# Germanic Origins from the Perspective of the Y-Chromosome

By

Michael Robert St. Clair

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor in Philosophy
in
German
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Irmengard Rauch, Chair
Thomas F. Shannon
Montgomery Slatkin

Spring 2012

# Abstract

# Germanic Origins from the Perspective of the Y-Chromosome

by

Michael Robert St. Clair

Doctor of Philosophy in German

University of California, Berkeley

Irmengard Rauch, Chair

This dissertation holds that genetic data are a useful tool for evaluating contemporary models of Germanic origins. The Germanic languages are a branch of the Indo-European language family and include among their major contemporary representatives English, German, Dutch, Danish, Swedish, Norwegian and Icelandic. Historically, the search for Germanic origins has sought to determine where the Germanic languages evolved, and why the Germanic languages are similar to and different from other European languages. Both archaeological and linguist approaches have been employed in this research direction. The linguistic approach to Germanic origins is split among those who favor the *Stammbaum* theory and those favoring language contact theory. *Stammbaum* theory posits that Proto-Germanic separated from an ancestral Indo-European parent language. This theoretical approach accounts for similarities between Germanic and other Indo-European languages by posting a period of mutual development. Germanic innovations, on the other hand, occurred in isolation after separation from the parent language. Language contact theory posits that Proto-Germanic was the product of language convergence and this convergence explains features that Germanic shares with other Indo-European languages. Germanic innovations, on the other hand, are potentially a relic of an era before language convergence.

Contemporary models of Germanic origins have gravitated towards language contact theory to explain the position of Germanic within the European linguistic tapestry. However, this theoretical approach is very dependent on the historical record for assessing the influence of language convergence. This dissertation utilizes genetic data, primarily single nucleotide polymorphism from the human Y-chromosome, for overcoming this inherent weakness of language contact theory. With genetic data, the linguist can now assess the influence of prehistoric language convergence by tracing prehistoric population expansions. Based on the available genetic data, the evolution of Germanic during the European prehistory may have been shaped by the convergence of Proto-Basque, Proto-Indo-European, Proto-Afroasiatic, and perhaps to a lesser extent, Proto-Uralic.

# DEDICATION

Dedicated to Dr. David Rood, a great man and Professor of Linguistics at the University of Colorado at Boulder.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgements

Thanks to those who have provided grants and fellowships: the Max Kade Foundation, the Institute of European Studies (University of California at Berkeley), the Department of German (University of California at Berkeley), and the Graduate Division (University of California at Berkeley). Thanks to those who have provided employment: the Department of German (University of California at Berkeley), the University Library (University of California at Berkeley), and the Department of History (University of California at Berkeley. I wish to acknowledge the members of my dissertation committee: Professors Irmengard Rauch, Thomas Shannon and Montgomery Slatkin. A special thanks to James Spohrer, Curator of the Germanic collection. This project would not have been possible without the support of Alexandra, my wife. Finally, a special thanks to the members of my church.

# Chapter One
# Dissertation Overview

My task in this dissertation is to demonstrate that genetic data are a useful tool for evaluating contemporary models of Germanic origins. To the best of my knowledge, this dissertation represents one of the first attempts by a linguist to utilize genetic data for exploring how the Germanic language family evolved during the European prehistory. This approach to examining Germanic origins reflects that the use of genetic data for such a purpose has only been available within the last decade. Most of the genetic data provided in this dissertation stem from population reports that describe world-wide mutational variation found on the non-recombining region of the human Y-chromosome. These reports started to appear in the year 2000 following technological advances at the end of the last century that facilitated the identification of molecular markers.

This dissertation is divided into seven chapters. Chapter Two provides an overview of the traditional approaches to examining the origins of Germanic languages: linguistics and archaeology. The same chapter also provides three reasons why the use of Y-chromosome may have been underutilized. First, the methodology behind population genetics requires further explication so that the non-geneticist can evaluate the usefulness of this tool. Secondly, the nomenclature used to Y-chromosome variation has been subject to standardization and revision since 2000. Lastly, the source of Y-chromosome data is extremely fragmented, appearing in over 200 population reports published in approximately forty different scientific journals.

Chapter Three explicates the methodology for interpreting Y-chromosome data. The non-recombining region of the Y-chromosome provides a means of tracing prehistoric migration and settlement. Chapter Four overcomes the problems of inconsistent and revised nomenclature by lumping the Y-chromosome data into ten population expansions during the European pre-history. Chapter Five provides a survey of population studies that ultimately demonstrate the usefulness of genetic data for the linguist. In Chapter Six, four contemporary models of Germanic are evaluated utilizing the data from Chapters Four and Five. The dissertation conclusion, found in Chapter Seven, stresses that contemporary models of Germanic origins are clearly more receptive to language contact theory. As such, genetic data has become a useful tool for evaluating these models as they are able to overcome an inherit weakness of this theoretical approach to language variation. With genetic data, the linguist now has a much clearer picture of prehistoric language convergence that may have shaped the evolution of Early Germanic.

# Chapter Two
# Germanic Origins: Issues and Approaches

## 2.0 Chapter Introduction.

Perhaps the troubling aspect of any examination of language origins is that it seems so speculative, that the researcher lacks the security of attested language change as well as the historical record. Nevertheless, researchers still continue to use the resources at hand to render their best guess as to how Germanic languages may have evolved in the prehistory. In this chapter, I will discuss the traditional tools for examining Germanic origins, linguistics and archaeology. I will then introduce a new source of data for exploring this topic, population genetics. This chapter also presents a section explaining how the search for the origins of Germanic languages has become a controversial topic.

## 2.1 Typology.

Linguists classify the Germanic languages as a sub-group of the Indo-European language family. The figure below provides an overview of the Indo-European language family. Around 600 Indo-European languages are spoken. Because of space constraints,

**Figure 2.1 Indo-European.**

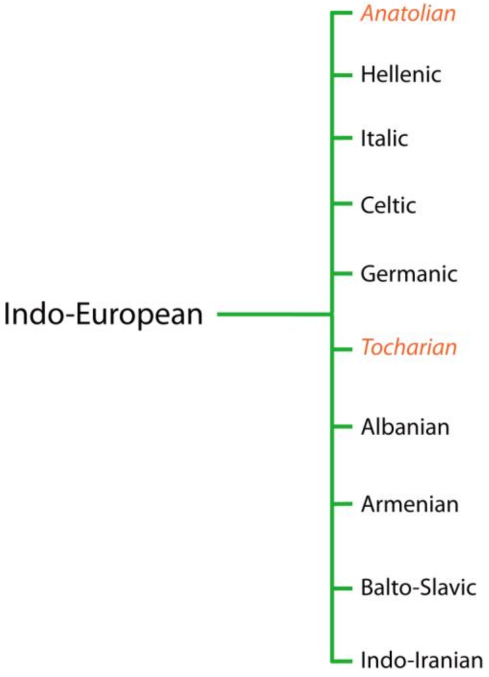


Figure 2.1 focuses on the evolution of the ten main branches of the Indo-European language family. Two of the ten daughter languages, Anatolian and Tocharian, appearing in orange italics, are now extinct. Hellenic (Greek), Armenian, and Albanian are single language branches of the Indo-European language family. Celtic includes Irish, Welsh and Breton. The Italic languages include Latin, as well as the Romance languages, Spanish, Portuguese, French, Italian and Romanian. The Balto-Slavic branch separates into two sub-branches, Baltic and Slavic. The Baltic branch consists of Latvian and Lithuanian. Slavic languages include Russian, Ukrainian, Polish, Czech, and Bulgarian. The Indo-Iranian branch includes Persian and Hindi, languages spoken in Asia. (See also Beekes 1995: 11-33 for a more thorough discussion of Indo-European languages.)

Several languages fall under the classification of "Germanic," including Dutch, Frisian, German, Norwegian, Danish, Icelandic, Swedish, and English, the language of this dissertation. Figure 2.2 (below) provides an overview of the Germanic branch of the Indo-European language family, reflecting the evolution of Germanic languages from an ancestral Proto-Germanic language. Proto-Germanic diverged into three sub-branches, North, West and East Germanic. Modern English and Modern Standard German are West Germanic languages.

The term "Proto-Germanic" describes the first Germanic language, which arose in prehistoric times in Northern Europe. Proto-Germanic is not attested in the historical record, but rather based on linguistic reconstruction. Early Germanic, on the other hand, was first attested 2,000 years ago with short inscriptions found on combs, brooches, drinking horns, spearheads, sword scabbards, shield bosses and other personal items. These inscriptions mostly conveyed a personal name or perhaps some sort of magical incantation using, as a character source, the runic alphabet, also known as futhark. These letters may have been borrowed from an alphabet used for a non-Germanic language spoken in northern Italy (Todd 1992: 120). Although very controversial, some date the oldest Early Germanic inscription to the first century BC, and more specifically the following runic words found on a helmet discovered in Negau on the Austrian/Slovenian border: *harigasti teiwa* (e.g. Schultze 1986: 329). According to Waterman, scholars differ on the meaning of this inscription; he suggests "to the god Harigast" (1976: 21). Another example of Early Germanic is the Golden Horn of Gallehus, which an example of one the few runic inscriptions not conveying a message rooted in magic or religion. This drinking horn, produced in the fourth century, contains the following inscription: *ek hlewagastir holtigar horna tawido*. This inscription means "I Hlewagast of the Holting clan made this horn" (Waterman 1976: 22). Part of the significance of this inscription is that it provides a glimpse of Early Germanic syntax and morphology.

