England, Cornwall, Wales, and Ireland), they display a high level of expected diversity based on *S*, and a lower level of expected diversity based on estimates using *k*.

Table 8. Summary statistics describing the mtDNA HVS-I sequence diversity (np 16090-16397) within populations the number of samples (n), the expected heterozygosity (*H*), the number of pairwise differences between haplotypes (π), and the expected diversity estimated by *k* and *S*.

Sample	n	*H*	s.d. (*H*)	π	s.d. (π)	θ_k	95% C.I. (θ_k)	θ_S	s.d (θ_S)	Tajima's	Fu's FS
KC Games	80	0.9677	0.0125	3.7943	1.931	71.7110	(45.6858-114.3927)	11.5082	3.2313	-2.2004	-26.021
Scotland	895	0.966	0.0037	4.007	2.21688	130.1041	(111.3963-151.6475)	15.1896	2.96154	-2.07693	-24.938
Germany	582	0.971	0.0043	3.7536	2.09696	n/a	n/a	16.4198	3.3429	-2.23165	-25.165
England	242	0.9651	0.0078	3.9717	2.20672	140.4169	(108.0302-182.7598)	15.1713	3.4966	-2.24347	-25.521
Cornwall	92	0.9603	0.0128	3.5201	2.00354	51.2470	(33.8323-77.8340)	10.6016	2.93554	-2.15228	-26.162
Wales	92	0.9259	0.0208	3.1424	1.8204	34.1656	(22.4883-51.7604)	9.22734	2.59788	-2.1053	-26.404
Ireland	300	0.9567	0.008	3.6212	2.03685	113.3193	(89.3976-143.5096)	13.8549	3.12902	-2.20313	-25.612
Italy	248	0.9717	0.007	4.562	2.48906	175.8613	(135.3743-229.0425)	15.7671	3.60795	-2.16002	-25.234
France	379	0.9713	0.0053	3.8587	2.14928	164.6082	(133.3325-203.1740)	16.1205	3.47385	-2.25248	-25.36
Poland	473	0.9644	0.0058	4.144	2.28465	164.1568	(135.6427-198.4943)	16.1835	3.38705	-2.17528	-25.137
Denmark	38	0.9346	0.0289	3.495	2.02361	20.9510	(11.2143-39.3554)	7.37817	2.48592	-1.83126	-14.09
Norway	629	0.9528	0.0064	3.8035	2.12061	162.9112	(137.4551-192.8251)	17.3776	3.4841	-2.25614	-25.145
Sweden	32	0.9879	0.0115	4.3992	2.48035	78.7792	(34.9157-190.9116)	9.18742	3.11626	-1.8895	-24.852
Belgium	33	0.9924	0.0094	3.5133	2.04117	110.7921	(46.6067-287.6055)	9.11669	3.07634	-2.21682	-25.956
Mean	294	0.9634	0.0103	3.8276	2.13579	109.156	(80.4132-158.7163)	13.0788	3.166161	-2.14248	-24.685

*Tajima's D are significant at $p < 0.05$, Fu's FS are significant at $p < 0.005$

Figure 20 displays the percent of haplotypes shared between the Kansas City Games population and the thirteen comparative European populations. Sixty-one percent of the KC games group shared a haplotype with Scotland. However, it should be noted that of these shared haplotypes, only three are found only in Scotland, while the others are shared and dispersed throughout Europe. Wales shares the largest percent of its haplotypes with the KC Games group at 50 percent, although this may be a reflection of the small sample size of 92 individuals. Ireland, with a total sample size of 300, is second to Wales and shares 51 percent of its lineages with the Kansas City Games group. It should be noted here that unlike the Scottish populations Irish females equaled or outnumbered males in their migration to the United States (Greenwood, 2008). As many of the maternal lineages in Ireland and Scotland are shared, it may be hard to differentiate between the two populations using this marker.

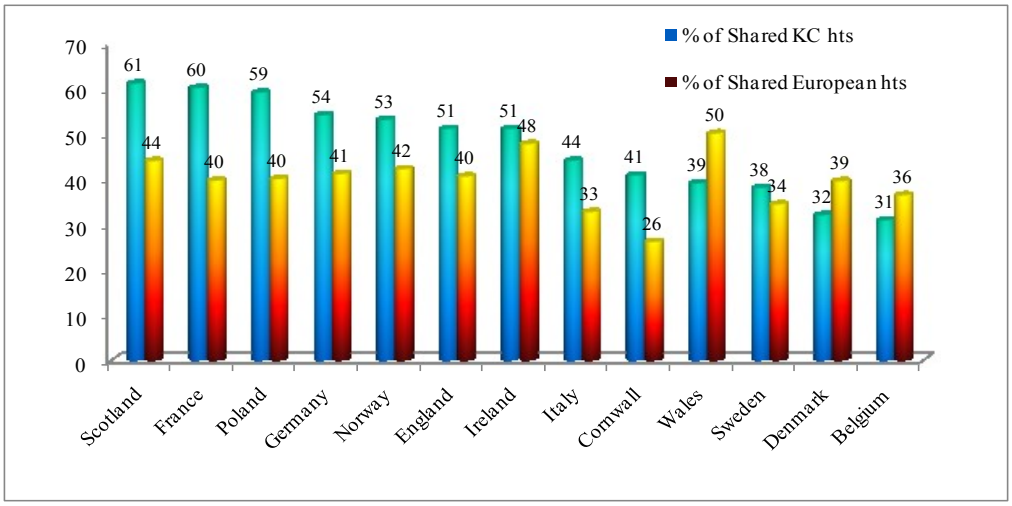


Figure 20. Percent of shared mtDNA HVS I haplotypes.

There is also a large amount of haplotype sharing seen when the sample size of a population is small, as can be seen with the amount of sharing with Belgium, Sweden, Denmark, and Wales.

A neighbor-joining tree was constructed to visualize the relationship between the KC Games group and the European comparative populations (Figure 21). The KC group clusters with Italy, Cornwall, Germany and Poland. England, Denmark and Norway, Scotland and Wales form a separate cluster in agreement with a shared geographical proximity and historical movements of people. It should be noted here that the Kansas City Games group is not clustered with Scandinavian countries which would be expected following their dissimilarity to Orkney and the Western Isles and Isles of Skye. The group seems to be more related to other populations besides Scotland, a further indication that the group is a mix of maternal European genes. As Scottish male immigrants outnumbered female immigrants to the New World, it is

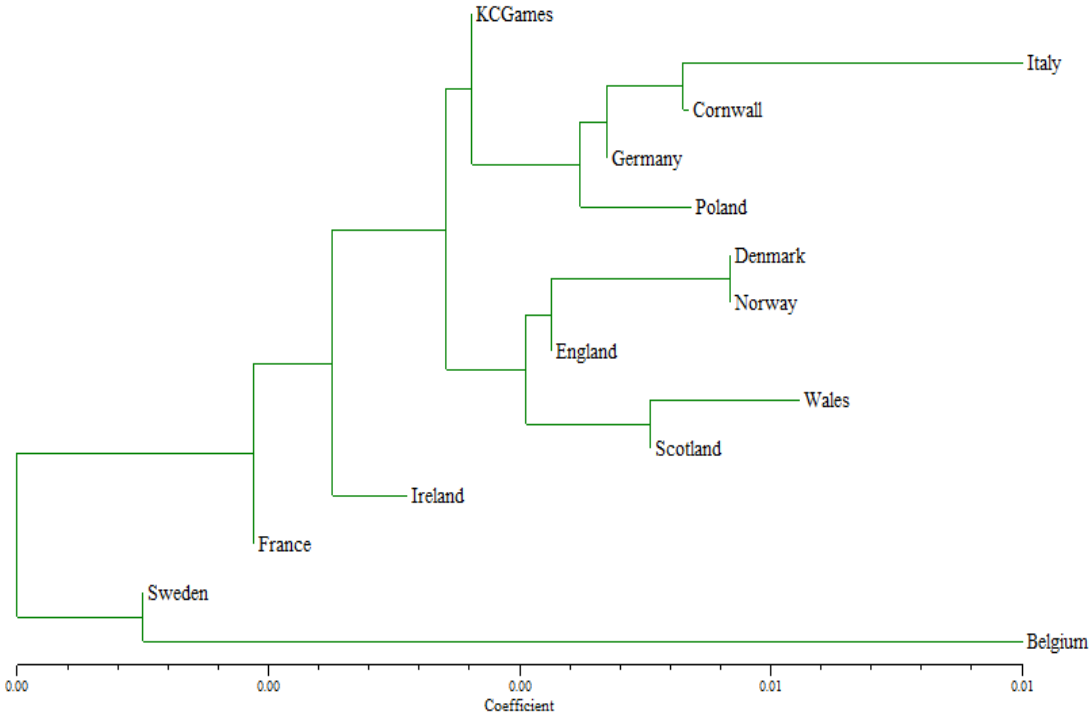


Figure 21. NJ tree based on Tamura and Nei's distance between European populations.

likely that males found mates who migrated from Italy, Cornwall, Germany, Poland, or some other region of Europe. Belgium is the least similar to the KC group in the tree. A two-way Mantel test comparing the original distance matrix and the cophenetic distance matrix gave an insignificant correlation value of 0.53231, indicating the there is some correlation between the original distance matrix and the distances represented in the NJ tree, but it is a poor fit to the data.

When Tamura and Nei's (1993) distances are used to compute a MDS plot, a similar relationship is seen. Figure 22 shows that the sample from Belgium is separated from the other populations at the far, upper right corner of the plot. This is probably due to a relatively high frequency of haplogroup K in the Belgium population. Another MDS plot was constructed excluding the Belgium sample, but a

high stress value and an inability to further differentiate the populations made it uninformative for this study. In Figure 22, the KC Games group is surrounded by Norway, Denmark, England, and France. Scotland, Germany, Ireland and Sweden are further from the Kansas City group but still form a cluster that includes all European populations except for Italy and Belgium. This tight clustering of several European populations is due to the amount of shared lineages between populations due to historical female migration. In sum, the KC Games group seems to be a mix of European maternal genes, an indication that their Scottish identity might be the result of some other association. The stress value for the plot was given as 0.14439, lower than the suggested goodness-of-fit maximum value of 0.217 for fourteen populations suggested by Sturrock and Rocha (2000).

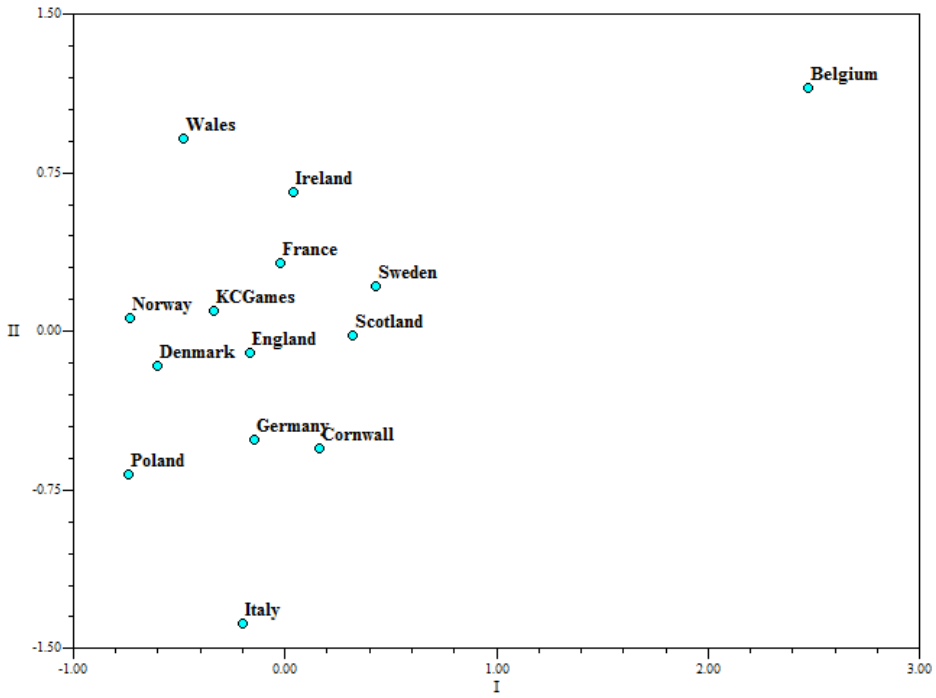


Figure 22. Multidimensional Scaling plot of Tamura and Nei's (1993) distance measure between populations, based on mtDNA control region sequences. Final stress = 0.14439.

Y-chromosome Analyses

A summary of the results of Y-STR analyses, Y haplogroup predictions, and their probabilities are shown in Table 9. A total of forty-seven samples provide information for this study, with Y-haplogroup assignments to G2a, I1, R1a and R1b. Predicted probabilities for four individuals designated to haplogroup G2a were above 95 percent. Samples assigned to haplogroup I1 (n=4) were given a probability of assignment above 90 percent, except for one sample lacking data at three loci that was given a probability of 73.6 percent. The one haplotype assigned to R1a, and all but two haplotypes assigned to R1b had a probability of correct assignment greater than, or equal to, 95 percent. One haplotype with missing data at three loci was given a probability of haplogroup assignment at 83.7 percent, while another haplotype missing data at four loci was predicted as belonging to haplogroup R1b with a probability of 58.8 percent. Four of these individuals were not used for comparative analysis due to missing data caused either by chelex interference with the chemical reactions used to analyze these samples or due the low levels of DNA in the samples when they were processed.

Table 10 provides the summary statistics for the eleven loci examined. During Y-STR analysis, the number of copies of each locus shows that DYS385a yielded the fewest repeats with n totaling 37. All other loci provided at least 43 copies for analyses. The number of alleles at each locus range from three to five, with an average of four alleles at each locus. The expected heterozygosity is highest for DYS390 ($H = 0.7497$) and lowest for DYS393 ($H = 0.4440$). All of the loci

Table 9. Y-STR haplotypes, predicted haplogroups, and probability of haplogroup assignment (x 100). * denotes haplogroups and probabilities assigned in Whit Athey's haplogroup Predictor (http://www.hprg.com/hapest5/) (Athey, 2006). A (-) indicates failed STR amplification.