**Figure 2.2 Germanic.**

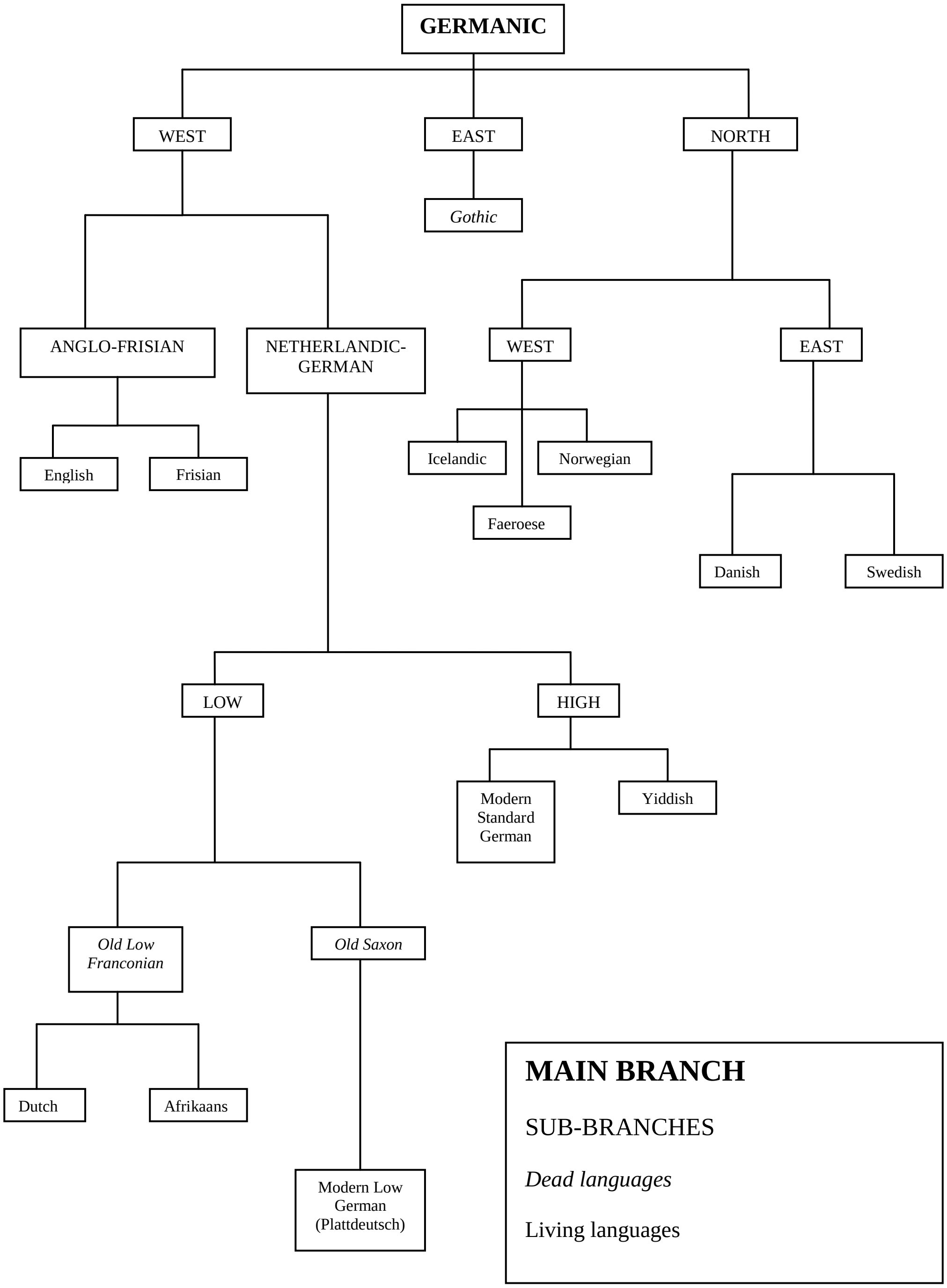


**Chart adapted from Pyles and Algio 1993: 68-69.**

The most significant attestation of Early Germanic is the Codex Argenteus, a manuscript copy of the bible written in Gothic, the language of the Goths, one of the Germanic tribe. Around 100 BC the Goths left the Germanic homeland and by 300 AD they had migrated to the edge of the Black Sea in modern-day Romania. In 341, Wulfila became Bishop of the Goths. In order to convert his people to Christianity he translated the bible into his native language. Since Wulfila was the first to write in Gothic, he had to create an alphabet, possibly adapting letters from the Runic, Greek and Latin (e.g. Rauch 2003: 4-5). While the original manuscript has disappeared, in the sixteenth century a manuscript copy of Wulfila's bible, the Codex Argenteus was found in the Abby at Werden in Germany. The document appears quite striking, elaborately ornamented in gold and silver letters on purple vellum. During the Thirty Years War, the codex was taken to Sweden, where it remains today at the University of Uppsala. Other copies of Wulfila's bible are found elsewhere; the Codex Carolinus in Wolfenbüttel, Codex Ambrosiani and Codex Turinensis in Turin, and Codex Gissensis in Egypt (Henriksen and van der Auwera 1994: 2). Meanwhile, the Gothic language has become extinct, perhaps last spoken somewhere in the Ottoman Empire in the sixteenth century (Rauch 2003: 12). Nevertheless, its importance continues in that linguists consider the language to be the best representation of Proto-Germanic in comparative grammar.

## 2.2 The Origins of Indo-European Languages.

Since Germanic languages are classified as Indo-European languages, the traditional starting point for exploring the prehistoric development of Germanic languages is to examine the origins of Indo-European languages. Linguists believe Indo-European languages are not indigenous to the European continent, but rather these languages came to Europe from Western Asia sometime during the prehistory. Today, two competing models of this language expansion have surfaced in the literature. Both models attempt to isolate the putative homeland of the speakers of the Proto-Indo-European languages and explain when they expanded into Europe. Marija Gimbutas, a Lithuanian archaeologist, proposed the Kurgan conquest model of Indo-European origins in a series of articles she wrote over a forty year period, ultimately compiled in *The Kurgan Culture and the Indo-Europeanization of Europe: Selected Articles from 1952-1993*. Her theory is often cited in linguistic texts as offering a plausible explanation of how Proto-Indo-European spread throughout Europe. Trask (1996: 360), for example, writes that while he does not find the Kurgan theory totally persuasive, "it is still the best solution we have and it refuses to go away." Gimbutas wrote her final article about the Kurgan conquest in 1993, which was published in 1997. This article, "The Fall and Transformation of Old Europe: Recapitulation 1993," reports that the Kurgan culture emerged somewhere in the Volga basin between 5000 and 4500 BC. An identifying trademark of this culture is a unique mortuary practice; they buried their dead in pits, which were then covered with a mound of dirt. In her final article, Gimbutas maintains (1997a: 354) that the Kurgans rode horses and raised herd animals within a patriarchal society. Around 4500 BC the Kurgans became more aggressive and began migrating to the west. In the area to the west, what Gimbutas often calls “Old Europe,” lived a “Goddess worshiping” culture, whose focus was “the perpetual functioning of the cycle of life, death and regeneration embodied by a central feminine force.” (351). Gimbutas asserts (358) that this culture could not resist the Kurgan invasion of warriors from the east who rode horses and who were better armed. During the conquest of Old Europe, the Kurgans imposed their language, Proto-Indo-European, upon the indigenous Europeans (364).

In her final discussion of the Kurgans, Gimbutas relies in part upon archaeological remains, primarily burial customs and pottery. She also reconstructs an Indo-European culture based on the comparative method and comparative mythology. The comparative method (cf. Section 2.3.1 below**)** examines grammatical similarities among attested Indo-European languages and attempts to reconstruct an original Indo-European form using plausible linguistic explanations. Gimbutas believes that an Indo-European lexicon can be reconstructed, which becomes the foundation of reconstructing elements of an Indo-European culture, such as social structure and economy. Comparative mythology uses a similar methodology and examines ideological similarities among attested Indo-European-speaking cultures, which is then used to reconstruct an Indo-European ideology. Other scholars also cite ideological changes in Europe as additional evidence of a Kurgan invasion (e.g. Fortson 2010: 18-49; Anthony 2007a: 463-466; Beekes 1995: 25-52).

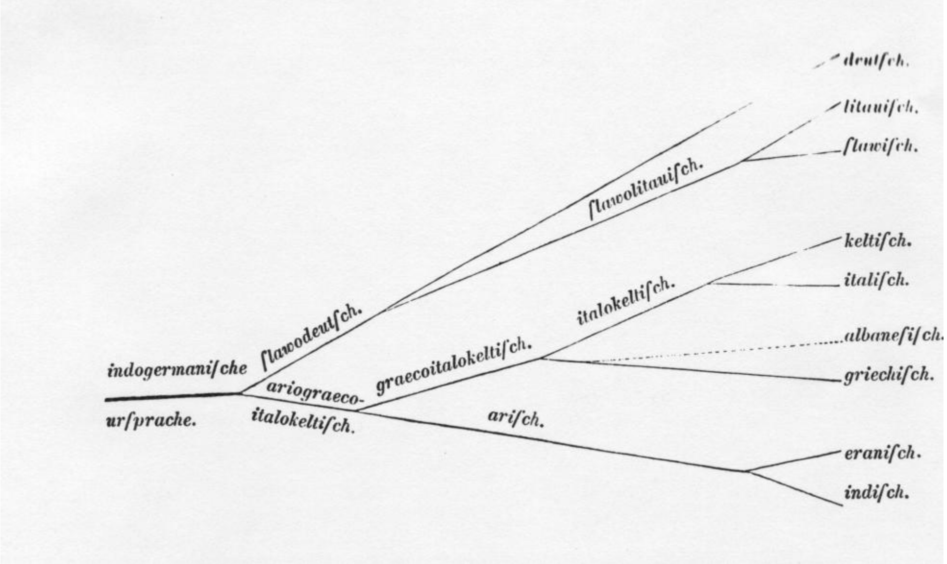
The alternative model of Proto-Indo-European origins, the language-farming model, was proposed by Colin Renfrew, a British archaeologist. In order to understand his model it is necessary to briefly discuss the origins of agriculture in Europe. Following the end of the last Ice Age, about twelve thousand years ago, agriculture arose independently in seven different areas of the world: Mexico, North America, South America, New Guinea, Sub-Saharan Africa, East Asia, and southwest Asia. The origins of agriculture in southwest Asia are found in what are now Turkey, Syria and Iraq. Starting about 10000 BC, people in this area began to cultivate a variety of crops, including cereals such as wheat, barley and rye, and pulses such as chickpeas and lentils. They also started to domesticate animals such as sheep, goats, cattle, and pigs. By 6500 BC, farming had spread from southwest Asia to Greece. Within four thousand years this new technology was adopted throughout most of the European continent. From Greece, the spread of farming followed two different routes. One route took farming technology along the northern Mediterranean coast, into Italy and finally the Iberian Peninsula. The other route introduced farming in the Balkans, central Europe, and eventually Scandinavia and Britain. (cf. Scarre 2005a: 176-199 and Scarre 2005b: 392-431 for a more thorough discussion of the spread of farming in Europe).

Renfrew published his model of Indo-European origins in 1987, in his book *Archaeology and Language: the Puzzle of Indo-European Origins*. He proposed (145-177) that Indo-European languages were introduced onto the European continent by farmers who migrated from present-day Anatolia into Europe during the Neolithic era. He asserts that over the course of two thousand years, the descendants of these farmers migrated into central and western Europe in search of suitable farmland. While the number of farmers involved in the initial migration to Greece was relatively small, their Indo-European speaking descendants were able to colonize southern and central Europe and displace the non-Indo-European speaking hunter-gathers, whose descendants had previously migrated into Europe during the Paleolithic, perhaps 30,000 years prior.

## 2.3 Germanic Origins from the Perspective of Linguistics.

The linguist Theo Vennemann (2000: 234-236) describes a traditional and an alternative approach for examining the origins of Germanic languages. The traditional approach considers Germanic as a further development of Proto-Indo-European. The alternative model views the origins of Germanic as the product of language contact between two or more languages. The traditional approach follows *Stammbaum* theory or "family tree" theory, proposed by the German linguist August Schleicher (1821-1868) in his two volume *Compendium der vergleichenden Grammatik der Indogermanischen Sprachen*, published in 1861 and 1862. In his *Compendium*, Schleicher provides a taxonomic diagram that posits a chain of events leading to the evolution of German, Lithuanian, Slavic, Celtic, Italic, Albanian, Greek, Iranian and Indic from a common Indo-European language. An image of Schleicher's *Stammbaum*, or 'family tree' is provided in Figure 2.3 below.

**Figure 2.3 Schleicher's *Stammbaum*.**



**Source: Schleicher 1861: 7.**

In the first volume of the *Compendium* (1861) Schleicher cites numerous phonological similarities found in German, Lithuanian, Slavic, Celtic, Italic, Albanian, Greek, Iranian and Indic to support his model. In the second volume, Schleicher addresses morphological similarities. Based partly on evolutionary theory, the basic idea behind the *Stammbaum* model is that similarities between languages are accounted for by a common ancestral language and a period of common development, whereas differences occurred after diverging from the ancestral language.

Language contact theory accounts for similarities between language as a product of borrowing or admixture, and differences as a lack of contact. The theoretical foundation of language contact theory stems from Johannes Schmidt and his 1872 book *Die Verwantschaftsverhältnisse der Indogermanischen Sprachen*. In this book, Schmidt voices his disagreement with *Stammbaum* Theory and its proponent, August Schleicher. Relying on words and word roots, Schmidt disputes *Stammbaum* Theory, arguing that similarities in Indo-European languages are not explained by divergence from a single ancestral proto-Indo-European language. He argues that similarities arise from innovations that eventually spread to neighboring languages. Schmidt proposed a wave model to explain similarities between Indo-European languages. This model posits that innovations spread due to religious, political, social or other reasons. Nevertheless, the intensity of an innovation loses it strength over distance, similar to the ripple effect when a stone it tossed onto to a lake (27-28). Thus if A, B, C, D, E, F, and G represent a continuum of languages over distance, an innovation in language C might extend to languages B and D, but not languages A and E.

In my opinion *Stammbaum* theory and language contact theory are not necessarily contradictory models of language change, but rather may present complementary approaches to language change and variation. I personally view *Stammbaum* theory as a process that could be labeled “divergence” and language contact theory as “convergence.” I would like to see both divergence and convergence brought under a single theoretical umbrella to explain language variation and change. Documented language change would probably reflect that both convergence and divergence drive language variation, and that the influence of each model varies over time within a given area. For example, I suspect that in Germany the number of German dialects increased steadily from the Middle Ages until the Napoleonic Wars as a result of increased political and religious fragmentation, which conforms to a divergence model of change. Following the Napoleonic Wars, I suspect dialect variation steadily decreased as the result of nationalism, according to a convergence model. Moreover, modern mass-media has probably accelerated the process of convergence.

### 2.3.1 Comparative Method and Stammbaum Theory.

Clackson’s 2007 book, *Indo-European Linguistics*, presents, in my opinion, an easy-to-read contemporary treatment of the *Stammbaum* model across all the Indo-European languages. This book presents arguments for this model using phonological reconstructions, as well as morphological, syntactic and lexical reconstructions. The term “reconstruction” points to a part of the grammar that has never been recorded. Thus, standard practice requires the linguist to use an asterisk to show that a form is reconstructed. To help the non-linguist understand the methodology behind linguistic reconstruction, I will turn to phonological reconstruction and use examples from an introductory linguistics book. This methodology is commonly known as the "comparative method." Reconstructed phonology starts by assembling cognate sets among attested languages. The term cognate reflects the idea that the meaning assigned to a group of sounds is completely arbitrary. Consequently, if several languages assign the same meaning to the same sounds, then these languages potentially have a common ancestral language. Consider the paradigm below in the table below.

**Table 2.1 Germanic Cognates and Turkish.**

| **English** | **German** | **Dutch** | **Swedish** | **Turkish** |
|---|---|---|---|---|
| man | Mann | man | man | adam |
| hand | Hand | hand | hand | el |
| foot | Fuß | voet | fot | ajak |
| bring | bringen | brengen | bringa | getir |
| summer | Sommer | zomer | sommar | jaz |

**Source: Murray 2001: 324.**

A linguistic interpretation of the data in Table 2.1 would posit that 'man', 'hand', 'foot', 'bring', and 'summer' are cognates in English, German, Dutch, and Swedish, whereas Turkish fails to produce a cognate for these words. Taking this a step further, English, German, Dutch, and Swedish may have diverged from a common ancestor, whereas the ancestral language for Turkish and English is different or far more distant in the past.

Once cognate sets have been assembled, the next step in phonological reconstruction is to seek systematic correspondences. Consider Table 2.2 below.

**Table 2.2 'Female Friend' in Spanish, Portuguese, Sardinian and French.**

| | **Spanish** | **Portuguese** | **Sardinian** | **French** |
|---|---|---|---|---|
| **Orthographic**: | amiga | amiga | amica | amie |
| **Phonetic**: | [amiγa] | [amiga] | [amika] | [ami] |

**Adapted from Murray 2001: 326.**

Here the linguist could posit that the striking similarities found in Spanish, Portuguese, Sardinian and French for 'female friend' resulted from divergence from a common ancestral language and the reconstructed Proto-Romance word *amika. Based on plausible phonological rules, the *k in Proto-Romance *amika underwent voicing and frication in Spanish, voicing in Portuguese, and deletion in French, whereas Sardinian retained the voiceless stop (Murray 2001: 326-329). The historical record indeed shows that French, Sardinian, Portuguese, and Spanish are a product of Roman conquest. As one would expect, the etymology of the Spanish word *amiga* finds its origins in Latin, and more specifically the word *amīca* (Diego 1985: 53). However, Proto-Indo-European and Proto-Germanic are not attested in the historical record, and thus reconstruction remains the primary means of determining how these languages may have appeared.

The above discussion of linguistic reconstruction is a simplified example of what actually takes years of specialized training. Additionally, even with the comparative method, sometimes the decision as to whether a language is Indo-European may rest with linguistic intuition and consensus with most linguists (Clackson 2007: 3). Nevertheless, the non-linguist has a rough idea of the methodology used to support a *Stammbaum* model of Germanic origins, an approach still discussed in the literature (e.g. Ringe 2006).

The strength of the comparative method for reconstructing Proto-Indo-European and Proto-Germanic is found in attested Germanic languages. Throughout the history of Germanic, the evolution of sound change has been regular and systematic affecting entire classes of sounds, not just a few words. For example, in the historical evolution of Old High German to Middle High German, back vowels became fronted due to phonological phenomena known umlaut (Waterman 1976: 85-86). In the evolution from Middle High German to Early New High German, short vowels in open syllables underwent lengthening (c.f. Waterman 1976: 102-103). Nevertheless, the *Stammbaum* theory as model of Indo-European origins has its critics. For example, Lyle Campbell (2004: 164-166) lists three objections. First, this approach assumes a uniform Proto-European language over a vast geographical expanse, from Western Europe to India, without dialectal variation. Secondly, this approach assumes that ten daughter languages, including Proto-Germanic, simultaneously diverged from Proto-Indo-European. Finally, this approach assumes that after the daughter languages diverged, the speakers of these languages had no further contact with each other. In my opinion Campbell's argument suggests that *Stammbaum* theory defies the attested behavior of languages.

### 2.3.2 Language Contact Theory.

The most recognized proponent of modern language contact theory is probably the linguist Sarah G. Thomason. She (2001:1) defines language contact as "the use of more than language in the same place at the same time." One of the most appealing features of her approach to language contact theory is that the historical record indeed contains numerous examples of language contact induced change. Thomason (2001: 10) emphasizes that "language contact is the norm, not the exception. We would have a right to be astonished if we found any language whose speakers had successfully avoided contacts with all other languages for periods longer than one or two hundred years." According to Donald Winford (2003: 305-308), modern language contact theory presents three possible scenarios or outcomes that may occur when a language comes in contact with another language. The first scenario posits language maintenance, that a language remains largely intact with the exception of lexical borrowing. The second scenario posits that language contact may induce language shift, which states that within a geographic area a new language replaces another. In the third scenario, language contact may result in the creation of a new language, such as a pidgin or creole. The term pidgin can be explained as a hybrid language that sometimes emerges through commerce and trade among speakers of different languages. Here, the speech community remains rather small. The definition of a creole is somewhat controversial. For the sake of brevity, I would define creole as a pidgin with a much larger speech community. Creoles also tend to be passed from one generation to the next, whereas pidgins serve a temporary need and then disappear.

Interestingly, all the three language contact scenarios posited by Winford are attested in Germanic languages. Modern Standard German provides an example of language maintenance. Standard German has borrowed words from Latin, French and English, yet remains grammatically distinct in other areas, such as verb second word order and the strong/weak adjective distinction. The language history of Ireland supports language shift, the second scenario, where English has almost completely replaced Irish Gaelic, a Celtic language. Finally, even the formation of pidgins and creoles is attested in Germanic. Russenorsk provides an example of a pidgin that arose between Norwegian and Russian sailors in the nineteenth century (Broch 1927). One example of a creole is Negerhollands, a hybrid language that arose in the Virgin Islands in the eighteenth century, a combination primarily of Dutch and African languages (Rossem and Voort 1996).

One controversial aspect of language contact theory involves what is borrowable. The adoption of new words from other languages, i.e. lexical borrowings, is well attested. Modern English provides one of the best examples. Technically a Germanic language, English has borrowed heavily from French, Latin and Greek, and to a lesser extent, from Scandinavian, Celtic, Spanish, and Italian (Pyles and Algeo 1993: 286-311). However, linguists dispute the degree to which borrowing could affect the structure of language (e.g. Winford 2003: 61-63). Nevertheless, Thomason (2001: 63) takes the position that any linguistic feature of a language can be borrowed by another language. She responds to her critics by writing, “various claims can be found in the literature to the effect that this or that is un-borrowable, but counterexamples can be found (and have been found) to all of the claims that have been made to date.” An extreme example of structural change induced by language contact is found in Asia-Minor Greek, an Indo-European language. As the result of long-term heavy contact with speakers of Turkish, the morphology and syntax of Asia Minor Greek changed considerably. For example, common Greek has subject-verb-object word order. However, in certain circumstances Asia Minor Greek adopted the subject-object-verb word order found in Turkish. Perhaps more striking is that some dialects of Asia Minor Greek borrowed Turkish agglutinative morphology (Thomason and Kaufman 1988: 215-222).

One potentially huge disadvantage of language contact theory is its heavy reliance on the historical record. *Stammbaum* Theory (cf. Section 2.3.1**)**, as the reader may recall, utilizes the comparative method and linguistic reconstruction when the historical record has disappeared. Nevertheless, I would argue that attested Germanic still leaves traces of language contact induced change occurring before recorded history. One possible area that presents evidence of such change in prehistoric Germanic is the lexicon or vocabulary of Modern German. Scholars of Germanic linguistics (e.g. Vennemann 2000: 241; Waterman 1976:36) estimate that a third of the Modern German lexicon lacks an Indo-European cognate. The authority in defining the non-Indo-European and purely Germanic component of the German lexicon is Schirmer and Mitzka (1969). Uniquely Germanic words fall within three categories: cereal production and animal husbandry, seafaring and fishing, and legal terminology. Schirmer and Mitzka (46-50) give the following examples of Modern German words having a Germanic origin: *Brot* 'bread', *Schaf* 'sheep', *Hafen* 'harbor', *Dorsch* 'cod', *Volk* 'people' and *König* 'king.' The strong verbs of modern standard German may also leave traces of prehistoric language contact induced change. In his examination of Germanic strong verbs, Robert Mailhammer concluded that "46.5% of all Germanic strong verbs do not have an accepted Indo-European etymology" (2007: 167-187). Mailhammer suggests (2007: 175) that convergence of a non-Indo-European language with Proto-Indo-European may explain the identifying feature that strong verb ablaut is for the Germanic languages. Moreover, Mailhammer (2007: 199) suggests that language contact between speakers of Punic (a Semitic language spoken by the Carthaginians) and Germanic may explain the unique role played by ablaut in Germanic strong verbs. He asserts that the systemization and function of ablaut found in Germanic strong verbs are typologically more similar to that found in Semitic languages than that in Indo-European languages. His views prompt arguments for further research in Germanic languages from the perspective of language contact theory.

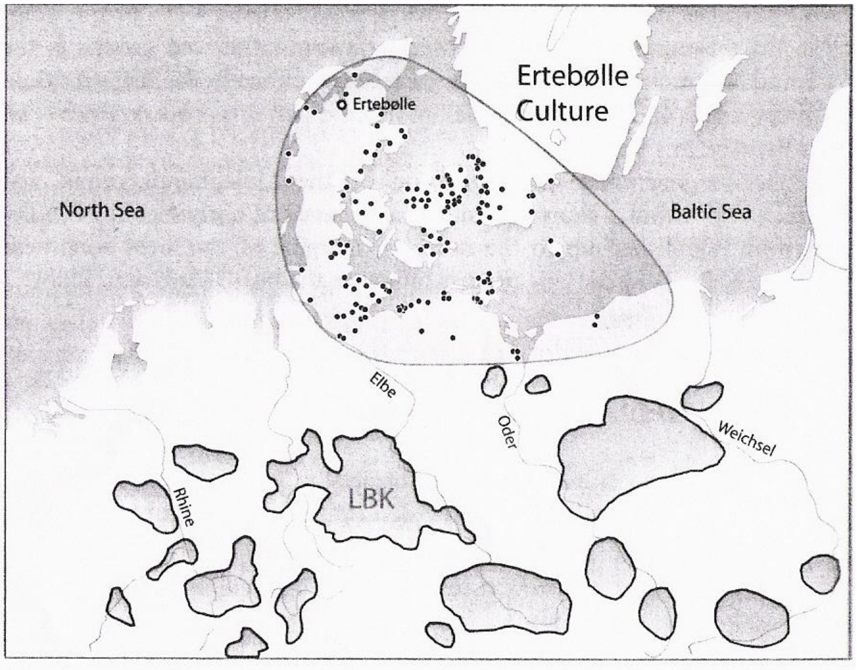
### 2.4 Germanic Origins from the Perspective of Archaeology.