DYS19	DYS385a	DYS385b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	Haplogroup*	Probability*
13	-	-	13(10)	30(27)	24	10	11	13	10	12	E1b1b	100
15	10	14	12(9)	30(27)	22	10	10	14	10	11	G2a	100
15	-	14	12(9)	30(27)	22	10	11	14	10	11		96.8
15	10	14	12(9)	30(27)	22	10	11	14	10	11		99.7
16	-	-	12(9)	29(26)	22	10	11	14	10	11		95.7
14	11	13	12(9)	29(26)	22	10	11	13	10	11	I1	98.3
14	-	-	13(10)	30(27)	22	-	11	13	10	12		92.3
16	-	14	12(9)	28(25)	22	10	11	13	10	11		96.8
15	-	-	12(9)	28(25)	24	10	11	13	10	-	I1/I2b	73.6/17.8
14	13	16	14(10)	29(26)	23	10	11	12	9	11	J2a1b	97.5
15	13	16	12(9)	28(25)	24	10	11	12	11	11	J2b	82.7
15	11	14	13(10)	33(30)	25	10	11	13	11	10	R1a	100
13	11	14	12(9)	28(25)	24	11	13	14	12	12	R1b	100
13	11	14	13(10)	29(26)	23	11	14	13	12	12		100
14	10	13	13(10)	29(26)	25	11	14	13	12	12		100
14	11	13	13(10)	29(26)	24	11	13	13	12	12		100
14	11	13	13(10)	29(26)	25	11	14	13	12	12		100
-	11	13	-	-	-	10	-	13	-	12		99.6
14	11	13	13(10)	30(27)	24	10	13	13	12	12		100
14	11	14	13(10)	29(26)	23	11	14	13	12	11		100
-	11	14	-	-	24	10	-	-	-	11		95
14	11	14	13(10)	29(26)	24	10	13	13	12	11		100
14	11	14	13(10)	29(26)	24	10	13	13	12	11		100
14	11	14	13(10)	29(26)	25	11	13	13	12	11		100
14	11	14	13(10)	29(26)	-	12	13	14	12	12		100
14	11	14	13(10)	30(27)	24	11	13	12	11	12		100
14	11	14	13(10)	30(27)	24	11	13	12	12	13		100
14	11	14	13(10)	30(27)	25	10	13	13	13	11		100
14	11	14	14(11)	30(27)	25	11	12	13	12	12		100
14	11	14	14(11)	30(27)	25	11	13	13	12	11		100
14	11	15	-	29(26)	23	11	13	13	12	13		100
14	11	15	13(10)	29(26)	23	11	13	13	12	11		100
14	11	15	14(11)	33(30)	24	11	13	12	12	11		100
14	11	17	12(9)	29(26)	24	11	13	13	12	12		100
14	12	14	14(11)	28(25)	25	11	13	13	12	12		100
14	-	-	13(10)	29(26)	-	-	13	-	12	11		100
14	-	-	13(10)	29(26)	25	11	14	13	12	12		100
15	10	17	12(9)	28(25)	23	11	13	13	12	12		100
15	11	14	13(10)	28(25)	23	11	13	13	12	11		100
15	11	-	13(10)	29(26)	23	10	13	13	12	12		100
15	11	14	13(10)	29(26)	23	12	13	13	12	12		100
15	11	15	12(9)	28(25)	24	11	13	13	12	12		100
15	11	15	13(10)	29(26)	25	11	13	13	12	12		100
-	11	-	12(9)	28(25)	25	10	13	13	12	13		100
12	11	15	-	-	-	12	12	-	9	11	R1b/J2a1	58.8/38.9
15	-	13	13(10)	29(26)	-	11	11	-	12	11	R1b/R1a	83.7/15.7

*Predicted in Whit Athey's Haplogroup Predictor

Table 10. Summary statistics for the Kansas City sample at eleven loci

Y-STR	# of copies	# of Alleles	Allele Range	Exp. Het	G-W stat
DYS19	45	5	12-16	0.6111	1.0000
DYS385a	37	4	10-13	0.4460	1.0000
DYS385b	39	5	13-17	0.6437	1.0000
DYS389I	44	3	12-14	0.5867	1.0000
DYS389II	45	4	28-33	0.6758	0.6667
DYS390	43	4	22-25	0.7497	1.0000
DYS391	45	3	10-12	0.5687	1.0000
DYS392	46	5	10-14	0.6348	1.0000
DYS393	44	3	12-14	0.4440	1.0000
DYS438	46	5	9-13	0.5575	1.0000
DYS439	44	4	10-13	0.5941	1.0000
Mean	43.45	4.09	n/a	0.591995	0.9697

have a G-W statistic of one except DYS389II which has a G-W value of 0.6667. This value in DYS389II is due to the absence of individuals with 31 and 32 repeats at the loci.

Figure 23 provides a visual comparison of the haplogroup composition for the Kansas City Games group and the four regions of Scotland, examined by Goodacre et al. (2005). The data from Capelli et al. (2003) were excluded from this comparison because haplogroups R1a and R1b could not be distinguished from one another. Haplogroup G is solely found in the Kansas City Games sample. Halpogroup R1a is found at a lower frequency in the Kansas City Games group when compared to Scotland. Haplogroups D/E, largely represented by the subclade E1b1b (formerly E3b), is found at similar frequencies in the KC group as the Scottish populations, except for the Western Isles and Skye where it is absent. The combined frequency of haplogroups I/J in the Kansas City Games group, at 17 percent, is above the mean.

This frequency falls between the ranges of I/J seen in the Shetland (19 percent) and the Scottish mainland with a frequency of 16 percent. The Northwestern Coast of Scotland has an I/J frequency of 12 percent, and the Western Isles and the Isle of Skye exhibit a frequency of 11 percent.

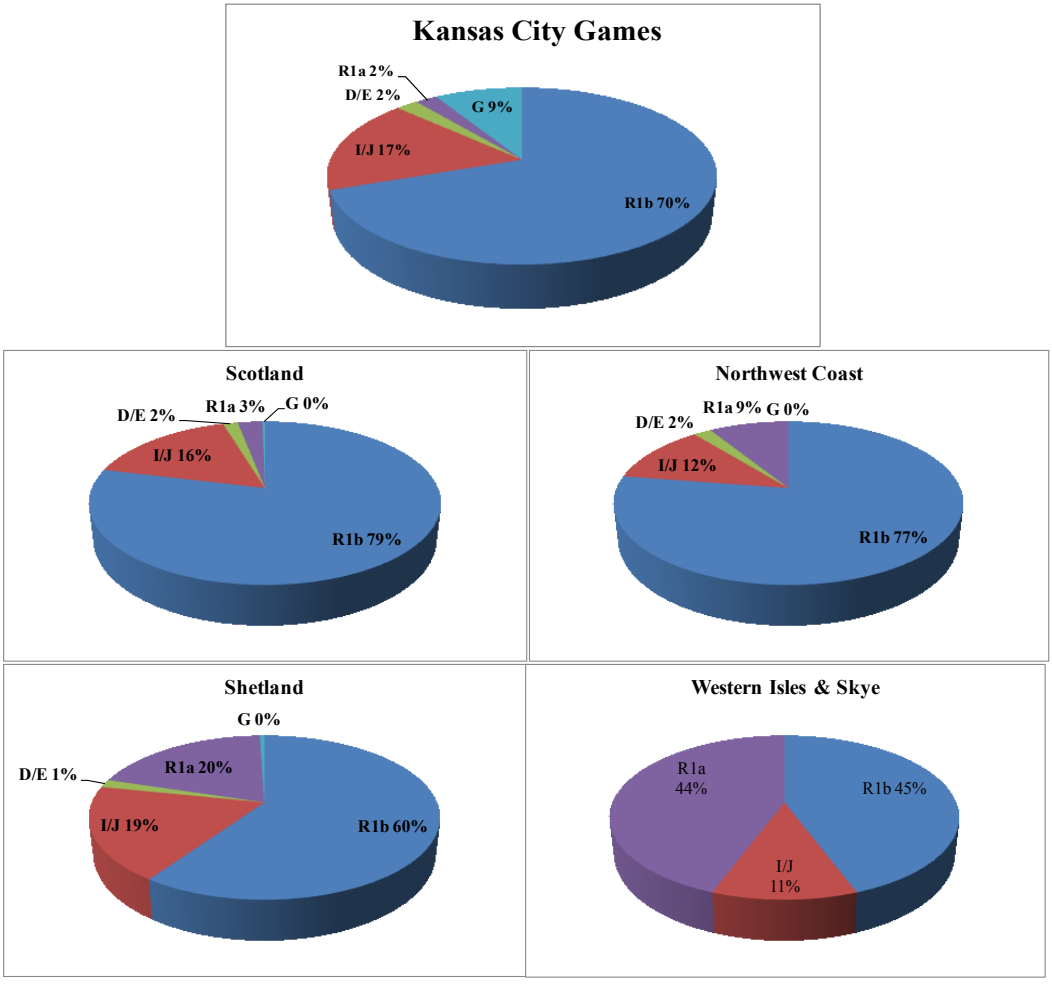


Figure 23. Y-haplogroup frequencies for the Kansas City Highland Games population and four regions of Scotland (Goodacre, et al., 2005).

Table 10. Summary statistics of Y-STR expected heterozygosity (*H*) and the mean number of pairwise differences (π) for loci DYS19, DYS390, DYS391, DYS392 and DYS393. The number of samples (n) and the number of haplotypes (#hts) are given.

Population	n	#hts	*H*	sd(*H*)	π	sd(π)
KC Games	43	30	0.9812	0.0089	2.1872	1.2344
Scotland Mainland	495	99	0.9059	0.0082	2.1683	1.2046
Orkney	121	48	0.9507	0.0089	2.6208	1.4111
W Isles & Skye	160	47	0.9285	0.0120	2.4722	1.3431
Shetland	256	58	0.9544	0.0053	2.6605	1.4232
Northwest Coast	154	50	0.9339	0.0095	2.3401	1.2851
Mean	205	55.3	0.94243	0.0088	2.4082	1.3169

Summary statistics of Y-STR diversity of forty-three Kansas City Games males and five Scottish populations are given in Table 10. Although the sample size of the KC Games group is small (n=43), the *H* is higher here at 0.9812 than in any Scottish population. This indicates that some of the samples may have a paternal ancestry outside of Scotland. The population with the second highest *H* is Shetland. The lowest diversity is seen in the Scottish mainland where *H* equals 0.9059. The number of pairwise differences between haplotypes (π) in the Kansas City games group is low compared to the Scottish populations.

The frequency of shared Y-STR haplotypes with Scottish populations is higher than the frequency of shared mtDNA haplotypes in the Kansas City Games group. Figure 24 shows the percent of shared Kansas City and Scottish haplotypes based on five loci. Seventy-seven percent of the Kansas City Games haplotypes are shared with the Scotland mainland, 74 percent are shared with Shetland, and 60, 53,

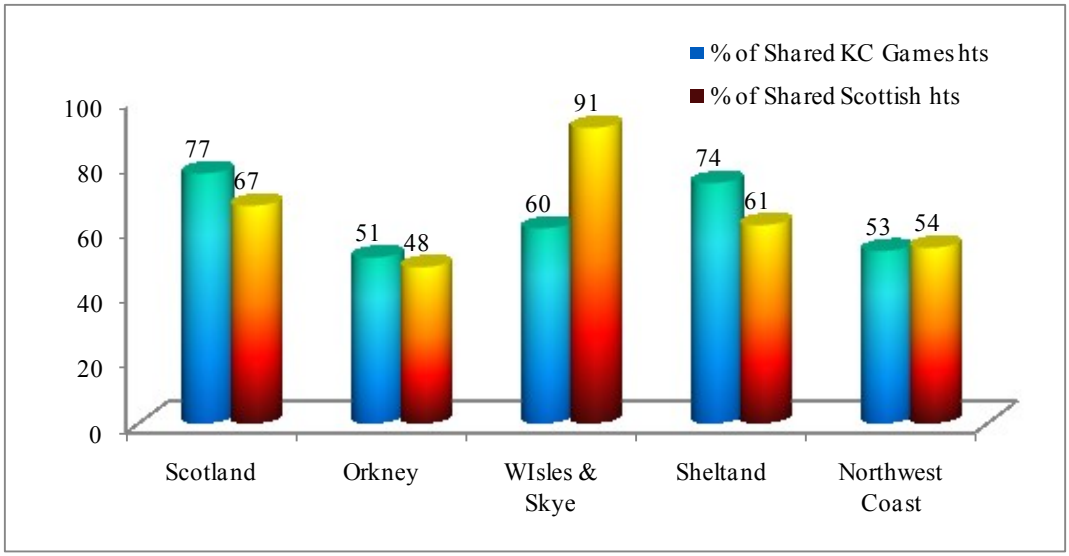


Figure 24. Percent of shared 5 loci Y-STR haplotypes between Scotland and the Kansas City Games group.

and 51 percent are shared with the Western Isles and the Isle of Skye, the Northwest Coast, and Orkney respectively. The Western Isles and the Isle of Skye share the largest number of haplotypes with the Kansas City games group with over 90 percent matching a KC Games haplotype. This may reflect the heavy historical migrations from these islands, noted by Murdoch (1998), and the absence of economic opportunities for Highland men following the Highland Clearances. At 48 percent, Orkney shares the least number of haplotypes with the KC group.

A NJ tree was generated using Slatkin's R_{ST} distances for microsatellite data and is presented in Figure 25. In this plot, the Kansas City Games group clusters with the northern island populations of Shetland, Orkney, the Western Isles and the Isles of Skye. This grouping is consistent with a similarity of the KC Games group to populations that were the most likely to have participated in the migrations across the

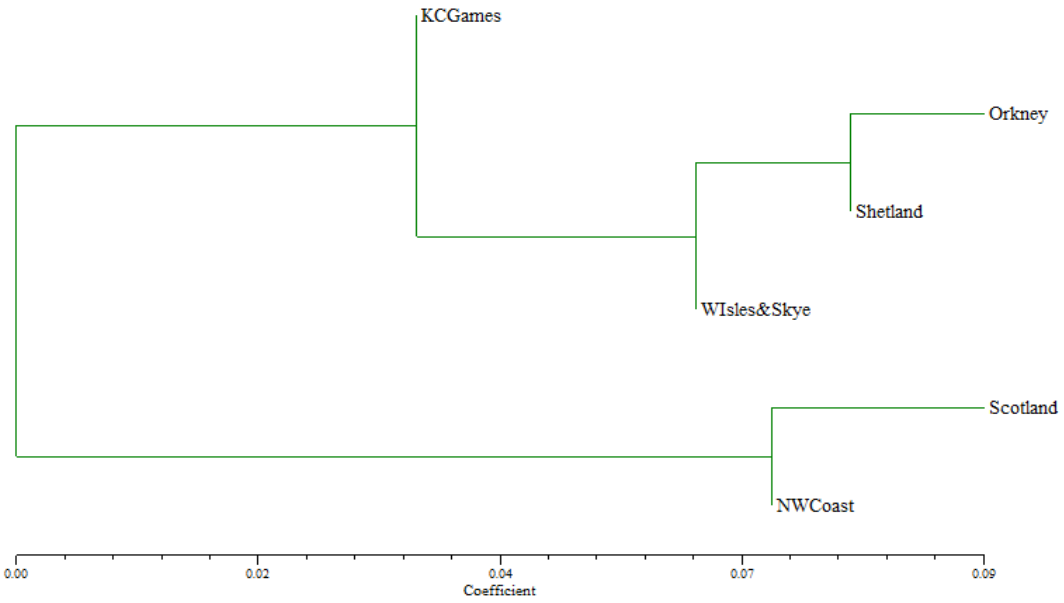


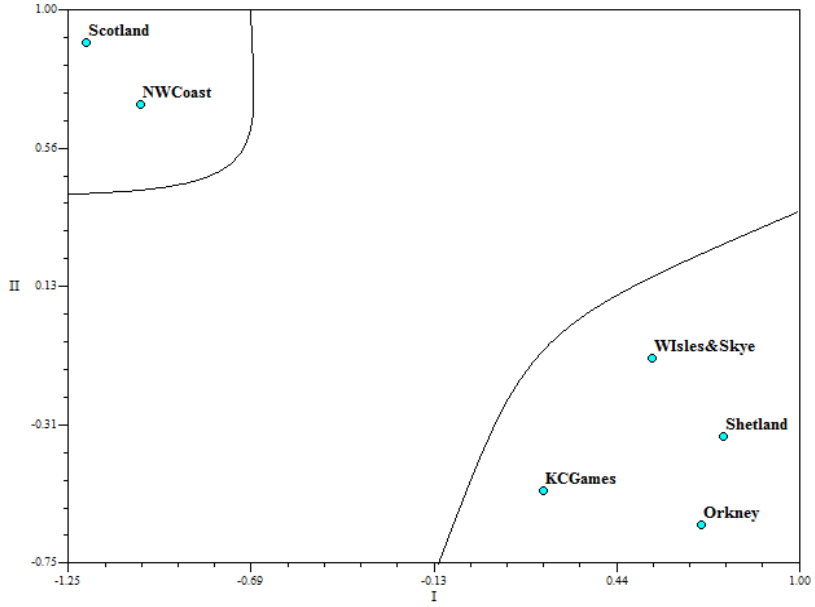
Figure 25. NJ tree based on corrected R_{ST} distances between the Kansas City Games group and five Scottish populations.