In English, the word "Germanic" sounds very similar to the word "German." Modern German, however, makes a distinction using the terms *der Germane* and *der Deutsche*. Thus, reader should not confuse the terms *Germanic tribes* or *peoples* with modern-day *Germans* for both terms are ethnically distinct. In my opinion, a convenient starting point for the emergence of the Germans as an ethnic group is the ninth century and the division of Charlemagne's kingdom among his grandsons. The Germanic tribes, on the other hand, appeared in the historical record around 113 BC near the current Austrian/Italian border, where the Cimbri and Teutones, two Germanic tribes, fought a battle against the Romans. According to Seyer (1988: 37), the Romans did not initially distinguish the Germanic tribes from the Celts. Rather, the Romans made this distinction about 50 years later. According to Seyer (41), the word *Germanic* first appeared in the work of the Roman historian Poseidonious to describe the "border neighbors of the Celts on the right bank of the Rhine." After making their grand entrance onto the historical record, the Germanic tribes fought largely amongst themselves, and often with the Romans, for territory and plunder. The story of the Germanic tribes eventually ended with the Saxon Wars at the end of the eighth century, when Charlemagne defeated a Saxon uprising led by Widukind.

Stepping away from recorded history, this dissertation now explores if the Germanic tribes had a prehistory. Since this covers a period before recorded history, the traditional approach to answering such a question is archaeology and the artifacts left behind by prehistoric people. Interestingly food waste, known as kitchen middens, may present the archaeological starting point for researching the prehistoric Germanic peoples. The kitchen middens in Denmark were left behind by the Mesolithic Ertebølle culture. By 5400 BC, the Ertebølle culture had already constructed permanent settlements along the coastline of northern Germany, Denmark and southern Scandinavia (e.g. Hartz 2007: 573). The construction of permanent settlements represents atypical Mesolithic behavior, providing evidence that the Mesolithic inhabitants of the Germanic homeland had indeed found a unique survival strategy by exploiting locally available marine resources. Hunter-gatherer cultures generally do not build permanent settlements, but rather this cultural behavior is Neolithic and associated with the rise of agriculture. My research in Demark revealed that mussels are one factor in understanding how the Ertebølle culture found a unique survival strategy. The kitchen waste left by this culture consists mostly of mussel shells. The Ertebølle culture had access to an abundant supply of mussels, which are highly nutritious. Even today along the shore of the Limfjord, a fjord adjacent to some of the Ertebølle settlements, an abundance of mussels is still found. An abundant supply of locally supplied marine resources meant that the Ertebølle culture could remain in one location. The same food source may have also increased fertility and had corresponding effect on population density.

In 1849, Danish scientists established a commission to examine the kitchen middens left behind by the prehistoric inhabitants of Denmark. Kristian Kristiansen (2002: 21-22) defines this commission as the beginning of modern archaeology in Denmark. Perhaps even more significant, Kristiansen (12-13) asserts that archaeology in Denmark was the beginning of archaeology as an international and independent discipline. Danish archaeology, in turn, influenced the work of Gustaf Kossinna (1858-1931), the father of German anthropology. In his 1896 article, "Die vorgeschichtliche Ausbreitung der Germanen in Deutschland," Kossinna (1) cited the work of Jens Jacob Asmussen Worsaae (1821-1885), one of the founders of Danish archaeology and a key player in the kitchen middens commission. Kossinna (1) used Worsaae's research, along with his own research, and that of the other archaeologists of his time, to dispute linguistic interpretations that placed the homeland of Germanic peoples in Asia. Rather than Asia, Kossinna (14) identified Northern Germany, Denmark and southern Sweden as the homeland of Germanic peoples. Since Kossinna made his proposal, linguists (e.g. Nielsen 1989: 39) have consistently identified this area of Europe as the homeland of the Germanic peoples and the geographic point of origin for Germanic languages.

**Figure 2.4 Ertebølle Settlements 5000 BC.**



**Source: Hartz et al. 2007: 570.**

## 2.5 The "Kossinna Syndrome."

Prior to the Second World War, the prehistory of the Germanic peoples represented a potential research direction among scholars. However, since the war this research direction has become taboo, especially among German academics. The avoidance of Germanic prehistory is sometimes labeled the "Kossinna Syndrome." Nevertheless, it appears that some German academics believe that it is time for a change. In 2000, Sabine Wolfram, a German archaeologist, published an article maintaining that the "Kossinna Syndrome," has hindered German archaeology, and that German archaeologists have a severe handicap compared to their British and American counterparts. Using the term *Vorsprung durch Technik*, meaning 'progress through technical detail,' she (184) criticizes the lack of theoretical debate in German archaeology, that German archeology is mostly focused on empirical and data-oriented approaches, and avoids building models. A German-American archaeologist, Bettina Arnold, offered a similar critic of German archaeology. She ( 2000: 401) wrote "there is no theoretical foundation [in German archaeology] apart from the need to distance all archaeological research from theory." In my opinion, Wolfram's and Arnold's criticism is meant to address how German archaeology is essentially devoid of any ethnographic assessment of archaeological remains. Wolfram attributes the avoidance of theoretical debate to a failure of German archaeology to openly discuss the "ideological misuse of archaeology during the Third Reich." According to Wolfram (184-185), rather than discussing this issue, German archaeologists have turned Gustaf Kossinna into a scapegoat. For example, Karl Heinz Otto (1988: 29), a German archaeologist in the former East Germany, writes:

> The result of Kossinna's archeological-historical approach to archaeological theory was extraordinarily tragic. His archaeological theory, tainted by his nationalistic, ethno-centric, and racist views, led to the use of Germanic prehistory as support for imperialism and ultimately a fascist ideology … Kossinna's archaeological work asserted an undisturbed development of Germanic ethnicity since the Mesolithic, where Southern Scandinavia, Denmark and Schleswig-Holstein are seen as "purely Germanic soil" since the Late Neolithic.[1]

Commenting on Kossinna's theoretical approach to archaeology, Bettina Arnold (1990: 464), writes:

> The groundwork for an ethnocentric German prehistory was laid by Gustaf Kossinna … [who] proposed cultural diffusion as a process whereby influences, ideas and models were passed on by more advanced peoples to the less advanced with which they came in contact. This concept, wedded to Kossinna's *Kulturkreis* theory, the identification of geographical regions with specific ethnic groups on the basis of material culture, lent theoretical support to the expansionistic policies of Nazi Germany.

Heinz Grunert (2002: 339), Kossinna's biographer, writes:

> His ... monographs, above all the *German Prehistory - an Essential National Science* and the several versions of work published under the title *Old-Germanic Cultural Greatness* ... were objectively suitable to whitewash a national socialist ideology (or for what passed as an ideology) lacking in substance with a coat of scientific authority. Kossinna apparently submitted proof of an alleged German historical right to expand into other middle and eastern European territories. He contributed to fusing ethnic and national identity with race. In the process he strengthened the maxim of alleged racial superiority and the cultural supremacy of Germans over other cultures. In doing so he delivered important arguments for the justification and legitimatization of Nazi politicians, who were demagogic, ethnocentric and finally genocidal.

The term "Kossinna Syndrome" actually comes from an article published by Günter Smolla, a German archaeologist, in 1979/80. Smolla used this term to object to those who attacked Kossinna, both personally and professionally, in the post-war era. Smolla (1979/80: 8) takes the position that Kossinna was a "normal scholar." He writes:

---

[1] All translations from German are mine.

> Kossinna remains a significant yet rather difficult scholar, retaining both positive and negative character traits. His work shows both factual and methodological mistakes, but also correct and even stimulating findings. Apparently he was a 'normal' scholar. As soon as he recognized that he could proceed with the archaeological record where the linguistic evidence had hit a dead-end, he passionately attempted to present a dynamic picture of prehistory using ethnic groups whose identity were partially gleaned from areas that they had apparently occupied or vacated. The fact that he used terminology such as Germanic peoples or Indo-Europeans reflects a tradition that he encountered during his university education. He was essentially a product of his generation, adapting his theories to existing preconceptions, not one who re-invented the wheel.

My exhaustive review of Kossinna's life and work has failed to find any evidence of sympathy or association with the Nazi Party. Moreover, he died in 1932 before the Nazi Party came to power. On the contrary, after spending countless hours trying to uncover the great sin of German archaeology, I conclude that Kossinna has indeed become a convenient scapegoat. In my opinion, an honest assessment of the disgraceful direction that German archaeology took between 1933 and 1945 would fail to find fault with Kossinna, but rather would fault a Nazi organization called *Ahnenerbe*.

The word *Ahnenerbe* is difficult to translate into English. Some (e.g. McCann 1994: 79) translate the term as "ancestral inheritance." In my opinion, "German heritage" would be a better translation. I will simply use the German word for this organization. *Ahnenerbe* was created in 1934 by Heinrich Himmler, a leading figure in the Nazi regime. In Nazi Germany, *Ahnenerbe* allocated almost all of the funding for archaeological projects investigating the prehistory of Germanic peoples. These projects were funded to promote a belief in German racial superiority. Some of these projects were simply bizarre. For example, *Ahnenerbe* investigated a rumor that the ancient Germans procreated during the mid-summer so children could be born in the following Spring. The goal of the inquiry was to develop guidelines for the German soldiers in their duty to produce racially pure offspring (Pringle 2006: 121). Another bizarre research project involved "world ice theory," that an ancient catastrophe had blanked the entire world with ice except for a few remote areas at high altitude. An expedition to Bolivia was planned to determine if ancient Germans in the New World survived the catastrophe (Pringle 2006: 178-180). After the start of the Second World War, military projects took center stage in *Ahnenerbe* research. At this point, *Ahnenerbe* research turned from bizarre to inhuman. Dr. Sigmund Rascher, an *Ahnenerbe* researcher, conducted ghastly medical experiments on prisoners at the Dachau concentration camp (Kater 1974: 231-245). Another project, led by Dr. August Hirt, also an *Ahnenerbe* researcher, assembled a collection of skeletons using the corpses of prisoners that were gassed at the Natzweiler-Struthof concentration camp (Kater 1974: 245-255).

Clearly, the controversy surrounding Gustav Kossinna and the emergence of *Ahnenerbe* have tainted the search for Germanic origins. Scholars who explore Germanic origins must recognize that this research direction has the potential for abuse among those agenda is ethnocentric. In my opinion, an ethical and scholarly approach to exploring Germanic origins first views this language group as simply a co-equal member of the global linguistic tapestry. Secondly, the goal of this inquiry should merely attempt to better understand global language variation.

## 2.6 Germanic Origins from the Perspective of the Y-Chromosome.

The goal of this dissertation is to demonstrate that genetic data, especially Y-chromosome data, are a useful tool for evaluating models of Germanic origins. My research has uncovered a single published report that used Y-chromosome data for exploring Germanic origins. The paper was published in 2008 by Kalevi Wiik, a phonetician and professor emeritus at the University of Turku in Finland. He posits (83) that emergence of Proto-Germanic involved language shift from Uralic to Germanic. Nevertheless, in my opinion the potential of population genetics still remains a research direction that has not been fully appreciated by some researchers. Perhaps one explanation is that this research has only emerged in the last decade. Geneticists began to focus on human molecular variation about thirty years ago. However, the pace of this inquiry finally accelerated in the late 1990s with the development of new technology such as Denaturing-High Performance Liquid Chromatography (D-HPLC), a development that has made the detection of Y-Chromosome mutations "easy, fast and inexpensive" (Francalacci and Sanna 2008: 60). This development triggered a flood of population reports, beginning in 2000, describing world-wide Y-haplogroup variation and the evolutionary history of various human populations.

During my research, I found that Y-chromosome data are currently very fragmented, which may also explain why this research direction remains unrecognized by some in the academic community. For this dissertation I actually had to gather my data from approximately two hundred and forty published reports. Another huge problem for the some researchers is the nomenclature system used by geneticists to describe Y-chromosome variation. This system has been at times inconsistent, and subject to refinement and revision. Initially, geneticists used several different nomenclature systems to describe Y-chromosome haplogroups. For example, Rosser (2000) used "Haplogroup 3" to describe the current R-M17 mutation, whereas Semino (2000a) used "Eu 19." In 2002 the Y-chromosome Commission standardized the nomenclature and "R1a1" became the cladistic label for the R-M17 mutation. In 2008, Karafet and others revised the nomenclature system and, for example, the cladistic description for the M178 mutation changed from "N3a" to "N1c1." In addition to the two official revisions I just described, a number of "unofficial" revisions have also taken place, where a group of researchers rather than a specific organization change the nomenclature. Moreover, researches sometimes use a nomenclature that is different from the official standard, or different from that used by another researcher. Please refer to the table below.

**Table 2.3 Nomenclature for Haplogroup I Sub-Clades.**

| | Scandinavian | Balkan | Central European | Sardinian |
|---|---|---|---|---|
| Rootsi et al. 2004 | I1a1-M253 | I1b-P37.2 | I1c-M223 | I1b2-M26 |
| Underhill et al. 2007 | I1-M253 | I2a2-M423 | I2b1-M223 | I2a1-M26 |
| Karafet et al. 2008 | I1-M253 | I2a-P37.2 | I2b-M223 | I2a2-M26 |
| Battaglia et al. 2009 | I1-M253 | I2a1-M423 | I2b1-M223 | I2a2-M26 |
| Mirabal et al. 2009 | I1-M253 | I2a-P37.2 | I2b1-M223 | I2a1-M26 |
| Pala et al. 2009 | I1-M253 | I2a1-M423 | I2b1-M223 | I2a2-M26 |

As shown in Table 2.3, Rootsi and others defined four common sub-clades of the I-M170 mutation in 2004. In 2007, Underhill and others revised the classification. However, Karafet did not recognize the new I-M423 mutation, and used the old I-P37.2 mutation. For the I-M26 mutation, Karafet also used a different cladistic description, “I2a2” rather than “I2a1.” Battaglia, in 2009, used the I-M423 mutation from Underhill, but with a different cladistic description. The same report adopted Karafet’s cladistic description for the I-M26 mutation. In 2009, Mirabel and others continued to use the I-P37.2 mutation rather than the Underhill I-M423 mutation, but nevertheless used the Underhill cladistic description for the I-M26 mutation, rather than the one used by Karafet and Battaglia. In the 2009 report by Pala and others, the researchers adopt the same nomenclature as Battaglia.

A final reason why Y-chromosome data may be underutilized is that the target audience for published research in this area has been largely geneticists. In my opinion, the methodology of population genetics needs to be explicated so that a wider audience can evaluate the potential of this new research. This task will be undertaken in the next chapter.

## 2.7 Chapter Conclusion.

This chapter defines Germanic as a branch of the Indo-European language family. As such, the origins of Germanic languages are linked to the origins of Indo-European languages. Today, the linguist encounters two different models attempting to explain the putative homeland and the expansion of Indo-European languages across Europe, either the diffusion of agricultural technology across Europe, or alternatively the Kurgan expansion. Traditionally, two different linguistic approaches have been used to explain why Germanic is a part of the Indo-European language family, *Stammbaum* Theory and Language Contact Theory. Either Germanic diverged from Proto-Indo-European and developed independently, or alternatively, two or more languages converged to produce Germanic. The archaeological approach to Germanic origins posits southern Sweden, Denmark and northern Germany as the putative Germanic homeland. Unfortunately, archaeological research into the prehistory of Germanic peoples has stagnated since 1945 and the end of the Second World War. Finally, this chapter introduces a potential new tool for exploring the origins of Germanic languages. However, this research remains underutilized due to the fragmented reporting of the data, and inconsistent nomenclature system, and a methodology that awaits further clarification.

# Chapter Three

# Why the Y? The Y-Chromosome as a Tool for Understanding Prehistoric Migration.

## 3.0 Chapter Introduction.

In this chapter I will explain how Y-chromosome data has emerged as a powerful tool for tracing prehistoric migration and settlement. By avoiding recombination, the Y chromosome provides a genetic record that is transmitted largely intact from one generation to the next. Nevertheless, single nucleotide polymorphisms, a type of genetic mutation, distinguish one Y-chromosome from the next. The term haplogroup is used to refer to these mutations. Moreover, short tandem repeats, another type of mutation, provides a means of dating the evolution of a haplogroup. By examining haplogroup frequencies in modern populations, and by having a rough idea when the various haplogroups arose, geneticists are able to postulate several population expansions that define the human prehistory. While the Y-chromosome represents one of several potential sources of genetic data for deciphering prehistoric human population expansion, this dissertation focuses on Y-chromosome data because of the volume of published data, and because this data is easier to understand.

## 3.1 Playing by-the-Rules.

In their article “The human Y chromosome: an evolutionary marker comes of age,” Mark A. Jobling and Chris Tyler-Smith, two geneticists, claim that the Y-chromosome does not play according to the rules of genetics, and for this reason has emerged as “a superb tool for investigating recent human evolution from a male perspective” (2003: 598). In order to understand how the Y-chromosome behaves differently from other human chromosomes, at least from the viewpoint of geneticists, it is necessary to briefly discuss Mendelian genetics, which is often part of high school and introductory college biology instruction. According to Mendelian genetics, we inherit our genes from both parents. However, the Y-chromosome plays by its own genetic rules in that it is only passed from a man to his son. The Y-chromosome is one of the two sex-chromosomes in the human genetic inventory, or human genome. The other sex chromosome is the X-chromosome. During human reproduction, two X-chromosomes yield female offspring, and an X-chromosome and a Y-chromosome yield male offspring. Consequently, a male can only inherit the Y-chromosome from his father. Another “rule” of Mendelian genetics is recombination. During human reproduction, the genetic cards are essentially “reshuffled,” or more precisely, recombination occurs. For example, Mendelian genetics would define hair color as a genetic trait or phenotype, and variations of this phenotype, such as blonde hair and red hair, as alleles. Because of recombination, parents may have blonde hair, but their child itself may potentially inherit red hair from a grandparent. However, the genetic material contained in the Y-chromosome, for the most part, escapes recombination, providing yet another example of failing to follow the rules of genetics.

In order to explain how the Y-chromosome avoids recombination, it is necessary to briefly discuss the evolutionary history of this chromosome. The sex-determining locus of the Y-chromosome not only codes for maleness in humans, but in all mammals. This section of the Y-chromosome, however, only represents a fraction of its entire length. During the evolutionary history of mammals, about 300 million years, the Y-chromosome has, in the words of some geneticists, slowly “degenerated” or degraded (Lahn et al. 2001: 211). When mammals first evolved, the Y-chromosome “behaved normally” in that the entire chromosome recombined with the X chromosome. Now, as the result of slowly evolving structural decay, about 95% of the entire length of the Y-chromosome has been damaged, emerging in what the geneticists call a “non-recombining region.” This large non-recombining region means that during human reproduction, very little genetic exchange occurs between the X and Y chromosome. Consequently, the Y-chromosome is transmitted from one male to the next largely intact.

## 3.2 Mutation.

So far this chapter has explained that the Y-chromosome is unique, partly due to uni-parental inheritance, and partly due to the absence of recombination. Consequently, men inherit a large section of genetic information that remains unaltered from their fathers. However, the non-recombining region of the Y-chromosome can and often varies from one Y-chromosome to the next. Geneticists describe this variation as mutation. In population studies examining Y-chromosome variation, two different types of mutation are particularly informative: single nucleotide polymorphisms and short tandem repeats.

Before discussing Y-chromosome mutations in detail, I want to emphasize a concept known as neutral selection. I emphasize this concept to avoid misconceptions that may arise among those whose knowledge of human genetics is rudimentary. Those who have taken an introductory biology or physical anthropology course have probably encountered the term “natural selection,” initially proposed by the naturalist Charles Darwin. This theory accounts for different animal and plant species based on fitness, or survival of the fittest. According to this theory, differentiation among species arose as the result of a mutation that enabled the plant or animal to survive in a given environment long enough to pass on its genes to the next generation. Y-chromosome mutations, however, are classified as selectively neutral, meaning they do not confer any reproductive advantage.

Another point also needs to be clarified before discussing Y-chromosome mutations. As explained earlier, Y-chromosome mutations are not reproductively advantageous. Likewise, these mutations are not disadvantageous. Introductory biology courses often emphasize that genetic mutations can be harmful or fatal to living organisms. For example, among humans one of the most recognized harmful genetic mutations is sickle cell anemia. In contrast to sickle cell anemia and other genetic mutations, Y-chromosome mutations are totally benign. This explains, partially, why Y-chromosome mutations survive, while many genetic mutations affect reproductive success and are consequently eliminated from the gene pool.

### 3.2.1 Single Nucleotide Polymorphisms.

As detailed above, the non-combining region of the Y-chromosome can vary from one male to the next because of mutations which are selectively neutral and benign. Furthermore, two different types of mutations have emerged as particularly informative in examining population history from a male perspective. One of these mutations is classified as a single nucleotide polymorphism. This mutation is also described in population reports as a unique mutational event, or as a haplogroup, or as a clade, and sometimes as an allele. For the non-geneticist, the use of so many essentially synonymous terms certainly poses a challenge in understanding the literature.

Focusing now on the term single nucleotide polymorphism, it is useful to discuss the structure of DNA, short for deoxyribonucleic acid (cf. Figure 3.1 below). The structure of DNA resembles that of the non-recombining region of the Y-chromosome. This molecular “ladder” has “rails” formed by alternating sugar and phosphate molecules The “rungs” of this ladder, known as nucleotides, are formed by bonding two molecules having a nitrogenous base; either adenine and thymine, or guanine and cytosine. The order of the bonding can alternate, meaning the nucleotides appear in one of four different combinations: adenine/thymine, thymine/adenine, guanine/cytosine and cytosine/guanine. A single nucleotide polymorphism occurs when one of the rungs of our molecular ladder changes, or mutates. An example of a mutation would be a nucleotide reversal from adenine/thymine to thymine/adenine. Sometimes a mutation entails the addition or deletion of a nucleotide. These single nucleotide polymorphisms are sometimes referred to in population reports as unique mutational events, because they are so rare they only occur once during human evolution. The evolutionary rate of mutation for single nucleotide polymorphisms is estimated to be about $10^{-8}$ per base pair per generation (Novelletto 2007: 140). Since the non-recombining section of the Y-chromosome has about 60 million molecular “rungs,” or base pairs, geneticists have a vast region of genetic information to harvest the evolutionary history of human males.