Atlantic to the United States, particularly those from Skye. The Scottish mainland and the Northwest Coast form a separate cluster, sharing the same branch on the NJ tree. However, the maximum distance between any two points on the plot was found between Orkney and the Kansas City Games group. A two-way Mantel test between the original distance matrix and the cophenetic distance matrix gave a correlation value of 0.83273, showing a relatively good fit to the data.

A MDS plot showing the relationship between the Kansas City Games Group and the Scottish regions was constructed and are shown in Figure 26. In this plot, the KC group clusters with the northern islands of Scotland including Orkney, Shetland, and the Western Isles and the Isle of Skye, consistent with the results from the NJ-tree shown above. The Scottish mainland and the Northwest Coast form a cluster in the upper left corner, and both of these groups show the greatest diversity of alleles at the

five loci examined. Dispersal along the first axis is influenced by variation at the DYS392 locus, where the Scottish mainland has the highest frequencies of alleles 13 and 14, and the lowest frequencies of alleles 11 and 12. In contrast, Shetland exhibits the opposite with the lowest frequencies of 13 and 14 and the highest frequencies of alleles 11 and 12. DYS19 also contributes to this axis, with allele 14 having a relatively high frequency in Scotland compared to the other populations. The second axis is separated by allele variation at the DYS391 locus. Orkney displays the highest frequency of allele 10 and the lowest combined frequency of alleles 11 and 12 when compared to all other groups represented in the plot. The variation at each locus can be seen in Figure 27. For clarity, the Northwest Coast and the Western Isles and Isle of Skye were left out of these plots. The final stress for the MDS plot was 0.01652, below the suggested maximum stress value (Sturrock & Rocha, 2000).

Figure 26. MDS Plot of Kansas City Games group and Scottish Populations based on corrected R_{ST} values. Final stress=0.01652.



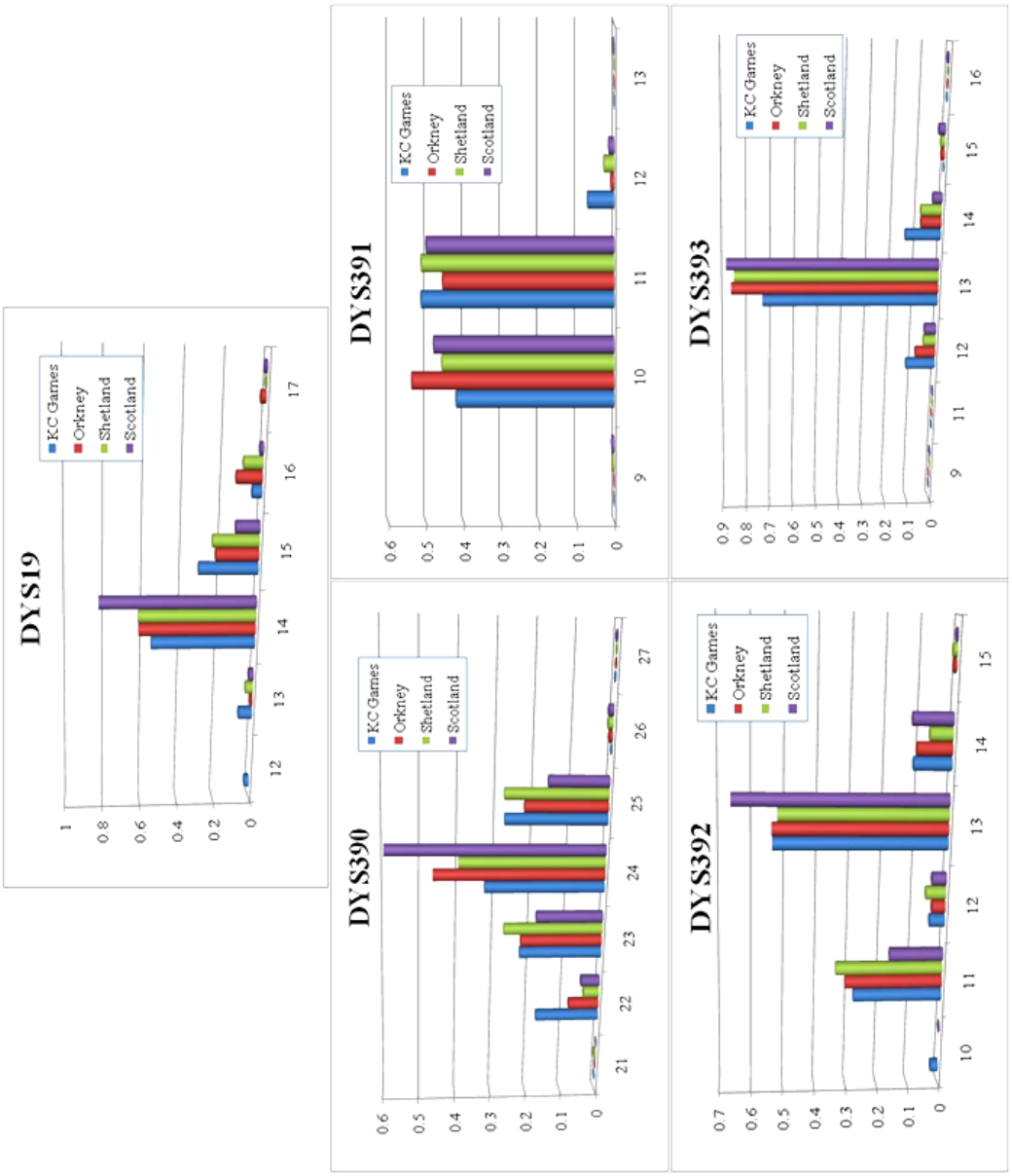


Figure 27. Allele distributions for DYS19, DYS390, DYS391, DYS392, and DYS393 in the Kansas City Game group, Orkney, Shetland, and the Scottish mainland (Scotland).

Table 9. AMOVA results for comparison of three groups (KC Games, the Scottish Mainland (including the Northwest Coast), and the Scottish Isles (Orkney, Shetland, and the Western Isles and Isle of Skye) based on data from five loci.

Source of Variation	d.f.	Sum of squares	Variance Components	Percent of Variation	*F*-Statistic
Among groups	2	52.543	0.07233	2.56	$\Phi_{CT} = 0.02646$**
Within Groups	3	8.307	0.00057	0.02	$\Phi_{SC} = 0.00021$*
Within Populations	11223	3253.721	2.66044	97.33	$\Phi_{ST} = 0.02667$*

*p <0.05, **p < 0.001

An AMOVA (analysis of molecular variance) was computed to examine the genetic structure of the Kansas City Games group and the five Scottish regions under study. Three groups were used in this analysis including the KC Games sample as one group, the Scottish mainland which also consisted of the Northwest Coastal region, and the Scottish Isles which consisted of Orkney, Shetland, the Western Isles and the Isle of Skye. Based on this analysis, only 2.56 percent of the variation can be explained due to variation among the groups, with a significant p-value that was less than 0.001 (See Table 10). The amount of variation among populations within the groups was also significant, but only accounted for 0.02 percent of the total variation in the sample. As with the AMOVA results for mtDNA sequences, the largest contributor to the overall variance (97.33 percent) was found within the populations, showing that there is little variation among groups.

For a broader view of the relationship of males in the KC Games sample with Europe, further analyses were computed to compare the male KC Games group with

thirteen European populations. The summary statistics of all European populations are displayed in Table 10, and show that the Kansas City Games group has a high level of genetic diversity (*H*) despite its relatively small sample size. Based on five loci, the diversity of the KC group is 0.9812, higher than any other population analyzed. When this measure is expanded to include seven loci, *H* increases to 0.9919, a number higher than all other populations compared in this study, except Italy. However when five or seven loci are used, the number of pairwise differences between haplotypes (π) is low for the KC group when compared to the rest of Europe. This result is expected if some of these males who claim Scottish identity are not Scottish through paternal ancestry. If this is the case, then you would expect to see a high level of diversity in the small sample because the group would represent a variety of European populations instead of a group with a single geographical origin.

Table 10. Within population Y-STR expected diversity (*H*) and number of pairwise differences between haplotypes (π) for five loci (DYS19, DYS390, DYS391, DYS392, DYS393) and seven loci (DYS389I and DYS389II included). The number of samples (n) and number of haplotypes (#hts) are given.

Population	n	#hts	#hts	*H*	sd(*H*)	*H*	sd(*H*)	π	sd(π)	π	sd(π)
KCGames	42	30	36	0.9812	0.0089	0.9919	0.0074	2.1872	1.2344	4.0592	2.0662
Scotland	1186	199	n/a	0.9222	0.0041	n/a	n/a	2.4143	1.3117	n/a	n/a
Germany	3442	534	1064	0.978	0.0008	0.9915	0.0005	2.9364	1.5394	4.2012	2.0873
Netherlands	275	92	136	0.9495	0.0062	0.9767	0.0038	2.6599	1.4226	3.8293	1.9364
Sweden	708	185	304	0.9655	0.0032	0.9818	0.0023	2.8146	1.4876	4.1884	2.0837
Belgim	125	57	87	0.9525	0.0106	0.9855	0.0047	2.7241	1.4560	3.9391	1.9867
Poland	1313	305	464	0.9661	0.0022	0.9878	0.0010	2.6591	1.4044	3.7299	1.8845
Norway	300	107	164	0.9654	0.0049	0.9859	0.0028	2.7157	1.4467	4.0831	2.0414
Italy	1340	351	692	0.9804	0.0016	0.9933	0.0008	3.1012	1.6117	4.3326	2.1448
Denmark	63	30	43	0.9493	0.0138	0.9811	0.0076	2.5422	1.3850	3.7921	1.9355
England	247	82	143	0.9468	0.008	0.9824	0.0040	2.5753	1.3860	3.7942	1.9174
France	208	102	138	0.9696	0.0061	0.9897	0.0025	2.9734	1.5613	4.1993	2.0943
Ireland	107	39	61	0.9178	0.018	0.9601	0.0122	2.2892	1.2655	3.3312	1.7238
Mean	720	163	277.667	0.957254	0.0068	0.984	0.00413	2.6609579	1.424	3.9566	1.9918

*** White columns are for values calculated using 5 loci. Gray columns indicate values calculated for 7 loci.**

The percent of shared Kansas City haplotypes with European populations, and shared European haplotypes with the Kansas City Games group are shown in Figure 28. Again, this analysis examines the relationships based on a five allele system and a seven allele system. As expected, the amount of haplotype sharing decreased as the number of STRs used increased. The Kansas City group shares over 70 percent of its seven allele haplotypes with Germany. The population that shares the most of its seven allele haplotypes with KC Games group is Ireland with 41 percent. All of these haplotypes belong to haplogroup R1b, which attains frequencies over 90 percent in some parts of Ireland (Sykes, 2006). When the KC group is compared to Scotland using five STRs, they shared 81 percent of the Scottish-American haplotypes, and the Scottish samples included in this analysis shared 79 percent of its haplotypes with the Kansas City group. However, matches of YSTR haplotypes alone are not necessarily an indication of Scottish paternal ancestry. Six of the individuals belonging to haplogroup R1b share a haplotype with all five regions of Scotland and all of the European populations examined in this study. These individuals carry some of the most common Y-STR haplotypes found in Europe. There are also two individuals, with Scottish surnames and a listed clan affiliation, who do not share a haplotype with any population examined.

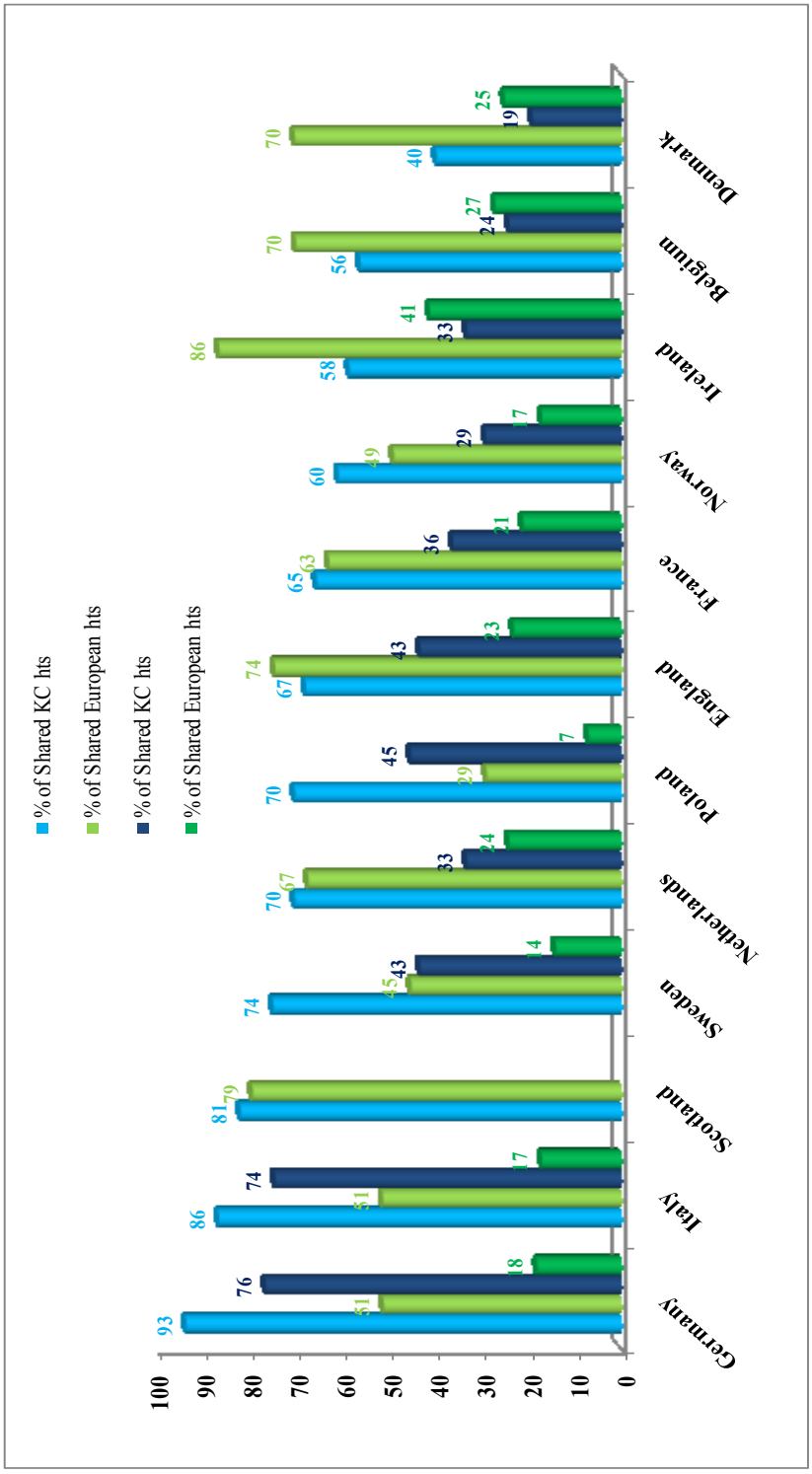


Figure 28. Percent of shared five and seven loci Y-STR haplotypes. Light blue and light green bars represent matches at 5 loci, dark blue and dark green represent matches at 7 loci.

Figure 29 provides a NJ-tree of the Kansas City Games group and thirteen European populations. The North Atlantic Islands form a cluster that includes England, Scotland and Ireland. The Netherlands, Denmark and Belgium encompass this group forming the next cluster. The Kansas City Games group is the final addition to the group. Poland and Norway make up separate cluster, with Poland having the greatest distance from the Kansas City Games group. These clusters are in rough agreement with geographical proximity and known historical relationships between the populations included. A two-way Mantel test between the corrected R_{ST} distance matrix and the cophenetic distance matrix produced by the tree gave a correlation value of 0.56689, showing the distortion of data in the matrix. This implies that another form of representing the data might better describe the relationship between these populations.