As explained in the above paragraph, single nucleotide polymorphisms are a common mutation occurring in the non-combining region of the Y-chromosome. Geneticists comb the non-recombining region to identify these mutations. The presence or absence of these mutations, or single nucleotide polymorphisms, can distinguish the Y-chromosomes of one male population from the next. Another and perhaps more common label for a single nucleotide polymorphism is the term haplogroup. Haplogroups are reported in population reports using a nomenclature system first standardized in 2002 by the Y Chromosome Consortium. This standard uses an uppercase letter to identify major haplogroups. An uppercase letter followed by a combination of numbers and lower case

**Figure 3.1 The Structure of DNA.**

*The fact that nucleotide bases vary is central to population genetics.*

```
P                   P
 \                 /
  S --- A --- T --- S
 /                 \
P                   P
 \                 /
  S --- C --- G --- S    }nucleotide
 /                 \
P                   P
 \                 /
  S --- G --- C --- S
 /                 \
P                   P
 \                 /
  S --- T --- A --- S
 /                 \
P                   P
 \                 /
  S --- A --- T --- S
 /                 \
P                   P
```

**P = Phosphate** **S = Sugar** **C = Cytosine**
**G = Guanine** **T = Thymine** **A = Adenine**

letters is used to report sub-haplogroups. Haplogroup nomenclature often adds the mutation number, usually prefixed with the abbreviation “M,” and sometimes “P” or “V.” For example, one very common haplogroup found in Europe is I-M170, the “I” meaning haplogroup I, the M170 referring to mutation number 170. An example of a sub-haplogroup is the I1-M253 mutation, commonly found in Scandinavia. The “1” is used to classify I1-M253 as a sub-haplogroup of haplogroup I. Often the literature does not make a formal distinction between major haplogroups and sub-haplogroups, and thus I-M170 and I1-M253 would simply be reported as “haplogroups.” Haplogroups and their subgroups represent terminology used to build phylogenetic trees, hierarchical relationships between polymorphisms, very much akin to language trees utilized by linguists. Since the methodology used to build these hierarchical relationships is called cladistics, the terms clade and subclade are sometime used to label haplogroups and sub-haplogroups.

### 3.2.2 Short Tandem Repeats.

Hierarchical relationships between single nucleotide polymorphisms, or haplogroups, are partially defined by the age of the polymorphism, which is also considered a unique mutational event. To obtain a rough date for a unique mutational event, geneticists use another type of Y-chromosome mutation called short tandem repeats, often referred to as “microsatellites” in the literature. Short tandem repeats are defined as repeated units of one to six nucleotide base pairs. The molecular rungs of the molecular ladder in Figure 1.1 mutate in such a way that a section of nucleotides repeat over and over again. For linguists, “stuttering,” a speech impediment may be a useful analogy for this type of mutation (Cavalli-Sforza 2000: 82). Compared to single nucleotide polymorphisms, short tandem repeats have a much faster rate of mutation. Moreover, the rate of mutation is believed to occur at a relatively constant rate, estimated at $6.9 \pm 1.3 \times 10^{-4}$ per base pair per generation (Zhivotovsky et al. 2004: 54-55). Consequently, geneticists can obtain a rough date for a unique mutational event by counting the number of short tandem repeat mutations that accrued since the emergence of a new haplogroup. This is very much akin to dating a tree by counting the number of rings. Perhaps another useful analogy is the barnacles that accumulate on the bottom of a boat; the number of barnacles that have accumulated provide a clue as how long the boat has been in the water. Short tandem repeats also have other applications in genetic research and are insightful for examining more recent population histories, and affinities among and between populations.

## 3.3 Population History

The non-recombining region of the Y-chromosome accumulates single nucleotide polymorphisms, which are considered unique mutational events, which are reported as haplogroups. Moreover, short tandem repeats provide a molecular clock for determining when a unique mutational event occurred. In order to discuss how these mutations assist the researcher in identifying prehistoric migration and settlement of human populations, it is now necessary to present three additional concepts: population, polymorphism, and drift. The most basic definition of a population is a group of individuals in a given geographic area who reproduce. Most modern populations are polymorphic, meaning that more than one Y-chromosome haplogroup is found in the population. However, when populations are small and isolated, characteristic of most populations in the human prehistory, the haplogroup composition tends to be less polymorphic because of genetic drift. The term drift refers to a leveling of genetic variation between members of a population. For the purposes of this discussion, this means that most of the male members of the population have the same Y-chromosome haplogroup.

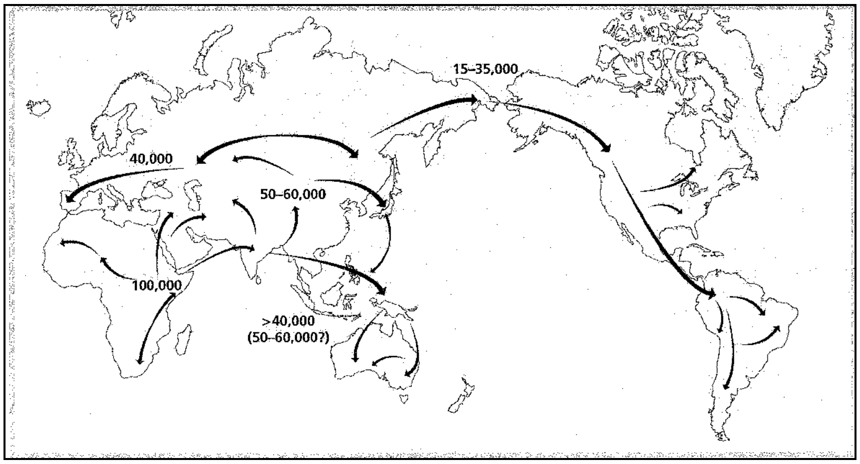
The term “founder effect” illustrates a type of genetic drift and is useful concept in understanding the development of genetic differentiation between populations. Founder effect describes a situation when a group of people separates from a larger population. The new group may have a different haplogroup composition than their ancestors. For example, the ancestral group may have been 60% haplogroup F and 40% haplogroup I. The group that separates, on the other hand, may have been 40% haplogroup F and 60% haplogroup I. The ancestral population, will over time, have an increasingly greater proportion of F haplogroups to I haplogroups due to genetic drift, whereas drift produces an increasingly greater proportion of I haplogroups to F haplogroups in the new population.

Another type of genetic drift that sometimes results in genetic differentiation between populations is “bottleneck.” This term describes a situation where perhaps disease or a natural disaster suddenly reduces the size of a population that is isolated and relatively small. This sudden reduction in population reduces the amount of haplogroup variation, and like founder effect, accelerates drift. Again, drift effects small populations more than large populations. The Black Plague, for example, killed about a quarter of the Europeans during the Middle Ages, yet did not produce a “genetic scar” or bottleneck because of the large human population at the time (Zerjal et al. 2002: 466).

By examining haplogroup frequencies among modern populations residing in different regions of the world, and having a rough idea when a haplogroup evolved, researchers can often determine how and when a prehistoric group may have migrated. A migration consists of a geographic point of origin and a geographic point of termination separated by distance. Sometimes the point of origin has the greatest frequency of a particular haplogroup, and over distance the frequency of this haplogroup diminishes. For example, haplogroup J2-M172 arose in the Near East during the Mesolithic. During the Neolithic this group migrated to Western Europe, and along this route the frequency of haplogroup J decreased because of admixture with other groups already living in the new territory. However, some prehistoric migrations show an opposite pattern or cline of haplogroup frequencies, where the point of origin has the lowest frequency of a certain haplogroup, and the terminal end of the migration has the highest. Haplogroup I1-M253, for example, represents a prehistoric migration from the current Spanish/French border to Scandinavia. Along this route the frequency of Haplogroup I1a increases, perhaps because this group moved into unoccupied territory, or perhaps this group acquired a novel survival strategy that gave them a reproductive advantage. When determining the point of origin for a prehistoric migration, geneticists often turn to the archaeological record. Also, geneticists look for the amount of variation that a particular haplogroup may have, and posit that the point of origin is where a haplogroup has the most variation. The correlation between the point of origin and greater genetic variation is based on the assumption that greater genetic variation is a product of greater time depth.

By examining the worldwide distribution of Y-chromosome haplogroups, and the frequency of these haplogroups from one region to the next, geneticists have developed an important tool for tracing prehistoric migration and settlement. Y-chromosome data point to human origins in Africa about 140,000 years ago (Cruciani et al. 2011: 815). If the evolutionary history of the human Y-chromosome were represented as a tree, the base of this tree would be haplogroups A and B. These haplogroups are confined to sub-Saharan Africa, and found in Bantu, Khosian and Pygmy populations (Underhill et al. 2001: 47). Climbing further up the tree, haplogroup CR-M168 separated from haplogroup B, and then separated into haplogroups C, DE and F. Going a step further, haplogroup F-M89 represents the main out-of-Africa human migration about 45,000 years ago.

**Figure 3.2: Out of Africa – The Story of Human Migration as Shown by Genetic Variation.**



*Source: Cavalli-Sforza and Feldman 2003 p. 270*

### 3.4 Other Molecular Markers.

The term “marker” refers to a section of DNA. Geneticists have found several different markers for measuring genetic variation among human populations. Among these markers is the Y-chromosome. In this section I will explain why I have focused on the Y-chromosome data in this dissertation instead of classical genetic markers or mitochondrial DNA. This discussion begins with a 1919 study published by Ludwik and Hanka Hirschfeld, two researchers at a military hospital, who were the first to use a genetic marker to assess human variation. The researchers utilized ABO blood groupings, a type of human protein variation, to find patterns of variation among different nationalities and ethnic groups (Hirschfeld and Hirschfeld 1919). Proteins are long chains of DNA, or peptides consisting of amino acids. Protein markers used to assess human genetic variation are generally referred to as "classical markers" in the literature. In 1994 Cavalli-Sforza and others published the most comprehensive study of human variation based on classical markers. In their study, the researchers conceded, however, that another type of marker, mitochondrial DNA, would be a better choice for population studies, but at the time a sufficient number of haplogroups had not yet been discovered (Cavalli-Sforza et al. 1994: 9-10).

In 1985 geneticists developed a new technique for sequencing DNA called Polymerase Chain Reaction (PCR). This and other innovations allowed geneticists to focus on molecular markers, rather than on classical markers, as a source of human variation. Molecular markers are nucleotide variations, whereas classical or protein markers utilize a much longer section of DNA. Because of vastly improved sequencing techniques, at the end of the 1990s reports describing European mitochondrial DNA variation began to appear (e.g. Richards et al. 1998). Soon thereafter, in 2000, reports of European Y-chromosome variation also appeared (e.g. Semino et al. 2000a). Mitochondrial DNA and the non-recombining region of the Y-chromosome have certain features in common. Both types of DNA avoid recombination, and thus are transmitted largely intact from one generation to the next. The source of variation for both types of DNA is mutations. These mutations are commonly described as haplogroups. Haplogroups from both DNA markers form a phylogeographic model reflecting the origins of anatomically modern human beings in Africa and their ultimate dispersal into Asia and other regions of the world. Finally, dating techniques have been developed for both types of DNA to estimate when a population expansion may have occurred. However, the fundamental difference between mitochondrial and Y-chromosome haplogroups is that Y-chromosome data reflects the evolutionary story of males, and mitochondrial DNA reflects the evolutionary history of females. Since the Y-chromosome determines male gender, Y-chromosome variation is only transmitted from father to son (cf. Section 3.1). Mitochondrial DNA, on the other hand, is found in the cell wall, and since the sperm cell lacks a cell wall, this type of DNA is only inherited from the mother.