Multidimensional scaling was also used to examine the relationship between the Kansas City Games group and European populations, and is shown in Figure 30. Unlike the plot of mitochondrial distances between populations, the plot based on distances between Y-STR haplotypes follows a general geographical cline of Eastern to Western Europe. There are also groupings of major European regions, and these are labeled in Figure 30 as Eastern Europe, Central Europe, Scandinavian, Dutch, and North Atlantic Islands. The Kansas City Games group clusters with the Dutch populations of Belgium and the Netherlands. The KC group is also centered among the Dutch, Central European, North Atlantic, and Scandinavian groups, all of which

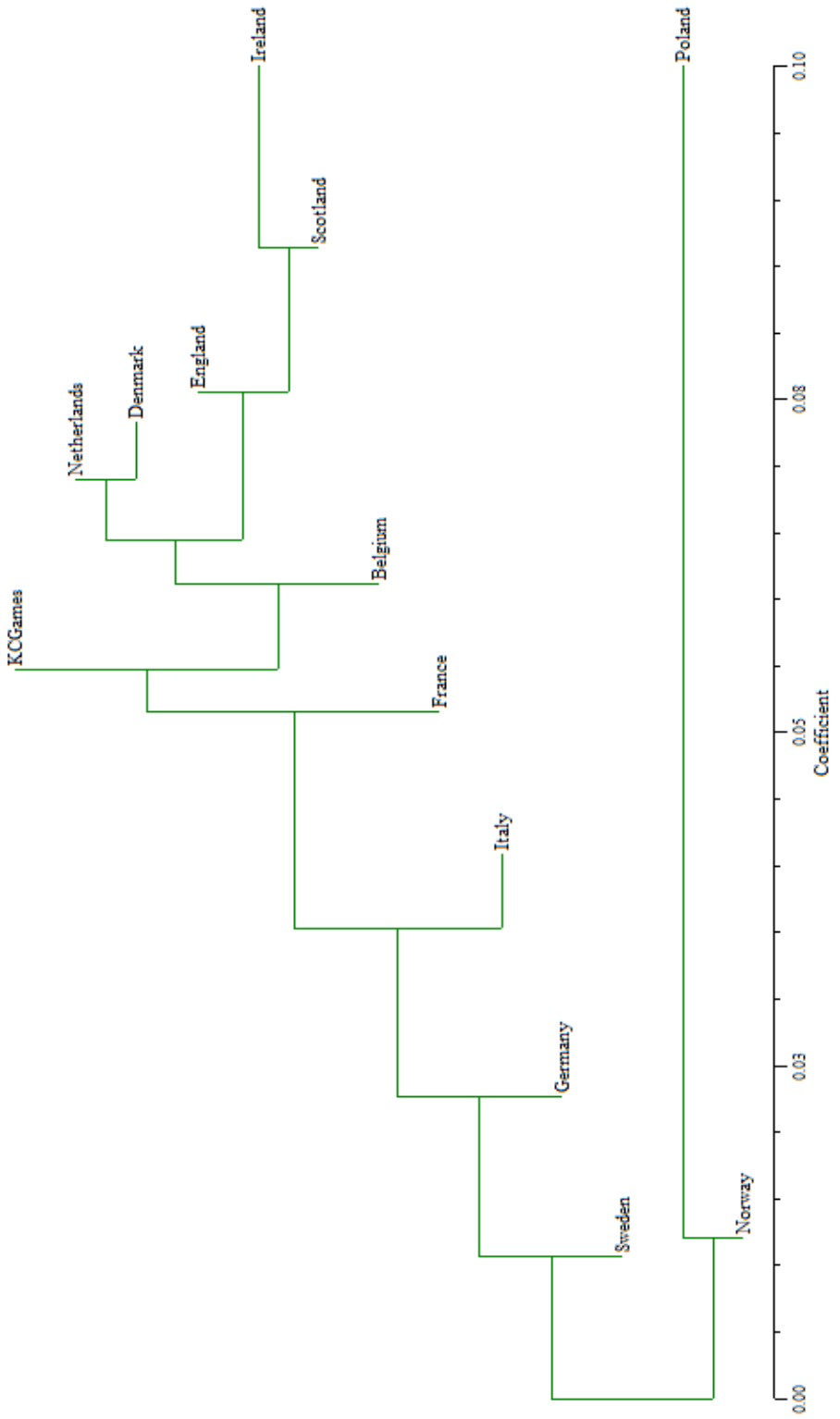


Figure 29. NJ tree based on RST distances between European populations.

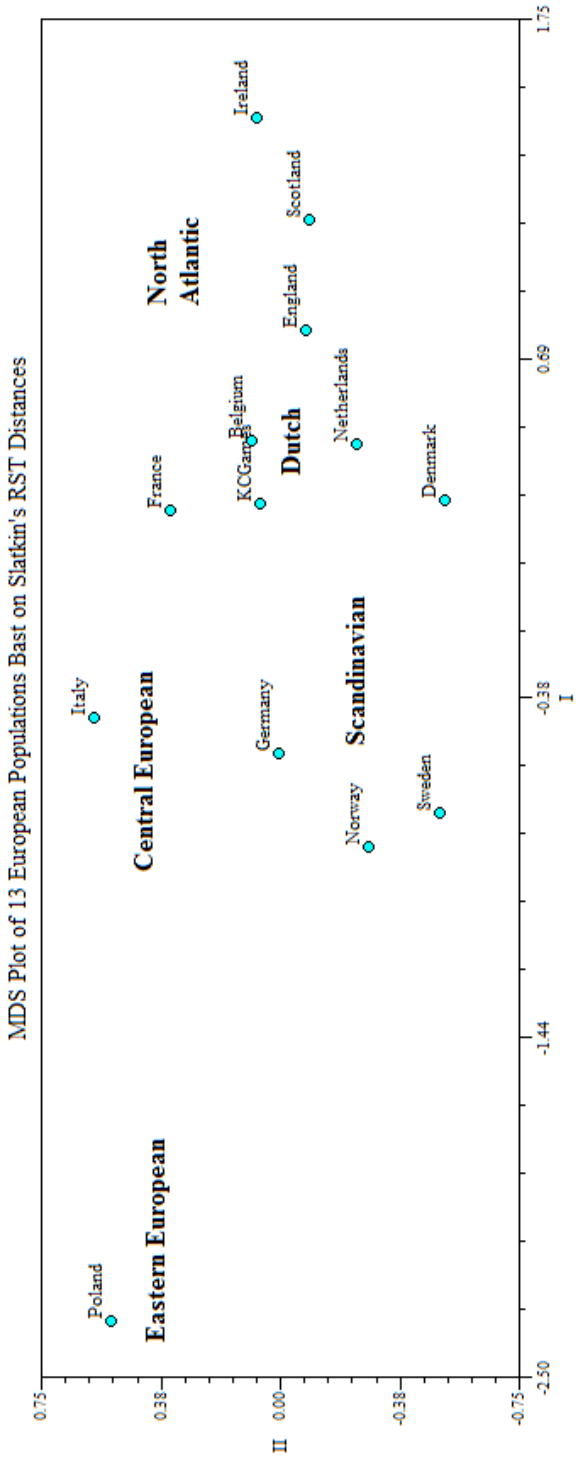


Figure 30. MDS Plot based on Slakin's R_{ST} Distances of European populations using five loci. Final stress = 0.03274.

share 40 percent or more of these region's haplotypes. Poland is separated from the Scandinavian and Central European clusters along the first axis of the plot as a result of its high frequencies of haplogroup R1a. The axes and groupings on the plot may further be explained by unique allelic frequencies at specific loci. For example, the first axis is a gradient of DYS19 alleles 14 and 16. In the Polish sample, allele 14 has a low frequency (0.19512) compared to the other samples, and allele 16 is found at the highest frequency (0.33003). The reverse is true of Ireland which has the highest frequency of allele 14 (0.83178) and a low frequency of allele 16 (0.02804). A similar distribution is seen for alleles at the DYS391 locus where Poland has a high frequency of allele 10 when compared to other populations (see Appendix B).

Chapter Summary

Y-STR data appears to be a better indicator of Scottish ancestry than mtDNA data. Haplotype sharing with all regions of Scotland is higher for YSTR haplotypes at 77 percent, than for mtDNA haplotypes at 56 percent. Y-STR haplotypes, as with mtDNA haplotypes, are shared between populations of Europe. However, the Y-haplotypes are more localized in their distribution, and are not as widespread on the European continent. The Scottish-American mtDNA haplotypes more closely resemble those found on the Scottish mainland, a region which shares a large portion of its lineages with continental Europe. When compared to European populations in a MDS plot, the KC Games group appears in the center of the European populations. From this, one can say that the KC Games group as a whole appears European.

However, identifying a particular region or population to which it most resembles is not possible with the data obtained in this study. Analyses utilizing Y-STR data indicate that the Scottish-American group most closely resembles regions of the Scottish Isles than to the Scottish mainland. This would be the expected origin of Scottish males based on historical migrations. However, the KC group as a whole more closely resembles Dutch populations when compared to European populations on a MDS plot. These results indicate that a Scottish cultural identity is more likely to be passed through a paternal, rather than maternal, relationship. Summary tables of all markers, matching haplotypes in Scotland, and clan affiliations are provided in Appendix C.

Chapter Five: Discussion

This chapter presents the results of the study within the context of Scottish identity as claimed by the Kansas City Highland Games group. First, this chapter addresses maternal and paternal origins of individuals of this study, and discusses the differences when compared to different regions of Scotland based on the historical record. This chapter also discusses other possible contributors to the Scottish-American gene pool. The possibility that each individual is of Scottish maternal or paternal decent has been evaluated based on a combination of molecular data, surname information, and an association with a Scottish clan.

Do individuals in this study have maternal markers found in Scotland?

The historical record indicates that individuals with a maternal ancestry from Scotland should largely be from the Irish settlement of Ulster, Scots of the Border Clans, or have ancestry from the Western Isles or Central Highlands of Scotland (Murdoch, 1998; Landsman, 1999; McDonald & McDonald, 1980). Female migration from Scotland was largely in the form of family immigration, as single females tended to seek work in local economic activities such as domestic labor or textile manufacturing in Scotland's industrial centers, and seasonal agricultural or fishery work in the Lowlands or coastal areas. In contrast to Scottish female movement, the number of single women from Ireland that came to the colonies

outnumbered the number of Irish men (Gordon, 2002). While exact numbers of how many of these women were from the Ulster plantations are not available, there were probably many single females involved in the large movement from Ulster to the New World.

Appendix A provides a distribution of the clans of participants in the mtDNA analyses in this study. Clans found in the Highlands of Scotland *versus* the Lowlands of Scotland are present at roughly a 4:1 ratio, which would be expected since the samples were collected at the Kansas City Highland Games. As expected from the historical record, many of these clans are located along the Scottish/English border, in the Western Isles of Argyll, the Isle of Skye, and the central Highlands of Inverness. Given these distributions, we would expect to see a maternal relationship to the Western Isles and the Isles of Skye and the Scottish mainland.

Examination of HVS-I sequence information and RFLP data indicate that 60 percent of the individuals in this study have HVS-I haplotypes that are found throughout Scotland, in particular the Scottish mainland and the Northwest coast of Scotland. As with the rest of Europe, the Kansas City Games group has a high frequency of haplogroup H, which was dispersed through Europe following the repopulation of Europe out of the Iberian refugia. The KC group does not show a relationship to the Western Isles or the Isles of Sky in either the NJ tree or MDS plot based on Tamura and Nei's genetic distances. The haplotypes that are found in Scotland mainland are from the mtDNA haplogroups H, V, J, K, and U. However, as noted in the results chapter, there are only three haplotypes that are found exclusively

in Scotland. The rest are found throughout Europe, and are probably the reason that the KC Games group has a high level of expected diversity when compared to the mean expected diversity of all of the populations in this study.

Do individuals claim Scottish ancestry along the maternal line?

Maybe, but the mix of European haplotypes, and the location of the Kansas City Game group in the MDS plot in Figure 22, suggests that the maternal ancestry of these participants as a whole are co-mingled with broader European ancestry. Single male immigrants from Scotland outnumbered female immigrants to the New World, and it is likely that they had to find mates from other immigrating groups. Also, many of the mitochondrial lineages that are found in Scotland are shared with several European populations, and this sharing would only partially be resolved through further RFLP or SNP analyses that would more fully characterize each haplotype. Not only are the lineages shared throughout Europe, but the diversity of all mtDNA lineages in Europe, particularly haplogroups H and U5, make it difficult to determine if the Scottish comparative samples are an accurate representation of the diversity of Scotland. Helgason et al. (2001) noted that several studies have indicated that there are a large number of mtDNA lineages throughout Europe. This variation makes it difficult for any study to capture the true diversity within a population, regardless of a large sample sizes. The importance of this finding in relationship to this study is that there may be several mtDNA lineages within Scotland that have not yet been identified, especially in areas of the British Isles that have not been as heavily

sampled as the mainland. Furthermore, recent gene flow, and a possible loss of diversity due to migration, has left Scotland with few private mtDNA lineages. Obtaining more complete genealogical information from each of the participants would have helped to distinguish between a haplotype match that is due to its high frequency in Europe or a haplotype found within Scotland. Genealogical information may also have helped to identify lineages that have not been recognized as Scottish in previous studies. Combined, genealogical and genetic information could have indicated which mtDNA lines are of Scottish origin. Another possible improvement to the study would be to have comparative samples that represented the Highlands and Lowlands separately, instead of the Northwest Coast and Scottish mainland. However, this information was not available in the literature, and all of the individuals from both the Highlands and the Lowlands would have to be screened for migrations to and from each region over the last three hundred or so years.

Do individuals who claim Scottish identity have Scottish Y-markers**?**

Similar to female emigrants of Scotland, a large portion of males from the Central Highlands, and the Western Isles moved to the colonies. Many Scotch-Irish males, descendants of Lowlanders brought to Ireland in the 1600s, immigrated to the New World from Ulster (McDonald & McDonald, 1980). From the historical record one would expect to find males of paternal descent from Scotland most similar to

these regions of Scotland, with some bias towards the Highland since the samples were obtained from the Kansas City Highland Games.

As expected, all of the major haplogroups found in Scotland, R1b, R1a, I, J, E1b1b, were detected in the Kansas City Games group. Haplogroup R1b is the most common haplogroup found in the Kansas City Games group at 70 percent, a high frequency typical of Western European populations due to its spread by male foragers following the LGM. These groups are less genetically influenced by the later movements of people during the Neolithic and the spread of agriculture. The most surprising results are the three individuals who belong to haplogroup G2a, a haplogroup that is not seen in the comparative Scottish populations in this study, but has been found in Scotland at very low frequencies. It has been suggested that this haplogroup was the result of Roman interactions in Britain, and later migrations from the descendants of the Romans to Scotland. Two of these three individuals have common Scottish surnames and Scottish clan affiliations. These results suggest that the genetic signatures are probably a mark of today's Scottish population, influenced by many different movements of people and shaped by disease, war and outmigration, instead of a Scottish ancestral claim based on a non maternal/paternal relationship or one that is strictly cultural in nature. The other individual belonging to the G2a haplogroup has a surname that is found in Scotland, but more commonly seen throughout England.

As a whole, analysis of Y-markers showed a greater resemblance to Scottish populations than did mtDNA markers. Haplotype matching of five loci DYS19,

DYS390, DYS391, DYS392, and DYS393, show that the KC games group shares the largest percent of its haplotypes with the Scottish mainland. However, of the Scottish groups, the Western Isles and the Isles of Skye share 91 percent of its haplotypes with the KC group. These results alone are insufficient to support a paternal line through Scotland, since many of these haplotypes are found in populations throughout Europe. However, this strong resemblance to groups that have historically contributed to the American gene pool, and the listed clan affiliations that are distributed throughout these islands are indications that many individuals from the KC sample are of direct Scottish paternal descent.