Researchers have attempted to use other molecular markers, such as autosomal DNA and X-chromosome DNA, to circumvent the inherent weakness of Y-chromosome and mitochondrial DNA. Autosomal and X-chromosome DNA have the potential of elucidating the evolution of the entire human species, rather than just men or women. However, these markers require more sophisticated theoretical analysis, and partly for this reason, geneticists have focused on Y-chromosome and mitochondrial data. Data from mitochondrial and Y-chromosome DNA can also be represented by a tree, which makes this data easier to understand. Consequently, mitochondrial and Y-chromosome data still remain the genetic tools most commonly used for interpreting human population history. However, mitochondrial DNA fails to provide the same degree of resolution for investigating human expansions in Europe since the end of the last Ice Age (Richards et al. 2002: 1168; Barbujani and Bertorelle 2001: 23). Geneticists cite human behavior as a possible explanation for differences in mitochondrial and Y-chromosome data. Differences in male and female behavior may include differing migration patterns, differences in reproductive success, differences in mortality, polygamy versus polyandry, or patrilocality versus matrilocality (Rubicz et al. 2007: 150-151 ; Peričić et al. 2005a: 502-503).[2] An alternative explanation for differences in mitochondrial and Y-chromosome DNA data may stem from the size of these markers. Mitochondrial DNA consists of about 16,000 base pairs, whereas the non-recombining region of the Y-chromosome is much larger with 60 million base pairs. According to Underhill and Kivisild, because of its larger size, the Y-chromosome inherently better preserves cladistic relationships than mitochondrial DNA (2007: 551). This is analogous to image quality when a ten megapixel camera is chosen over a two megapixel camera. Nevertheless, a recent improvement in mitochondrial DNA dating techniques may improve the resolution of this marker (Soares et al. 2010).

The reason why the Y-chromosome takes center-stage in this dissertation is that this marker currently provides the most published genetic data for deciphering human migration in Europe within the last 12,000 years. Nevertheless, mitochondrial DNA will play an important supporting role in this dissertation. Moreover, I would like to emphasize that the mitochondrial DNA evidence does not generally contradict the Y-chromosome data, but rather Y-chromosome data often provides a more detailed picture of prehistoric population expansions in Europe.

### 3.5 Chapter Conclusion.

This chapter has introduced several key concepts that explain how the Y-chromosome has emerged as a tool for investigating prehistoric migration and settlement. Prehistoric populations tended to have their own genetic signature due to drift. Consequently, haplogroup variation over a given geographical distance in modern populations provides information as to the origin and direction of a past migration. From a worldwide perspective, Y-chromosome data point to human origins in Africa and a later out-of-Africa migration to the other continents. The next chapter will focus on the haplogroups that tell the story of the human settlement of Europe since this dissertation is committed to clarifying the origins of Germanic languages. The story of human migration and settlement of Europe, from the Y-chromosome perspective, is told by haplogroups R, I, N, E, J and G. The current distribution of these haplogroups in the modern European gene pool resulted from three principal population expansions in the prehistory: the initial migration of anatomically modern humans into Europe during the Pleistocene, about 30,000 years ago; the human re-colonization of Europe following the last Ice Age, about 12,000 years ago; and finally, the arrival of agriculture during the Neolithic, beginning about 8,000 years ago.

---

[2] Patrilocality means that people live closer to fathers place of birth, and matrilocality means living closer to the mother's place of birth.

# Chapter Four
# Y-Chromosome Haplogroups R, I, N, E, J and G . Population Expansions in the Paleo-, Meso-, and Neolithic.

## 4.0. Chapter Introduction.

This chapter explains how six of the major Y-chromosome haplogroups (R, I, N, E, J and G) shaped the demographic prehistory of the European continent. My source of information for these haplogroups stems primarily from approximately two hundred and forty population reports published online in respected peer review journals.

To tell the story of the European prehistory from a Y-chromosome perspective, in this chapter I transform six different major Y-chromosome haplogroups into ten different population expansions. Haplogroups E, G, J, and N describe each a single population expansion, whereas haplogroup R has two sub-haplogroups with their own demographic history, and haplogroup I has four informative sub-haplogroups with their own history. In my discussion of the data, I identify and label each population expansion using the main haplogroup name and the place where the genetic evidence for the expansion attains its highest frequency. For example, the I-M253 sub-haplogroup is labeled the "Scandinavian I-Group." The I-M253 mutation attains its highest frequency in Scandinavia, hence the word "Scandinavian." The term "I-Group" reflects that I-M253 is part of the larger I haplogroup. My demographic labels presents a strategy for unifying data that often have been reported under differing and inconsistent nomenclature schema.

To help the reader in achieving a rough overview of my data, Figure 4.1 (below) provides a diagram that reflects the evolution of Y-chromosome haplogroups defining the European prehistory. Each circle represents a population. Within each circle is the Y-chromosome mutational signature of the population. The arrows represent the emergence of a group from an ancestral population. The population identified with the CR-M168 mutation (top-center part of the diagram) represents the evolution of our species in Africa. One descendant population has the F-M89 mutation. The F-M89 population represents the main out of Africa migration of our species. Descendant populations of the F-M89 group terminate in the following mutations:

- M423 (the Balkan I-Group)
- M26 (the Sardinian I-Group)
- M223 (the Central European I-Group)
- M253 (the Scandinavian I-Group)
- J-M172 (Near Eastern J-Group)

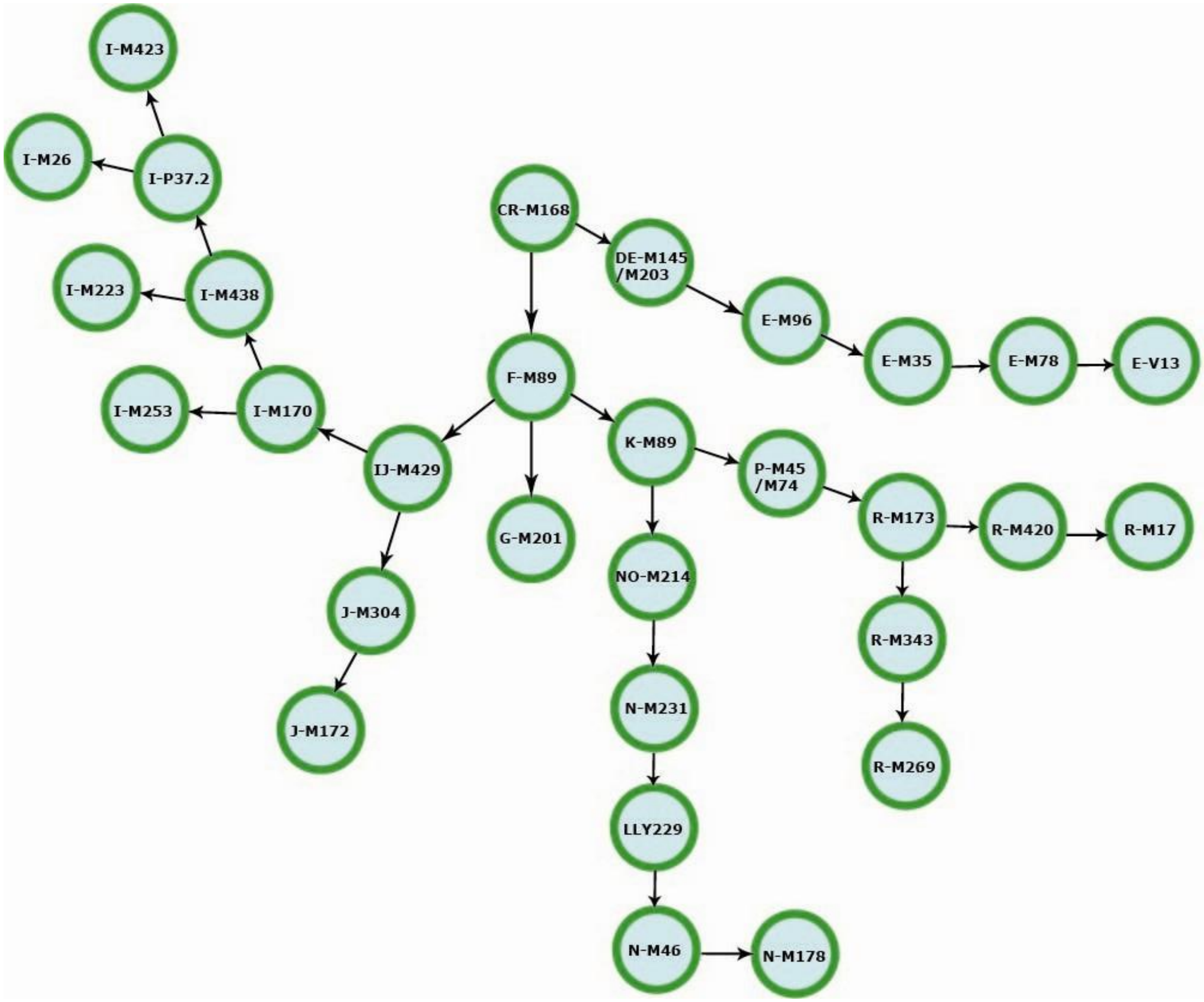
- G-M201 (Caucasus G-Group)
- N-M178 (Finno-Baltic N-Group)
- R-M269 (Western European R-Group)
- R-M17 (Eastern European R-Group)

The other population emerging from the CR-M168 mutation (human origins in Africa) is identified with the DE-M145/M203 mutation. This population, in turn, terminates with a population having the E-V13 mutation (the European E-Group).

For each population expansion, I also identify the formal current cladistic description of mutation that describes the event. For example, the current cladistic description for the Western European R-Group is R1b1b2-M269. However, most of my discussion of the Y-chromosome data alternates between the current mutation number (i.e. "the R-M269 mutation") and the demographic label (i.e. "the Western European R-Group"). This seems to facilitate a more elegant discussion of the data. Moreover, when discussing haplogroup frequencies, the term "low" means that less than 10% of the population has a given haplogroup, "elevated" describes a frequency between 10% and 19%, "moderate" 20% to 49%, and "heavy" is 50% and greater. Each population expansion is supported by a table in the Appendix that lists the genetic data, first by region, then by current geo-political boundary, followed by percentage of the population having the genetic signature for the population expansion, followed the nomenclature used in the report, and finally the source of the data.

Certain definitions are needed to understand the data in the sections that follow. The term "Paleolithic" refers to prehistoric Europe prior to the end of the last Ice Age, about 12,000 years ago. The Mesolithic began when the last Ice Age ended (about 12,000 years ago) and ended with the adoption of farming. Since people in Europe adopted farming at different times in the prehistory, the Mesolithic/Neolithic transition occurred roughly 8,500 years ago in southeastern Europe, and as late as 5,000 years ago in northern Europe.

**Figure 4.1 An Overview of Haplogroups that Define the European Prehistory.**



## 4.1 Haplogroup R.

Haplogroup R is an ancient Y-chromosome haplogroup distributed throughout western Eurasia. In a 2001 report, Peter Underhill, a geneticist at Stanford University, details the evolutionary history of the R-M173 mutation. Underhill starts with the main out-of-Africa migration, about 45,000 years ago, with a group having the F-M89 mutation. After migrating into the Levant, the F-M89 group split into several different groups. One of the F-M89 splinter groups was a population having the R-M173 mutation, which may have arisen in northern Asia. From northern Asia, those having the R-M173 mutation dispersed into Europe, the Caucasus, the Middle East, Central Asia, Pakistan and northern India (Underhill et al. 2001: 53-54).

Two sub-haplogroups, currently identified in the literature as R1b1b2-M269 and R1a1a-M17, have emerged as the markers of choice for describing R haplogroup variation in Europe (Balaresque et al. 2010; Underhill et al. 2010). This dissertation will label haplogroup R1b1b2-M269 as the Western European R-Group, and haplogroup R1a1a-M17 as the Eastern European R-Group. Again, I would like to emphasize that my labeling describes where the defining mutation for the population expansion attains its maximum frequency, not where the defining mutation arose. Also, as explained previously, I believe that the use of such labels will provide a more stable descriptor for the data.