The distribution of clans reported by male participants reflect the historical records, and are included Appendix A. The Highland to Lowland clan distributions of the male participants of this study are at a ratio of 4.56:1, slightly higher than the ratio of participants included for mtDNA testing. This may be due to a higher number of individuals who claim direct Scottish paternal descent over individuals that may be of Scotch-Irish descent. The high Highland to Lowland clan variation is the result of the location of the sampling which celebrates the success of the Highlander. Of the 41 individuals that listed a Highland clan as their clan affiliation, 13 belong to clans that are found in the Western Isles (Clan Stewart, Clan MacDonald and Clan Gunn), the Isle of Skye (Clan MacDonald), and Orkney (Clan Sinclair and Clan Gunn). Clan affiliation is usually tied to paternal relationships, and these regions are found to be most similar to the Kansas City Games group when compared to other parts of Scotland in a NJ tree and MDS plot (Figures 25 & 26). Shetland is an

exception to this due to its long affiliation with Vikings, and the resulting dissipation of the clan system there. However, the Shetland group also clusters with the KC group and the other islands of Scotland.

The high levels of Nei's gene diversity ($H = 0.9812$) found in the small sized sample indicates that there are some haplotypes from other parts of Europe mixed with the Scottish haplotypes of the Kansas City Games group samples. These results rely on a few loci, and a different picture may emerge if additional markers were employed. Search for the full haplotypes were performed at various web databases such as ysearch.org, ybase.org, and yhrd.org, revealing that samples with matches at five loci in Scotland did not match at 11 loci. In a few cases no matches were found in any population in Europe. This could be due to the sampling of the databases, or it could be the result of the loss of diversity in the Scottish population since the 1700s. If a more complete STR and SNP profile were analyzed, it is likely that a majority of the KC Group would match haplotypes in Scotland or be a single mutational step away from founding haplotypes.

What's in a name?

Another important source of information that can be considered in studies involving the Y chromosome is an individual's surname. Males pass on both Y chromosomes and surnames to all of their sons. A study by King et al. (2006) was performed in the British Isles and used presumably unrelated individuals of the same

surname to determine if sharing a surname increased the chances of having the same Y-haplogroup. The study showed that sharing a surname, even a common occupational surname such as Smith, significantly increased the chance that two individuals with that surname will also share a haplogroup. For rare surnames, 47 percent of the pairs shared a common haplogroup, and the rate of sharing is even higher for the rare haplogroups. Haplogroups such as R1a, G, D, E, J2 and K are often shared by individuals of the same surname supporting the concept that surname sharing is often due to common ancestry (King, et al., 2006).

The summary tables listed in Appendix C provides a method of identifying individuals who have a surname found in Scotland in 1881, the regions where those surnames were most common, the clan affiliation indicated on participant's questionaires, and the distribution of the haplotype that was found in Scotland by Capelli et al. (2003) and Goodacre et al. (2005). Out of the 43 participants used in Y-STR analyses, 24 have a five loci haplotype, surname, and clan affiliation that identifies them as paternally Scottish. Approximately one half of these individuals also have mtDNA haplotypes that are found in Scotland. Fifty-six percent of the male participants in this study are identifying themselves as Scottish through a paternal relationship. Approximately one half of those males may also be able to claim Scottish descent through a maternal relationship. Only five of the individuals who have a mtDNA and Y-chromosome haplotype found in Scotland claim more than one clan affiliation. This is strong support for a Scottish identity through a paternal line.

The MDS plot and NJ tree based on Slatkin's R_{ST} distances cluster the Kansas City Games group with the Western Isles and Isle of Skye, Orkney, and Shetland. Based on the historical record, the clusters with the Western Isles and the Isle of Sky makes sense due to the heavy immigration to the United States from these areas. The relationship to Orkney and Shetland is less clear, especially Shetland where the clan system dissipated during Viking interactions. However, it is the influence of the Vikings that is heaviest in all of these regions of Scotland, as the paternal genetic signatures show. Surnames in this region display marks of Scandinavian influence with names ending in –son and –s common, particulaly in Shetland and Orkney and explain part of this relationship (See Apendix B for distribution of surnames). The KC group shares about one half of its surnames with surnames that make up 75% of the population in the Shetland Islanders. According to the census bureau of Scotland (http://www.scrol.gov.uk/scrol/common/home.jsp), Shetlanders have a significantly different distribution of surnames when compared with the rest of Scotland. If the KC games sample is similar in both surnames and Y-STR polymorphisms to Shetland and the other Scottish Isles due to Viking influence, it explains their grouping on the MDS plot.

What about individuals who do not share maternal or paternal markers with Scotland?

Some of the haplotypes found in this study do not appear to be of Scottish origin. However, not finding a match in the Scottish population does not exclude a

sample originating from Scotland. Instead, this could be a result of sampling in the comparative data or a result of kin-structured migration. As mentioned above, mtDNA diversity is high in Europe and there is some chance that lineages found in this study were brought to the United States by individuals from Scotland, yet their particular haplotypes are not represented in the comparative samples due to a small sample size. For Y-chromosome diversity, inclusion of SNPs would have added more conclusive evidence for Scottish paternal ancestry, especially in distinguishing between subclades of haplogroup R1b. Kin-structured migration is a form of migration in which migrating groups are related individuals and thus carry similar genes (Mielke & Fix, 2006). While early movements of Scottish immigrants were typical of single males, family based movements from areas of the Potato Famine and the Highland Clearances put an economic strain on the population, became typical in the 1800s. As a result of these events, it is possible that much genetic diversity, including that found in the Y-chromosome and mtDNA, left Scotland and the Scottish Isles and can now only be found in emigrants from Scotland, who left for better opportunities overseas.

Chapter Summary

The results of this study indicate that individuals who attended the Kansas City Highland Games are more likely to claim ancestry from Scotland based on a paternal, rather than maternal, Scottish relationship. This is expected from the

historical record as greater numbers of single males from Scotland immigrated to the New World than single females. MtDNA analyses show that the KC group clustered with the Scottish mainland and the Northwest coast of Scotland, an indication that those of Scottish maternal descent are likely immigrants from the region of Inverness. However, many of these haplotypes are shared with other European populations, as evident through the comparisons with European populations, and this association may be due to a relationship with admixture between continental populations, both before and following migration to the New World, instead of Scotland. Information for the paternal side can be supported using clan affiliations and surname distributions. These three lines of evidence suggest that over one half of the males in this study are of likely Scottish descent, and that a small majority of the rest may also claim a male lineage from Scotland. However, there are individuals who do not share maternal or paternal lines with Scotland. These individuals may be claiming Scottish ancestry through some other affiliation not captured by mtDNA or Y-chromosome markers, or based on cultural affiliations developed in the United States.

Chapter Six: Conclusion

This thesis investigates Scottish identity through maternal and paternal lineages. Mitochondrial DNA analyses included the use of HVS-I sequences and haplogroup defining restriction fragment length polymorphisms to investigate maternal ancestry. Y-chromosome short tandem repeats were utilized in combination with surnames and clan affiliations to investigate paternal Scottish ancestry. Together these markers reflect population movements to the Americas from the 17th through 20th centuries. Table 11 summarizes the results of the study.

Maternal Scottish ancestry is not expected to be as common as paternal ancestry in this study, as the historical record of female migration to the New World indicates fewer female than male immigrants. Those who did migrate directly from Scotland, particularly with their families, are expected to largely represent lineages

Table 11. Summary of mtDNA and Y-chromosome analyses.

Results	mtDNA	Y-chromosome
Haplotypes found in Scotland	60 percent	77 percent
MDS plot with Scottish populations	Resembles Scottish Maniland	Resembles Scottish Isles
Haplotypes found in Europe	Most (at 60 percent) are found in Scotland	Most (at 93 percent) are found in Germany
Highest Percent of European haplotypes found in this sudy	50 percent of Welsh haplotypes	90 percent of Irish haplotypes (all R1b)
MDS plot with Eruroean populations	Clusters with several European populations	Clusters with French and Dutch groups

found in the central Scottish Highlands and the Western Isles. Scotch-Irish females, should resemble individuals from the Lowlands of Scotland. Analyses showed that the KC group shares 60 percent of its haplotypes with Scottish populations. Thirty percent of these haplotypes are found in the Western Isles and the Isle of Skye, lower than any other Scottish population except Orkney. The Scottish mainland shares 56 percent of its lineages with the sample. If these haplotypes are the result of Scottish ancestry, then it appears the KC Games group most closely resembles Scotland's central Highlands and Lowland regions. The expected heterozygosity ($H = 0.9734$) of the Kansas City Games group, and the number of pairwise differences between haplotypes ($\pi = 4.9555$) were high compared to regions of Scotland. These diversity measures indicate that there are maternal lines found in the KC Group that are not represented in the Scottish sample, suggesting that there are other populations contributing to the maternal line. MDS plots and NJ-trees based on HVS-I sequences support this, as the KC groups shares lineages with all European comparative populations, and cluster with other populations besides Scotland in both analyses. Due to the high levels of haplotypes sharing amongst European populations, mtDNA data cannot identify lineages that are particularly "Scottish."

Paternal ancestry is expected to be higher for males of Scottish decent as male immigrants outnumbered female immigrants in the historical records. These males are expected to represent lineages in the Scottish central Highlands and the Scottish Isles. Results show that the genetic diversity of the KC Games group is high when compared to Scotland, an indication that there are some other European paternal

lineages represented by the sample. However, in contrast to the maternal line, the KC groups shares a larger number of haplotypes with Scotland (77 percent). These haplotypes are most common in the Western Isles and Isle of Skye, and the Scottish mainland (which includes the central Highlands). Analyses based on Slatkin's R_{ST} distances, including MDS plots and NJ-trees indicate that the KC group most closely resembles regions of Scotland where out-migration has been documented. These regions are also the areas of Scotland that had the strongest Viking genetic influence (Helgason, et al., 2000). This would explain the clustering of Shetland and Orkney with the KC games group, despite lower levels of emigration to the New World.

This study also utilized surname and clan distributions in support of a larger paternal *versus* maternal Scottish ancestry. Individuals with the same surname are more likely to be related genetically than individuals who do not share surnames (King, et al., 2006). The same can be said of clan affiliations which are often recognized through a paternal line. Distributions of male participants' surnames and clan affiliations, particularly those that have haplotypes found in Scotland, are most frequent in the Western Isles and the Central highlands. Many of these surnames are a result of Scandinavian influence and are found throughout Shetland, Orkney, and other Scottish Isles. Over one half of the male participants had a Scottish surname, clan affiliation, and a Y-STR haplotype found in Scotland. This suggests a strong Scottish paternal component to the KC Games group.

This study provided a method for investigating cultural identity through molecular markers. In this example, maternal and paternal relationships can be examined using genetic data in support of the historical record. However, Y-chromosome data, typically less diverse and more locally distributed within Europe, seems better suited than mtDNA data for an investigation of Scottish ancestry. More refinement in genetic analyses, including the inclusion of more RFLP or sequencing in the mitochondrial genome and the inclusion of SNPs in Y-chromosome analyses, may provide more informative in future studies. Despite European admixture before and after settlement in the New World, and the use of limited markers, Scottish maternal and paternal genetic influences can still be detected in participants of this study.

Works Cited

Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., et al. (2004). The Molecular Dissection of mtDNA Haplogroup H Confirms That the Franco-Canabrian Glacial Refuge Was a Major Source for the European Gene Pool. *American Journal of Human Genetics , 75*, 910-918.

Adams, J., & Faure H, Q. m. (Eds.). (1997). *Review and Atlas of Palaeovegetation: Preliminary Land Ecosystem Maps of the World Since the Last Glacial Maximum.* Retrieved July 7, 2009, from http://www.esd.ornl.gov/ern/qen/adams1.html

Anderson, M. (1996). British Population History, 1911-1991. In M. Anderson (Ed.), *British Population History: From the Black Death to the Present Day* (pp. 359-409). Cambridge: Cambridge University Press.

Anderson, M. (1996). Population Change in North-Western Europe, 1750-1850. In M. Anderson (Ed.), *British Population History: From the Black Death to the Present Day* (pp. 191-279). Cambridge: Cambridge University Press.

Andrews, R., Kubacka, I., Chinnery, P., Lightowlers, R., Turnbull, D., & Howell, N. (1999). Reanalysis and Revision of the Cambridge Reference Sequence for Human Mitochondrial DNA. *Nature Genetics , 23* (2), 147.

Arnold, B., & Murray, M. L. (?). "Put Out the Geese, the Celts are Coming" Iron Age Migration and Social Change in Central Europe. *Ancient Travellers: Proceedings of the Twenty-Seventh Annual Conference of the Archaeological Association of the University of Calgary* , 111-118.

Athey, T. W. (2006). Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele-Frequency Approach. *Journal of Genetic Genealogy , 2*, 34-39.

Athey, T. W. (2005). Haplogroup Prediction from Y-STR Values Using an Allele-Frequency Approach. *Journal of Genetic Genealogy , 1*, 1-7.

Ayub, Q., Mohyddin, A., Qamar, R., Mazhar, K., Zerjal, T., Mehdi, S. Q., et al. (2000). Identification and Characterisation of Novel Human Y-chromosomal Microsatellites from Sequence Database Information. *Nucleic Acids Research , 28* (2), e8.

Bandelt, H., Forster, P., & Rohl, A. (1999). Median-joining Networks for Inferring Intraspecific Phylogenies. *Molecular Biology and Evolution , 1*, 37-48.

Barbujani, G., & Bertorelle, G. (2001). Genetics and the population history of Europe. *Proceedings of the National Academy of Sciences , 98* (1), 22-25.

Basu, P. (2007). *Highland Homecomings: Genealogy and Heritage Tourism in the Scottish Diaspora.* New York: Routledge Taylor & Francis Group.

Bonsall, C., Macklin, M. G., Anderson, D. E., & Payton, R. W. (2002). Climate Change and the Adoption of Agriculture in North-West Europe. *European Journal of Archaeology , 5*, 9-23.

Brittingham, A., & de la Cruz, G. P. (2004). *Ancestry: 2000, Census 2000 Brief.* U.S. Census Bureau, U.S. Department of Commerce Economics and Statistics Administration.

Bruce, D. A. (1996). *The Mark of the Scots: Their Astonishing Contributions to History, Science, Democracy, Literature and the Arts.* Caral Publishing Group.

Campbell, K. D. (2007). Geographic Patterns of Haplogroup R1b in the British Isles. *Journal of Genetic Genealogy , 3*, 1-13.

Campbell, K. D. (2007). Geographic Patterns of R1b in the British Isles - Deconstructing Oppenheimer. *Journal of Genetic Genealogy , 3* (2), 63-71.

Cann, R. (1988). DNA and Human Origins. *Annual Review of Anthropology , 17*, 127-143.

Cann, R. L., Stoneking, M., & Wilson, A. C. (1987). Mitochondrial DNA and Human Evolution. *Nature , 325*, 31-36.

Capelli, C., Redhead, N., Abernethy, J. K., Gratrix, F., Wilson, J. F., Moen, T., et al. (2003). A Y Chromosome Census of the British Isles. *Current Biology , 13*, 979-984.

Chikhi, L., Nichols, R. A., Barbujanii, G., & Beaumont, M. A. (2002). Y Genetic Data Support the Neolithic Demic Diffusion Model. *Procedings of the National Academy of Science , 99* (17), 11008-11013.

Cohn, R. L. (1995). Occupational Evidence on the Causes of Immigration to the United States, 1836-1853. *Explorations in Economic History , 32*, 383-408.

Creighton, C. (1891). Plague in Scotland, 1495-1603. In *A History of Epidemics in Britain* (pp. 360-370). Cambridge: The University Press.

Currat, M., & Excoffier, L. (2005). The Effect of the Neolithic Expansion on European Molecular Diversity. *Procedings of the Royal Society B , 272*, 679-688.

de Knijff, P., Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., et al. (1997). Chromosome Y Microsatellites: Population Genetic and Evolutionary Aspects. *International Journal of Legal Medicine , 10* (3), 134-140.

Department of History, University of Leiden. (2009, January 21). *History of International Migration.* Retrieved February 2, 2009, from The Migration to North America: http://www.let.leidenuniv.nl/history/migration/chapter52.html

Devine, T. (1979). Temporary Migration and the Scottish Highlands in the Nineteenth Century. *The Economic History Review, New Series , 32* (3), 344-359.

Devine, T. (1988). *The Great Highland Famine: Hunger, Emigration, and the Scottish Highlands in the Nineteenth Century.* Edinburg: John Donald Publishers LTD.

Dupanloup, I., Bertorelle, G., Chikhi, L., & Barbujani, G. (2004). Estimatine the Impact of Prehistoric Admixture on the Genome of Europeans. *Molecular Biology and Evolution , 21* (7), 1361-1372.

Eriksen, T. H., & Nielsen, F. S. (2001). *A History of Anthropology.* Sterling: Pluto Press.

Ewens, W. (1972). The Sampling Theory of Selectively Neutral Alleles. *Theoretical Population Biology , 3*, 87-112.

Excoffier, L. L., & Schneider, S. (2005). Arlequin ver. 3.0: An Integrated Software Package for Population Genetics Data Analysis. *Evolutionary Bioinformatics Online , 1*, 47-50.

Faux, D. (2008). *Shetland Islands Haplogroup R1b*. Retrieved April 2009, from Shetland Islands: DNA - Genealogy Project: http://www.davidkfaux.org/shetlandhaplogroupR1b.html

Finnila, S., Hassinen, I. E., Ala-Kokko, L., & Majamaa, K. (2000). Phylogenetic Network of mtDNA Haplogroup U in Northern Finland Based on Sequence Analysis of the Complete Coding Region by Conformation-Sensitive Gel Electrophoresis. *American Journal of Human Genetics , 66*, 1017-1026.

Fogleman, A. (1992). Migration to the Thirteen British North American Colonies, 1700-1775: New Estimates. *Journal of Interdisciplinary History , 22* (4), 691-709.

Fu, Y. (1997). Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics , 147*, 915-925.

Garza, J., & Williamson, E. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology , 10*, 305-318.

General Register Office for Scotland. (2005). *Scotland's Population 2004: The Registrar General's Annual Review of Demographic Trends: 150th Edition.* Edinburgh: General Register Office for Scotland.

Giles, R., Blanc, H., Cann, H., & Wallace, D. (1980). Maternal Inheritance of Human Mitochondrial DNA. *Proceedings of the National Academy of Science , 77* (11), 6715-6719.

Gonzalez, A. M., Brehm, A., Perez, J. A., Maca-Meyer, N., Flores, C., & Cabrera, V. M. (2003). Mitochondrial DNA Affinities at the Atlantic Fringe of Europe. *American Journal of Physical Anthropology , 120*, 391-404.

Goodacre, S., Helgason, A., Nicholson, J., Southam, L., Ferguson, L., Hickey, E., et al. (2005). Genetic Evidence for a Family-Based Scandinavian Settlement of Shetland and Orkney During the Viking Periods. *Heredity , 95*, 129-135.

Gordon, W. M. (2002). *Mill Girls and Strangers: Single Women's Independent Migration in England, Scotland and the United States, 1850-1881.* Albany: State University of New York Press.

Gray, M. (1955). The Highland Potato Famine of the 1840's. *The Economic History Review , 7* (3), 357-368.

Greenwood, M. J. (2008). Family and Sex-specific U.S. Immigration from Europe 1870-1910: A Panel Data Study of Rates and Composition. *Explorations in Economic History , 45*, 356-382.

Harding, R. M., & Sokal, R. R. (1988). Classification of the European Langueage Families by Genetic Distance. *Proceedings of the National Academy of Science, USA , 85*, 9370-9372.

Harpending, H., Sherry, S. T., Rogers, A. R., & Stoneking, M. (1993). The Genetic Structure of Ancient Human Populations. *Current Anthropology , 34* (4), 483-496.

Hatcher, J. (1996). Plague, Population and the English Economy, 1348-1530. In M. Anderson (Ed.), *British Population History: From the Black Death to the Present Day* (pp. 9-93). Cambridge: Cambridge University Press.

Helgason, A., Hickey, E., Goodacre, S., Bosnes, V., Stefansson, Ward, R., et al. (2001). mtDNA and the Islands of the North Atlantic: Estimating the Proportions of Norse and Gaelic Ancestry. *American Journal of Human Genetics , 68*, 723-737.

Helgason, A., Siguroardottir, S., Gulcher, J. R., Ward, R., & Stefansson, K. (2000). mtDNA and the Origin of the Icelanders: Deciphering Signals of Recent Population History. *American Journal of Human Genetics , 66*, 999-1016.

Helgason, A., Siguroardottir, S., Nicholson, J., Sykes, B., Hill, E. W., Bradley, D. G., et al. (2000). Estimating Scandinavian and Gaelic Ancestry in the Male Settlers of Iceland. *American Journal of Human Genetics , 67*, 697-717.

Houston, R. (1996). The Population History of Britain and Ireland 1500-1750. In M. Anderson (Ed.), *British Population History: From the Black Death to the Present Day* (pp. 95-190). Cambridge: Cambridge University Press.

Hunter, J., & Ralston, I. (1999). *The Archaeology of Britain: An Introduction from the Upper Paleolithic to the Industrial Revolution.* Routledge.

Jobling, M., Hurles, M., & Tyler-Smith, C. (2004). *Human Evolutionary Genetics: Origins, Peoples & Disease.* New York: Garland Science.

Johnson, N. P., & Mueller, J. (2002). Updating the Accounts: Global Mortality of the 1918-1920 "Spanish" Influenza Pandemic. *Bulletin of the History of Medicine , 76*, 105-115.

Jordan, W. C. (1996). *The Great Famine: Northern Europe in the Early Fourteenth Century.* New Jersey: Princeton Unitversity Press.

Jorde, L., Watkins, W., Bamshad, M., Dixon, M., Ricker, C., Seielstad, M., et al. (2000). The Distribution of Human Genetic Diversity: A Comparison of Mitochondrial, Autosomal, and Y-Chromosome Data. *American Journal of Human Genetics , 66*, 979-988.

Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., et al. (1997). Evaluation of Y-Chromosomal STRs: a Multicenter Study. *International Journal of Legal Medicine , 110*, 125-133.

Kearney, H. (2006). *The British Isles: A History of Four Nations.* New York: Cambridge University Press.

Kimura, M. (1969). The Number of Heterozygous Nucleotide Sites Maintained in a Finite Population Due to a Steady Flux of Mutations. *Genetics , 61*, 893-903.

Kimura, M., & Crow, J. (1964). The Number of Alleles that can be Maintained in a Finite Population. *Genetics , 49*, 725-38.

King, T. E., Ballereau, S. J., Schurer, K. E., & Jobling, M. A. (2006). Genetic Signatures of Coancestry Within Surnames. *Current Biology , 16*, 384-388.

Kivislid, T., Tolk, H., Parik, J., Wang, Y., Papiha, S., Bandelt, H., et al. (2002). The emerging limbs and twigs of the East Asian mtDNA tree. *Molecular Biology and Evolution , 19* (10), 173701752.

Kruskal, J. (1964). Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika , 29*, 1-27.

Landsman, N. C. (1999). Nation, Migration, and the Province in the First British Empire: Scotland and the Americas, 1600-1800. *The American Historical Review , 104* (2), 463-475.

Levi, J. M., & Dean, B. (2003). Introduction. In B. Dean, & J. M. Levi (Eds.), *At the Risk of Being Heard* (pp. 1-44). Ann Arbor: The University of Michigan Press.

Loogvali, E., Roostalu, U., Malyarchuk, B., Derenko, M., Kivisild, T., Metspalu, E., et al. (2004). Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Molecular Biology and Evolution , 21* (11), 2012-2021.

MacKenzie, J. M. (1998). Empire and National Identities The Case of Scotland. *Transactions of the Royal Historical Society , 8* (6), 215-231.

Macklin, M. G., Bonsall, C., Davies, F. M., & Robinson, M. R. (2000). Human-environment Interactions During the Holocene: New Data and Interpretations from the Oban Area, Argyll, Scotland. *The Holocene , 10*, 109-121.

MacLean, J. (2009). Chapter IV. Darien Scheme. In P. MacLean, *An Historical Account of the Settlements of the Scotch Highlanders in America prior to the Peace of 1783; Together with notices of Highland Regiments and Biographical Sketches* (pp. 60-74). Glasgow: BiblioBazaar, LLC.

Magnusson, M. (2000). *Scotland: The Story of a Nation.* USA: Harper Collins.

Manly, B. F. (1994). *Multivariate Statistical Methods.* New York: Chapmann & Hall/CRC.

Manly, B. (2005). *Multivariate Statistical Methods: A Primer.* New York: Chapman & Hall/CRC.

McDonald, F., & McDonald, E. S. (1980). The Ethnic Origins of the American People, 1790. *The William and Mary Quarterly , 37* (2), 181-199.

McEvoy, B., Richards, M., Forster, P., & Bradley, D. G. (2004). The Longue Duree of Genetic Ancestry: Multiple Genetic Marker Systems and Celtic Origins on the Atlantic Facade of Europe. *American Journal of Human Genetics , 75*, 693-702.

McGovern, T. H. (1990). The Archaeology of the Norse North Atlantic. *Annual Review of Anthropology , 19*, 331-351.

Mellars, P. (2006). Archaeology and the Dispersal of odern Humans in Europe: Deconstructing the "Aurignacian". *Evolutionary Anthropology , 15*, 167-182.

Mellars, P. (2004). Neanderthals and the Modern Human Colonization of Europe. *Nature , 432*, 461-465.

Mielke, J. H., & Fix, A. G. (2006). The Confluence of Anthropological Genetics and Anthropological Demography. In M. H. Crawford (Ed.), *Anthropological Genetics: Theory, Methods and Applications* (pp. 112-140). Cambridge, New York: Cambridge University Press.

Moffat, A. (2005). The Ice Domes. In *Before Scotland: The Story of Scotland Before History* (pp. 23-50). New York: Thames & Hudson.

Mogentale-Profizi, N., Chollet, L., Stevanovitch, A., Dubut, V., Poggi, C., Pradie, M., et al. (2001). Mitochondrial DNA Sequence Diversity in Two Groups of Italian Veneto Speakers from Veneto. *Annals of Human Genetics , 65*, 153-166.

Moore, L. T., McEvoy, B., Cape, E., Simms, K., & Bradley, D. G. (2006). A Y-Chromosome Signature of Hegemony in Gaelic Ireland. *The American Journal of Human Genetics , 78*, 334-338.

Murdoch, A. (1998). Emigration from the Scottish Highlands to America in the Eighteenth Century. *British Journal for Eighteenth-Century Studies , 21*, 161-74.

Nei, M. (1987). *Molecular Evolutionary Genetics.* New York: Columbia University Press.

Nei, M., & Kumar, S. (2000). *Molecular Evolution and Phylogenetics.* New York: Oxford University Press.

Newby, A. G. (2007). Ireland, Radicalism and the Scottish Highlands, c. 1870-1912. *Scottish Historical Review Monographs Series No. 15.* Edinburgh: Edinburgh University Press Ltd.

Novelletto, A. (2007). Y Chromosome Variation in Europe: Continental and Local Processes in the Formation of Extant Gene Pool. *Annals of Human Biology , 34* (2), 139-172.

Ohta, T., & Kimura, M. (1973). A Model of Mutation Appropriate to Estimate the Number of Electrophoretically Detectable Molecules in a Finite Population. *Genetic Research , 22*, 201-204.

Ohta, T., & M, K. (1973). A Model of Mutation Approriate to Estimate the Number of Electrophoretically Detectabe Alleles in a Finite Population. *Genetic Research , 22*, 201-204.

Oppenheimer, S. (2006). *The Origins of the British A Genetic Detective Story.* London: Constable.

Pakendorf, B., & Stoneking, M. (2005). Mitochondrial DNA and Human Evolution. *Annual Review of Genomics and Human Genetics , 6*, 165-183.

Pereira, L., Dupanloup, I., Rosser, Z. H., Jobling, M. A., & Barbujani, G. (2001). Y-Chromosome Mismatch Distributions in Europe. *Molecular Biology and Evolution , 18* (7), 1259-1271.

Poser, C. M. (1994). The Dissemination of Mutiple Sclerosis: A Viking Saga? A Historical Essay. *Annals of Neurology Supplement 2 , 36*, S231-S243.

Purvis, T. L. (1984). The European Ancestry of the United States Population, 1790: A Symposium. *The William and Mary Quarterly , 41* (1), 85-101.

Redd, A., & Stoneking, M. (1999). Peopling of Sahul: mtDNA Variation in Aboriginal Australian and Papau New Guinean Populations. *American Journal of Human Genetics , 65*, 808-828.

Redd, A., Clifford, S., & Stoneking, M. (1997). Multiplex DNA typing of short-tandem-repeat loci on the Y chromosome. *Biological Chemistry , 8*, 923-927.

Reidla, M., Kivisild, T., Metspalu, E., Kaldma, K., Tambets, K., Tolk, H.-V., et al. (2003). Origin and Diffusion of mtDNA Haplogroup X. *American Journal of Human Genetics , 73* (5), 1178-1190.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., et al. (2000). Tracing European Founder Lineages in the Near Eastern mtDNA Pool. *American Journal of Human Genetics , 67*, 1251-1276.

Roewer, L., & Epplen, J. (1992). Rapid and Sensitive Typing of Forensic Stains by PCR Amplification of Polymorphic Simple Repeat Sequences in Casework. *Forensic Science International , 53*, 163-171.

Roewer, L., Croucher, P. J., Willuweit, S., Lu, T. T., Kayser, M., Lessig, R., et al. (2005). Signature of Recent Historical Events in the European Y-Chromosomal STR Haplotype Distribution. *Human Genetics , 116*, 279-291.

Rogers, A., & Harpending, H. (1992). Population Growth Makes Waves in the Distrubution of Pairwise Genetic Differences. *Molecular Biology and Evolution , 9*, 552-569.

Roostalu, U., Kutuev, I., Loogvali, E., Metspalu, E., Tambets, K., Reidla, M., et al. (2007). Origin and Expansion of Haplogroup H, the Dominant Human Mitochondrial DNA Lineage in West Eurasia: the Near Eastern and Caucasian Perspective. *Molecular Biology and Evolution , 24* (2), 436-448.

Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., Help, H., Bermisheva, M., et al. (2004). Phylogeography of Y-Chromosome Haplogroup I Reveals Distinct Domains of Prehistoric Gene Flow in Europe. *American Journal of Human Genetics , 75*, 128-137.

Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., Alvantic, D., Amorim, A., et al. (2000). Y-Chromosomal Diversity in Europe is Clinal and Influenced Primarily by Geography, Rather than by Language. *American Journal of Human Genetics , 67*, 1526-1543.

Rubicz, R. C., Melton, P. E., & Crawford, M. H. (2006). Molecular Markers in Anthropological Genetic Studies. In M. H. Crawford (Ed.), *Anthropological Genetics* (pp. 141-186). Cambridge, New York: Cabridge University Press.

Rubicz, R., Melton, P., & Crawford, M. (2006). Chapter 7: Molecular markers in anthropological genetic studies. In M. Crawford (Ed.), *Anthropological Genetics: Theory, Methods, and Applications.* Cambridge University Press.

Schanfield, M. (2006). Applications of Molecular Genetics to Forensic Scienes. In M. H. Crawford (Ed.), *Anthropological Genetics* (pp. 235-276). Cambridge, New York: Cambridge University Press.

Schultting, R. J., & Richards, M. P. (2002). Finding the Coastal Mesolithic in Southwest Britain: AMS Dates and Stable Isotope Results on Human Remains from Caldy Island, South Wales. *Antiquity , 76* (294), 1011-1025.

Scottish Government. (2004, June). *Networking? It Runs in the Blood of the Celts*. Retrieved November 17, 2009, from Scotland: the Official Online Gateway: http://www.scotland.org/about/history-tradition-and-roots/features/culture/clan.html

Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V., 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. *American Journal of Human Genetics , 74*, 1023-1034.

Semino, O., Passarino, G., Brega, A., Fellous, M., & Santachiara-Benerecetti, A. S. (1996). A View of the Neolithic Demic Diffusion in Europe through Two Y Chromosome-Specific Markers. *Americcan Journal of Human Genetics , 59* (4), 964-968.

Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., et al. (2000). The Genetic Legacy of Paleolithic Homo sapiens sapiens in Extant Europeans: A Y Chromosome Perspective. *Science , 290*, 1155-1159.

Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J., & Barbrbujani, G. (2000). Geographic Patterns of mtDNA Diversity in Europe. *American Journal of Human Genetics , 66*, 262-278.

Slatkin, M. (1995). A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. *Genetics , 139*, 457-462.

Stoneking, M. (2000). Hypervariable Sites in the mtDNA Control Region are Mutational Hotspots. *American Journal of Human Genetics , 67*, 1029-1032.

Stringer, C. (2006). *Homo Britannicus: The Incredible Story of Human Life in Britain.* London: Penguin Group.

Sturrock, K., & Rocha, J. (2000). Multidimensional Scaling Stress Evaluation Table. *Field Methods , 12* (1), 49-60.

SWGDAM. (2004). Report on the Current Activities of Scientific Working Group on DNA Analysis Methods Y-STR Subcommittee. *Forensic Science Communications , 6* (3).

Sykes, B. (2006). *Blood of the Isles.* Toronto: Bantam Press.

Sykes, B. (2006). Chapter 12: The DNA of Scotland. In *Blood of the Isles* (pp. 186-218). Toronto: Bantam Press.

Sykes, B. (2006). Scotland. In *Blood of the Isles* (pp. 165-185). Toronto: Bantam Press.

Sykes, B. (2006). The Picts. In *Blood of the Isles* (pp. 177-185). Toronto: Bantam Press.

Symonds, J. (1999). Toiling in the Vale of Tears: Everyday Life and Resistance in South Uist, Outer Hebrides, 1760-1860. *International Journal of Historical Archaeology , 3* (2), 101-122.

Tajima, F. (1993). Measurement of DNA polymorphism. In N. Takahata, & A. Clark (Eds.), *Mechanisms of Molecular Evolution. Introduction to Molecular Paleopopulation Biology* (pp. 37-59). Tokyp: Japan Scientific Societies Press, Sinauer Associates, Inc.

Tajima, F. (1993). Simple Methods for Testing the Molecular Evolutionary Clock Hypothesis. *Genetics , 134* (2), 599.

Tamura, K., & Nei, M. (1993). Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees. *Molecular Biology and Evolution , 10*, 512-526.

Tamura, K., Dudley, J., Nei, M., & S, K. (2007). MEGA 4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution , 24*, 1596-1599.

Thompson, J., DG, H., & TJ, G. (1994). CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Research , 22*, 4673-4680.

Torroni, A., Bandelt, H.-J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., et al. (1998). mtDNA Analysis Reveals a Major Late Paleolithic Population Expansion

from Southwestern to Northeastern Europe. *American Journal of Human Genetics , 62*, 1137-1152.

Torroni, A., Bandelt, H.-J., Macaulay, V., Richards, M., Cruciani, F., Rengo, C., et al. (2001). A Signal, From Human mtDNA, of Postglacial Recolonization in Europe. *American Journal of Human Genetics , 69* (4), 844-852.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., et al. (1996). Classification of European mtDNAs From an Analysis of Three European Populations. *Genetics , 144* (4), 1835-50.

Torroni, A., Lott, M. T., Cabell, M. F., Chen, Y.-S., Lavergne, L., & Wallace, D. C. (1994). MtDNA and the Origin of Caucasians: Identification of Ancient Caucasian-specific Haplogroups, One of Which Is Prone to a Recurrent Somatic Dupliction in the D-Loop Region. *American Journal of Human Genetics , 55*, 760-776.

Torroni, A., Richards, M., Macaulay, V., Forster, P., Villems, R., Norby, S., et al. (2000). MtDNA Haplogroups and Frequency Patterns in Europe. *American Journal of Human Genetics , 66* (3), 1173-1177.

Tyfiled, L. A., Stephenson, A., F, C., Harvie, A., Bidwell, J. L., Wood, N. A., et al. (1997). Sequence Variation at the Phenylalanine Hydroxylase Gene in the British Isles. *American Journal of Human Genetics , 60*, 388-396.

Waddington, C., Bailey, G., Bayliss, A., Boomer, I., Milner, N., Pedersen, K., et al. (2003). *A Mesolithic settlement site at Howick, Northumberland: a preliminary report.* White Rose Consortium ePrints Repository. United Kingdom: Society of Antiquaries of Newcastle-upon-Tyne.

Watterson, G. (1975). On the Number of Segregation Sites. *Theoretical Population Biology , 7*, 256-276.

Weale, M. E., Weiss, D. A., Jager, R. F., Bradman, N., & Thomas, M. G. (2002). Y Chromosome Evidence for Anglo-Saxon Mass Migration. *Molecular Biology and Evolution , 19* (7), 1008-1021.

Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., et al. (2001). The Eurasian Heartland: A Continental Perspective on Y-Chromosome Diversity. *Proceedings of the National Academy of Science , 98* (18), 10244-10249.

Willis, K., & Whittaker, R. (2000). The Refugial Debate. *Science , 287* (5457), 1406-1407.

Winter, J. (1977). Britain's 'Lost Generation' of the First World War. *Population Studies , 31* (3), 449-466.

Withers, C. W. (1988). Destitution and Migration: Labour Mobility and Relief from Famine in Highland Scotland 1836-1850. *Journal of Historical Geography , 14* (2), 128-150.

Young, L. (1996). Spaces for Famine: A Comparative Geographical Analysis of Famine in Ireland and the Highlands in the 1840s. *Transactions of the Institute of British Geographers , 21* (4), 666-680.

Zlojutro, M. (2008). Mitochondrial DNA and Y-chromosome Variation of Eastern Aleut Populations: Implications for the Genetic Structure and Peopling of the Aleutian Archipelago. 221 pages. University of Kansas.

Appendices

Appendix A: Scottish Clan Distributions

Appendix A shows the distribution of clans identified by participants in this study for: a) individuals examined for mtDNA markers, and b) individuals examined for Y-chromosome markers. Image was adapted from Wikipedia.org.Clan distributions were obtained from Wikipedia.org at http://upload. wikimedia.org /wikipedia/commons/7/78/Scottish_clan_map.png. Information was also obtained for clans not listed on the above map from the following urls:

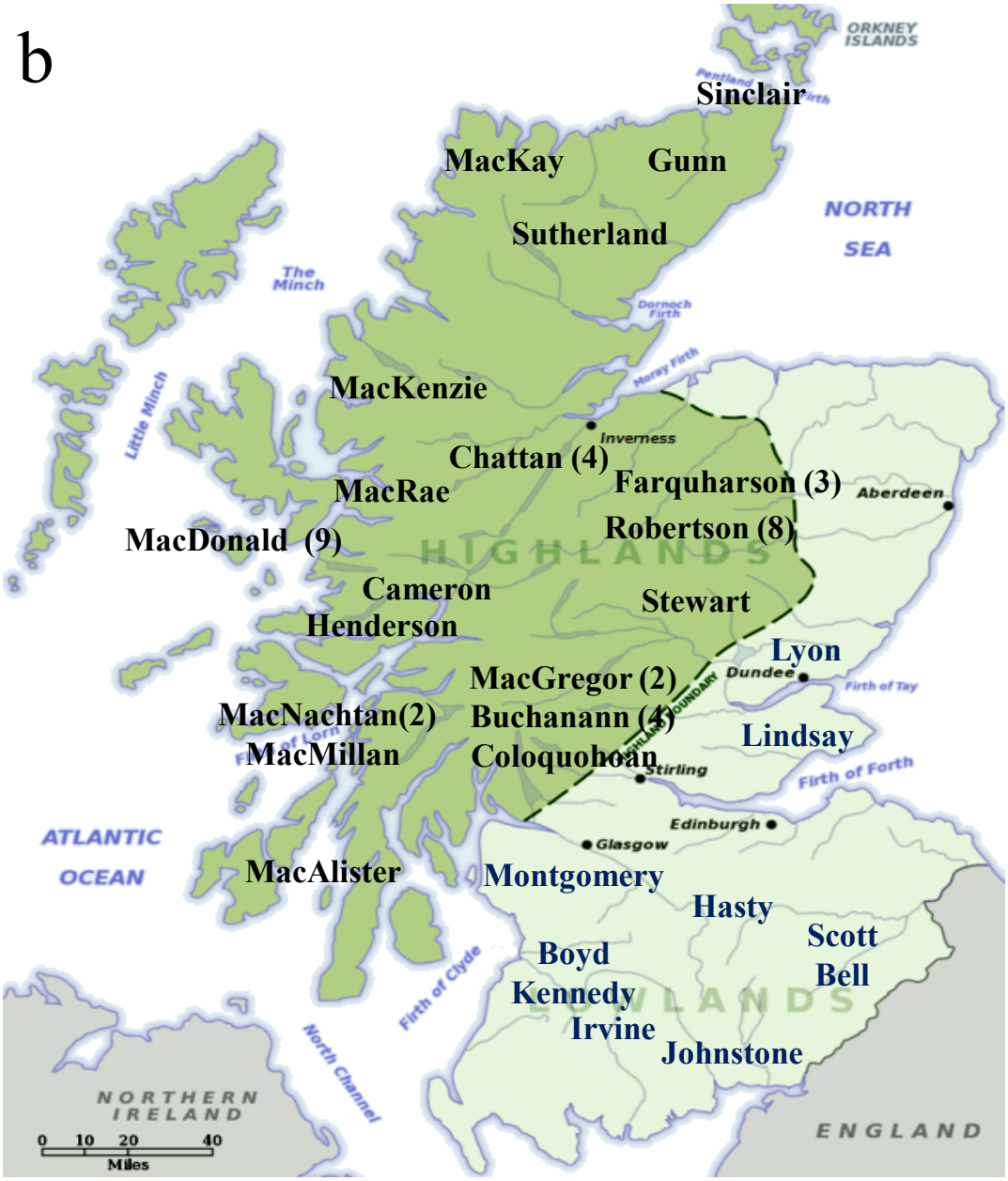
1. www.clanmacmillan.org : Clan MacMillan
2. www.electricscotland.com : Clans Bell, Fleming, MacTavish, and Lyon
3. www.scotweb.co.uk : Clan Irvine
4. www.clanmacnicol.org : Clan MacNicol
5. www.houseofnames.com : Clans Sellars and Hasty
6. www.clanboyle.org : Clan Boyle
7. www.clanboyd.info : Clan Boyd

Some clans, such as Clan MacDonald are distributed throughout Scotland, but are only labeled in one region. Clan names that are in green boxes are Scottish Highland Clans, while the Blue boxes indicate a Lowland Clan.

a


ORKNEY ISLANDS
Sinclair
MacKay
Gunn
NORTH SEA
Sutherland
The Minch
Dornoch Firth
Ross
Rose
Moray Firth
MacKenzie
MacIntosh
MacNicol
Little Minch
Fraser (3)
Inverness
Gordon
Chattan (5)
Farquharson (3)
MacRae
Aberdeen
Robertson (8)
Sellars
HIGHLANDS
MacDonald (14)
Cameron
Henderson
Stewart
Campbell (3)
Oliphant
MacAlpine
Lyon
MacFarlane
MacGregor (2)
Dundee
Firth of Tay
MacNachtan(2)
Buchanann (4)
Lindsay
Firth of Lorn
Boyle
Coloquohoan
Stirling
Fleming
Firth of Forth
MacMillan
Livingstone
Edinburgh
ATLANTIC OCEAN
MacTavish (3)
Glasgow
Montgomery
Cunningham
Hasty
Scott (4)
MacAlister
Boyd
Bell
Firth of Clyde
Kennedy
LOWLANDS
Irvine
North Channel
Johnstone
NORTHERN IRELAND
ENGLAND
0 10 20 40
Miles

b


ORKNEY ISLANDS
Sinclair
MacKay
Gunn
NORTH SEA
Sutherland
The Minch
Dornoch Firth
Moray Firth
MacKenzie
Little Minch
Inverness
Chattan (4)
Farquharson (3)
Aberdeen
MacRae
Robertson (8)
MacDonald (9)
HIGHLANDS
Cameron
Stewart
Henderson
Lyon
Dundee
Firth of Tay
MacGregor (2)
BOUNDARY
MacNachtan(2)
Buchanann (4)
Lindsay
MacMillan
Coloquohoan
Stirling
Firth of Forth
Edinburgh
Glasgow
ATLANTIC OCEAN
MacAlister
Montgomery
Hasty
Scott
Boyd
Bell
Kennedy
Firth of Clyde
LOWLANDS
Irvine
Johnstone
North Channel
NORTHERN IRELAND
ENGLAND
0 10 20 40
Miles

Appendix B: Y-STR Allele Frequencies

Y-STR allele frequencies for all European populations.

DYS19	KCGames	Scotland	Germany	Nether	Sweden	Belgium	Poland	Norway	Italy	Denmark	England	France	Ireland
10	-	-	-	-	-	-	-	-	0.00149	-	-	-	-
11	-	-	0.00029	-	0.00706	-	-	-	0.00149	-	-	-	-
12	0.02381	-	0.00174	-	-	-	-	-	0.00373	-	-	0.04808	-
13	0.07143	0.02530	0.06450	0.03273	0.06497	0.11200	0.03963	0.02667	0.14787	0.01587	0.03644	0.09615	0.01869
14	0.54762	0.73524	0.50349	0.72364	0.58898	0.62400	0.19512	0.52667	0.51307	0.69841	0.70445	0.54808	0.83178
15	0.30952	0.16189	0.22777	0.18182	0.23588	0.19200	0.21418	0.31333	0.24496	0.23810	0.18623	0.21154	0.12150
16	0.04762	0.06577	0.13771	0.04000	0.08757	0.05600	0.33003	0.12000	0.06423	0.03175	0.05668	0.06731	0.02804
17	-	0.01180	0.06363	0.01818	0.01554	0.01600	0.21875	0.01333	0.02315	0.01587	0.01215	0.02885	-
18	-	-	0.00058	0.00364	-	-	0.00229	-	-	-	0.00405	-	-
19	-	-	0.00029	-	-	-	-	-	-	-	-	-	-

DYS390	KCGames	Scotland	Germany	Nether	Sweden	Belgium	Poland	Norway	Italy	Denmark	England	France	Ireland
17	-	-	0.00029	-	-	-	-	-	-	-	-	-	-
18	-	-	0.00029	-	-	-	-	-	-	-	-	-	-
19	-	-	-	-	-	-	0.00076	-	-	-	-	0.00481	-
20	-	-	0.00145	-	0.00141	-	-	-	0.00075	-	0.00405	-	0.00935
21	-	0.00084	0.01482	0.00727	0.00565	-	0.00457	0.00333	0.02091	0.01587	0.00405	0.05288	-
22	0.175	0.05143	0.14643	0.18545	0.19915	0.128	0.05488	0.16	0.14712	0.22222	0.11741	0.10577	0.05607
23	0.225	0.22091	0.26119	0.32	0.37571	0.272	0.13643	0.31667	0.29574	0.39683	0.2753	0.23558	0.25234
24	0.325	0.4941	0.34311	0.36	0.26977	0.488	0.29497	0.29667	0.42644	0.26984	0.43725	0.47115	0.43925
25	0.275	0.21669	0.21296	0.10182	0.13418	0.08	0.46723	0.22	0.09933	0.09524	0.1417	0.11058	0.21495
26	-	0.01518	0.01772	0.02182	0.01271	0.032	0.0404	0.00333	0.00896	-	0.02024	0.01923	0.02804
27	-	0.00084	0.00174	0.00364	0.00141	-	0.00076	-	0.00075	-	-	-	-

DYS391	KCGames	Scotland	Germany	Nether	Sweden	Belgium	Poland	Norway	Italy	Denmark	England	France	Ireland
8	-	-	0.00058	-	0.00141	-	-	-	-	-	-	0.00481	-
9	-	0.00506	0.01801	0.02182	0.01695	0.032	0.00915	0.00667	0.06871	-	0.00405	0.05288	-
10	0.4186	0.47892	0.58425	0.51273	0.60593	0.52	0.58232	0.54667	0.56983	0.50794	0.51012	0.42308	0.36449
11	0.51163	0.49663	0.37885	0.43636	0.36017	0.448	0.40168	0.43667	0.34727	0.46032	0.47368	0.50481	0.61682
12	0.06977	0.01855	0.01656	0.02909	0.01554	-	0.00686	0.00667	0.01419	0.03175	0.0081	0.01442	0.01869
13	-	0.00084	0.00145	-	-	-	-	0.00333	-	-	0.00405	-	-
14	-	-	0.00029	-	-	-	-	-	-	-	-	-	-

DYS392	KCGames	Scotland	Germany	Nether	Sweden	Belgium	Poland	Norway	Italy	Denmark	England	France	Ireland
6	-	-	0.00029	-	-	-	-	-	-	-	-	-	-
7	-	-	0.00087	-	-	-	-	-	-	-	-	-	-
8	-	-	0.00058	-	-	-	-	-	-	-	-	-	-
9	-	-	0.00029	-	-	-	-	-	0.00075	-	0.00405	-	-
10	0.02326	0.00084	0.00494	-	0.00424	-	0.00305	-	0.00672	-	0.00405	-	0.00935
11	0.27907	0.23356	0.52963	0.34182	0.57627	0.352	0.7843	0.63333	0.48544	0.38095	0.22672	0.37019	0.11215
12	0.04651	0.05312	0.0584	0.08	0.05367	0.032	0.0404	0.05	0.10978	0.07937	0.06478	0.11058	0.05607
13	0.53488	0.59781	0.35706	0.53455	0.23305	0.576	0.13262	0.24667	0.35026	0.49206	0.61943	0.48077	0.6729
14	0.11628	0.10877	0.0398	0.04	0.12853	0.024	0.0343	0.07	0.03585	0.04762	0.07287	0.03365	0.1215
15	-	0.0059	0.00784	0.00364	0.00424	0.016	0.00381	-	0.00747	-	0.0081	0.00481	0.02804
16	-	-	0.00029	-	-	-	0.00152	-	0.00373	-	-	-	-

DYS393	KCGames	Scotland	Germany	Nether	Sweden	Belgium	Poland	Norway	Italy	Denmark	England	France	Ireland
9	-	0.00084	-	-	-	-	-	-	-	-	-	-	-
10	-	-	0.00029	-	-	-	0.00076	0.00333	0.00149	-	-	-	-
11	-	0.00084	0.00465	-	0.00141	-	-	-	0.01344	-	0.0081	-	-
12	0.12195	0.04216	0.10633	0.07273	0.0678	0.104	0.08155	0.04333	0.24496	-	0.04049	0.13462	0.03738
13	0.73171	0.87437	0.74346	0.79273	0.75424	0.752	0.80716	0.81333	0.59149	0.88889	0.80972	0.72115	0.84112
14	0.14634	0.05818	0.12522	0.11636	0.16102	0.112	0.10442	0.13333	0.13294	0.06349	0.09717	0.13462	0.1028
15	-	0.02108	0.01859	0.01818	0.01554	0.032	0.00534	0.00667	0.01568	0.04762	0.04453	0.00481	0.01869
16	-	0.00169	0.00116	-	-	-	0.00076	-	-	-	-	0.00481	-
17	-	0.00084	0.00029	-	-	-	-	-	-	-	-	-	-

Appendix C: Summary Tables

Summary tables of female and male participants, their clan affiliations, place of birth (POB), haplogroups (Hg) and haplotypes (Ht), and region where that haplotype is found: Scotland mainland (Sc), Northwest Coast (NW), Western Isles and Isle of Skye (WS), Orkney (Or), Sheltand (Sh). For male particpants, haplotype location and surname distributions from the year 1881, with shading indicating their prevalence in each region are also included (http://www.nationaltrustnames.org.uk/).

Female Participants		**Mitochondrial DNA**		**Shared ht with**			
Clann Affiliation	**POB**	**Hg**	**Ht (+16000)**	**Sc**	**NW**	**WS**	**Or**
not reported	PA	H	293				
Bell, MacDonald, Oliphant	NE	K	224, 311	Y	Y	Y	Y
Buchan, Robertson	WA	H	319	Y	Y		
Campbell	MO	H	183, 189, 225, 278				
Campbell	MO	V	298	Y			
Cunningham	IL	H	140, 274, 356				
Farquharsons, Buchanan	MO	H	60, 304	Y			Y
Fleming	MO	H	293,362				
Fraser	MO	W	223, 278, 292				
Fraser	VT	V	298	Y			
Gordon, Scott	KS	T	126, 294, 311				
Lindsay, Campbell	KS	H	CRS	Y	Y	Y	Y
Livingstone	MO	H	192, 239,311				
MacDonald	WV	J	126, 294, 296, 304	Y	Y		
MacDonald	KS	H	239	Y			
MacDonald	KS	H	147, 356				
MacKintosh	FL	V	261, 298, 311	Y	Y		
MacRae	OH	T	126, 153, 242, 294, 320				
MacRae, Irvine	IL	J	69, 126, 311		Y		
MacRae, MacNicol	OK	H	75				
MacTavish	NC	J	69, 126	Y	Y	Y	Y
MacTavish	KS	J	60, 69, 126, 265, 319				
MacTavish	MO	V	298	Y		Y	
Robertson	NM	H	304, 311				
Ross	RI	J	126				
Scott	KS	H	110, 162, 209				
Scott, Fraser	KS	U	75, 162				
Sellars*	AL	H	93, 304				

*Clan is not recognized by the Court of Lord Lyon

Male Participants			Mitochondrial DNA	Shared ht with				Y Chromosome		Shared ht/ surn with				
Clann Affiliation	POB	Hg	Ht	Sc	NW	WS	Or	Yhg	YSTR haplotype**	Sc	Or	WS	Sh	NW
not reported	TN	H	176					R1b	14, 11, 15, 13(10), 29(26), 23, 11, 13, 13, 12, 11	Y	Y	Y	Y	Y
not reported	MO	H	CRS	Y	Y	Y	Y	R1b	14, 11, 17, 12(9), 29(26), 24, 11, 13, 13, 12, 12	Y	Y	Y	Y	Y
not reported	WI	H	30, 362	Y	Y			R1b	14, 11, 14, 13(10), 29(26), 24, 10, 13, 13, 12, 11	Y	Y	Y	Y	Y
Boyd	MS	K	224, 311	Y	Y	Y	Y	R1b	15, 11, 15, 12(9), 28(25), 24, 11, 13, 13, 12, 12	Y	Y	Y	Y	Y
Boyle, Rose	KS	V	79, 298					no						
Buchan & Robertson	MO	H	319	Y	Y			R1b/R1a	15, - , 13, 13(10), 29(26), - , 11, 11, - , 12, 11	Y	Y	Y	Y	Y
Buchan & Robertson	MO	H	319	Y	Y			R1b	15, 10, 17, 12(9), 28(25), 23, 11, 13, 13, 12, 12	Y		Y	Y	
Cameron	MO	H	266, 320					R1b	15, 11, - , 13(10), 29(26), 23, 10, 13, 13, 12, 12	Y				
Chattan	KS	Unk	CRS	Y	Y	Y	Y	no						
Chattan	MO	Unk	CRS	Y	Y	Y	Y	G2a	15, 10, 14, 12(9), 30(27), 22, 10, 10, 14, 10, 11	Y	Y		Y	Y
Chattan	MO	Unk	CRS	Y	Y	Y	Y	G2a	15, - , 14, 12(9), 30(27), 22, 10, 11, 14, 10, 11	Y	Y		Y	Y
Chattan	MO	H	235, 291, 293	Y	Y	Y		G2a	15, 10, 14, 12(9), 30(27), 22, 10, 11, 14, 10, 11	Y	Y		Y	Y
Chattan	KS	U	93, 256, 270, 291					R1b	14, 11, 14, 13(10), 29(26), 24, 10, 13, 13, 12, 11	Y	Y	Y	Y	Y
Colquohoan	MO	H	129, 223, 391	Y	Y	Y	Y	R1b	- , 11, 13, - , - , - , 10, - , 13, - , 12					
Farquharson	MO	J	59, 126	Y	Y	Y	Y	R1b	14, 11, 14, 13(10), 29(26), 23, 11, 14, 13, 12, 11	Y			Y	Y
Farquharson	MO	J	69, 126	Y	Y	Y	Y	R1b	13, 11, 14, 13(10), 29(26), 23, 11, 14, 13, 12, 12					
Farquharson	MN	U	126, 294, 296, 304	Y	Y	Y		no						
Farquharson & Lyon	CA	H	368					I1	16, - , 14, 12(9), 28(25), 22, 10, 11, 13, 10, 11			Y		
Gunn & MacDonald	IL	H	64, 184, 189, 356					I1/I2b	15, - , - , 12(9), 28(25), 24, 10, 11, 13, 10 , -				Y	
Hasty	RI	H	CRS	Y	Y	Y	Y	R1b	14, - , - , 13(10), 29(26), 25, 11, 14, 13, 12, 12	Y		Y	Y	Y
Henderson	KS	H	93	Y	Y	Y		R1b	- , 11, - , 12(9), 28(25), 25, 10, 13, 13, 12, 13	Y	Y	Y	Y	Y
Irvine	IL	U	93, 192, 256, 270, 291					I1	14, - , - , 13(10), 30(27), 22, - , 11, 13, 10, 12	Y	Y	Y	Y	Y
Johnstone	MO	H	CRS	Y	Y	Y	Y	R1b	14, 12, 14, 14(11), 28(25), 25, 11, 13, 13, 12, 12	Y	Y	Y	Y	Y
Kennedy	CA	H	CRS	Y	Y	Y	Y	R1b	15, 11, 14, 13(10), 28(25), 23, 11, 13, 13, 12, 11	Y		Y	Y	
Lindsay	nr	no						R1b	14, 11, 14, 13(10), 30(27), 25, 10, 13, 13, 13, 11	Y	Y	Y	Y	Y

Surname: Least → Most

**DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439

Male Participants			Mitochondrial DNA	Shared ht with				Y Chromosome		Shared ht/surn with					
Clann Affiliation	POB	Hg	Ht	Sc	NW	WS	Or	Yhg	YSTR haplotype**	Sc	Or	WS	Sh	NW	Surname
MacAlister	OH	V	223, 240, 298					R1b	14, 11, 14, 14(11), 30(27), 25, 11, 12, 13, 12, 12			Y	Y		Least
MacAlpine	KS	U	192, 256, 270	Y	Y			no							
MacDonald	MO	H	CRS	Y	Y	Y	Y	R1b	13, 11, 14, 12(9), 28(25), 24, 11, 13, 14, 12 ,12						
MacDonald	MO	H	CRS	Y	Y	Y	Y	no							
MacDonald	IL	H	287	Y				G2a	16, - , - , 12(9), 29(26), 22, 10, 11, 14, 10, 11						
MacDonald	WA	J	75, 116, 126, 311					R1b	14, 11, 14, 13(10), 30(27), 24, 11, 13, 12, 11, 12	Y	Y		Y		Most
MacDonald	MO	K	224, 311	Y	Y	Y	Y	no							
MacDonald	OK	H	289							Y		Y	Y		
MacFarlane	MO	K	93, 224, 311, 319	Y				no							
MacKenzie	KS	H	189, 368,377					J2b	15, 13, 16, 12(9), 28(25), 24, 10, 11, 12, 11, 11	Y					
MacNachtan	Sc	U	75, 104, 189, 256, 270					no							
MacNachtan	MO	H	319		Y			E1b1b	13, - , - , 13(10), 30(27), 24, 10, 11, 13, 10, 12	Y				Y	
MacRae	KS	U	104, 192, 256, 270, 294					no							
MacRae & MacDonald	MO	J	69, 126, 145, 172, 223, 261	Y	Y			R1b	- , 11, 14, - , - , 24, 10, - , - , - , 11						
MacRae & MacDonald	OH	T	126, 153, 242, 294, 320					R1b	14, 11, 14, 13(10), 29(26), 25, 11, 13, 13, 12, 11	Y	Y	Y	Y	Y	
MacRae & MacDonald	MO	H	148, 293					R1b	14, 11, 13, 13(10), 30(27), 24, 10, 13, 13, 12, 12	Y	Y	Y	Y	Y	
MacRae & MacDonald	MO	no						R1b	14, 11, 14, 14(11), 30(27), 25, 11, 13, 13, 12, 11	Y	Y	Y	Y	Y	
MacRae, MacMillan, MacKay	MO	U	51, 129, 189, 194, 197, 362, 377					R1b	14, - , - , 13(10), 29(26), - , - , 13, - , 12, 11						
Montgomery	MO	H	189, 356, 375					R1b	15, 11, 14, 13(10), 29(26), 23, 12, 13, 13, 12, 12						
Robertson	MO	J	126, 294, 296, 304	Y	Y			I1	14, 11, 13, 12(9), 29(26), 22, 10, 11, 13, 10, 11	Y	Y	Y	Y	Y	
Robertson	MO	U	51, 129, 182, 183, 189, 233, 362					R1b	14, 10, 13, 13(10), 29(26), 25, 11, 14, 13, 12, 12	Y		Y	Y	Y	
Robertson	MO	H	CRS	Y	Y	Y	Y	R1b	14, 11, 13, 13(10), 29(26), 25, 11, 14, 13, 12, 12	Y		Y	Y	Y	
Robertson	KS	U	189, 270, 319, 392					R1b	14, 11, 15, - , 29(26), 23, 11, 13, 13, 12, 13	Y	Y	Y	Y	Y	
Robertson, MacGregor, Gunn	OK	H	CRS	Y	Y	Y	Y	R1b	14, 11, 15, 14(11), 33(30), 24, 11, 13, 12, 12, 11	Y	Y		Y		
Robertson, MacGregor, Gunn	MN	U	189, 270					R1b	14, 11, 14, 13(10), 29(26), - , 12, 13, 14, 12, 12						
Scott	OK	T	126, 163, 186, 189, 294	Y	Y		Y	R1b	15, 11, 15, 13(10), 29(26), 25, 11, 13, 13, 12, 12	Y		Y	Y		
Sinclair	MO	H	134, 356	Y	Y			R1b	14, 11, 13, 13(10), 29(26), 24, 11, 13, 13, 12, 12	Y	Y	Y	Y	Y	
Stewart	MO	H	176					R1a	15, 11, 14, 13(10), 33(30), 25, 10, 11, 13, 11, 10	Y	Y	Y	Y		
Sutherland	SC	no						R1b	14, 11, 14, 13(10), 30(27), 24, 11, 13, 12, 12, 13	Y	Y		Y		

**DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439