In 2000, two papers emerged as “seminal” in the description of European Y-chromosome variation (Novelletto 2007: 158). One of these papers (Semino 2000a) identified the Western and the Eastern European R-Groups as present in 50% of the European gene pool. According to the same report, both R groups descended from the R1-M173 haplogroup, which represents a migration of anatomically modern human beings from Asia to Europe roughly 40,000 years ago, during the Paleolithic (1155-1156). However, subsequent population reports take the position that more recent population expansions explain the current distribution of the R-M269 and R-M17 mutations in Europe.

### 4.1.1 The Western European R-Group.

The Western European R-Group is currently defined by the R1b1b2-M269 haplogroup. As shown by Table 1 in the Appendix, the R-M269 mutation attains its maximum frequency in Western Europe, with a frequency exceeding 50% for almost all the populations studied in this region. Moving eastwards to the Mediterranean Region, the Western European R-Group is found in the elevated to moderate range, 10-30% in the populations surveyed. Northern Italy is an exception to this pattern, where the R-M269 mutation exceeds 50% in the populations sampled. In central Europe, the Western European R-Group is found at a moderate frequency. In the Balkans, the frequency pattern is moderate. In the Middle East, the frequency is low, with the possible exception of Turkey, where the frequency is elevated. The frequency also remains low in south central Asia. Turning now to northern Europe, the Western European R-Group is found at moderate frequencies, with the exception of Finland, where the frequency is low. In Eastern Europe, the frequency is low to elevated. Finally, population reports describe the frequency of the R-M269 mutation as low in the Baltic Region.

Geneticists are currently divided as to source of the R-M269 mutation in Europe, whether the R-M269 mutation expanded throughout Europe from Iberia, the Middle East, or both regions. In 2010 Balaresque and others published a report strongly advocating a single-region expansion of the R-M269 mutation from Turkey during the Neolithic. Their argument is partly based on their estimated age of this mutation, which they date at roughly 8,000 years for populations in central Turkey, and roughly 5,500 years for populations in England and Ireland (6). In 2011, Myres and others published a study that also advocated a single Near-Eastern source for the R-M269 mutation. This study found additional downstream markers within the R-M269 mutation, R-U106 and R-S116. Both downstream markers represent almost all of the R-M269 variation in Europe. Furthermore, the distribution of both markers corresponds to the first agricultural expansion across Europe, carried by the Linearbandkeramik (LBK) culture (100). Interestingly, single-region expansion from Turkey was also taken by Hill and others a decade earlier in their discussion of Y-chromosome diversity in Ireland (2000: 351).

The mainstream opinion among geneticists may still attribute the distribution of the R-M269 mutation to a primary post-Ice Age expansion from Iberia and a secondary and perhaps later expansion from Turkey (Soares 2010: R178). Such a scenario was suggested by Peričić and others in 2005 based on short tandem repeat data in the Balkans ( 2005b: 1970-1971). This was also suggested by Cinnioğlu and others in 2004 based on short tandem repeat data, and more specifically, the distribution of haplotypes 15 and 35 in the Middle East and Europe (134). Al-Zahery and others also found that populations having the R-M269 mutation in the Middle East lack haplotype 15, whereas European populations with the M269 mutation have haplotypes 15 and 35. Al-Zahery and others suggest that haplotype 35 is older, and haplotype 15 is derived from haplotype 35. They further suggest this occurred when human populations were isolated during the last Ice Age (2003: 469). A report examining the presence of the R-M269 mutation in Lebanese Christians also suggests that short tandem repeat data may support two different source areas for the Western European R-Group. In this report by Zalloua and others, researchers found that European R-M269 mutations have the Western European Specific 1 (WESI) haplotype which is absent in Middle Eastern populations with the same mutation (2008b: 879). Finally, Morelli and others recently challenged the position taken by Balaresque and others, asserting that the Western European and Anatolian M269 mutations are distinguishable by unique short tandem repeat profiles, the Atlantic Modal haplotype versus the Eastern European haplotype (2010: 2). Morelli and others also assert that the dating method used by Balaresque and others was flawed, and that the date of the M269 mutation indicates a pre-Neolithic expansion of the R-M269 mutation in western Europe, and a second more recent expansion from Anatolia (2010: 5-8).

Based on additional data from Spain, the single-region expansion scenario (Balaresque et al. 2010 and Myres et al. 2011) may require reconsideration. The R-M169 and R-M153 mutations are two downstream mutations of R-M269 mutation found in populations of the Pyrenees Region of Iberia, a purported European refuge area from which human populations expanded following the last Ice Age. These mutations were recently dated by López-Parra and others at 7,383 years and 8,453 years respectively (2009: 48). This is substantially older than the date of the ancestral R-M269 mutation provided by Balaresque for populations in Spain and the Basque Region of France. Finally, the single source scenario, proposed by Balaresque, may require reconsideration based on the archaeological record of Scandinavia as well as M269 data from Sweden. In Sweden the R-M269 mutation was dated at 9,100 years by Karlsson and others (2006: 967). Moreover, in most of Europe, the expansion of agriculture during the Neolithic occurred as the result of farmers moving into uninhabited areas. However, in Scandinavia agriculture was adopted by people already living in the area (e.g. Bellwood 2005: 77-79). Consequently, the date of R-M269 variation in Sweden, as well as the indigenous adoption of agriculture in this area, undermine a single Neolithic population expansion as the original source of the R-M269 in Scandinavia. Balaresque and others does not provide any data for Sweden or Norway yet provide a date of 6,555 years for the M269 mutation in Denmark (2010: 6). The data from Denmark may represent an admixture of R-M269 variants from two different source regions, whereas the data from Sweden may reflect data for a single source for the M269 mutation, possibly a Mesolithic population expansion from Iberia.

### 4.1.2 The Eastern European R-Group.

Please refer to Table 2 in the Appendix. The Eastern European R-Group is currently defined by the R1a1a-M17 haplogroup. The R-M17 mutation attains its maximum frequency in Eastern Europe, particularly in Poland, followed closely by Russia and the Ukraine. In central Europe moderate frequencies of this R-Group are found. Population reports describe elevated to moderate frequencies of the R-M17 mutations for populations in the Balkans. In Western Europe, the frequency of the Eastern European R-Group drops significantly to very low levels. Surprisingly, moderate frequencies of the Eastern R-Group are reported for Norway and Iceland. Moderate frequencies are also found in the Baltic Region. Similar to Western Europe, low levels are reported for the Mediterranean Region, except for Greece, where the frequency is elevated. Low or slightly elevated levels of the Eastern European R-Group are reported in the Middle East and Caucasus. Finally, moderate frequencies of the R-M17 mutation are reported in some populations of India and Pakistan.

In their 2000 report, Semino and others (2000a: 1156) propose that the Eastern European R-Group arose in the Ukraine and may have spread throughout Europe with the Kurgan expansion and the associated spread of Indo-European languages (cf. Section 2.2) Since the 2000 report by Semino and others, several other researchers have endorsed this model for explaining the current distributions of the R-M17 mutation. However, the literature also presents an alternate expansion scenario from two different regions, northwestern India and Poland, occurring several thousand years prior to the purported

Kurgan expansion (see Section 5.4 for a more detailed discussion of various expansion scenarios for the R-M17 mutation).

## 4.2 Haplogroup I.

Haplogroup I-M170 is found in about 20 percent of European men (Rootsi et al. 2004: 128). This haplogroup is confined almost exclusively to Europe, rarely found in Asia or Africa. Furthermore, haplogroup I is the only European-specific haplogroup among the six informative haplogroups found on this continent (Underhill 2007: 33). According to Semino and others, haplogroup I has been in Europe since the Paleolithic, entering the continent 20,000 to 25,000 years ago from the Middle East. Semino and others also suggest that the spread of the I-M170 mutation in Europe may have been linked to the Gravettian culture (Semino et al. 2000a: 1156). According to Underhill and others, the I-M170 mutation descended from the F-M89 mutation, the haplogroup representing the main out-of-Africa migration (Underhill et al. 2000: 54). In a more recent population report, Battaglia and others identified the IJ-M429 mutation as an intermediate evolutionary step between the out-of-Africa haplogroup group (F-M89) and the I-M170 Europeans (2009: 825-826). People having the IJ-M429 mutation entered Europe from present-day Turkey before the last Ice Age, about 20,000 years ago. The IJ-M429 mutation evolved into haplogroup I-M170 in Europe, whereas the IJ-M429 mutation in Turkey was the ancestral mutation that later became haplogroup J-M304. During the Neolithic, haplogroup J group became the genetic signature of farmers and the expansion of agriculture from the Middle East into Europe (cf. Section 4.5).

Four sub-haplogroups of I-M170 mutation have deepened our understanding of Mesolithic Europe. One of these sub-haplogroups, defined by the I-M26 mutation, was discussed by Semino and others in their 2000a report. However, the three other sub-haplogroups, defined by the I-M253, I-M423 and I-M223 mutations, were first described about four years later, published by Rootsi and others in their 2004 report. According to Rootsi and others, these four sub-haplogroups of M170 describe 95% of haplogroup I variation in Europe (2004: 129). This dissertation will label the I-M253 mutation as the Scandinavian I-Group, the I-M223 mutation as the Central European I-Group, the I-M423 mutation as the Balkan I-Group, and finally the I-M26 mutation as the Sardinian I-Group.

### 4.2.1 Scandinavian I-Group.

Please refer to Table 3 in the Appendix. The Scandinavian I-Group is defined by the I1-M253 haplogroup and was first reported by Rootsi and others in 2004. This sub-haplogroup of I-M170 attains its maximum frequency in Scandinavia, where it is found at moderate levels. Low frequency levels of the Scandinavian I group are reported for central and western Europe. Perhaps surprising, elevated frequencies of the haplogroup are found in the Baltic Region. Lappalainen and others suggest (2008: 343-345), based on short tandem repeat data and the archaeological record, that the Scandinavian I-Group arrived in the Baltic region during the Neolithic from Sweden and northern Germany.

Low levels of the Scandinavian I-Group are also reported for Eastern Europe, including Russia. Perhaps the presence of the I-M253 mutation in Russia can be explained as a further expansion of those carrying the I-M253 mutation into the Baltic region during the Neolithic. Kalevi Wiik, a linguist and professor at the University of Turku in Finland, has an alternative suggestion and proposes that the Vikings carried the I-M253 mutation into Russia (Wiik 2008: 65). Finally, low frequency levels of the Scandinavian I-Group are reported for the Balkans and Mediterranean.

While an expansion of the I-M253 mutation from Scandinavia may have occurred during the Neolithic, Rootsi and others suggest that individuals with the I-M253 mutation arrived in Scandinavia during the Mesolithic, migrating from an Iberian refuge area along the current Franco-Spanish border after the glacial ice had retreated from central and northern Europe (Rootsi et al. 2004: 129). This is consistent with data provided by Underhill and others, who provide an average age of 8,000 years for the I-M253 mutation (2007: 39). A similar age for the I-M253 mutation was also reported by Lappalainen and others in their 2008 report, which examined Y-chromosome variation in the Baltic Region, as well as Finland and Sweden (2008: 340).

### 4.2.2 Balkan I-Group.

The Balkan I-Group is currently defined by the I2a1-M423 haplogroup. Please refer to Table 4 in the Appendix. The I-M423 mutation is found at high frequency in Bosnia-Herzegovina, and at moderate levels in other Balkan nations. In central and eastern Europe, low to elevated frequencies of this mutation are found. Throughout the rest of Europe, low frequencies are reported. Bara and others suggest (2003: 540) that the Balkan I-Group expanded from a refuge area near the Adriatic Sea following the last Ice Age. However, the Balkan I-Group may have expanded out of the Balkans during the Neolithic. Battaglia and others propose that men with the I-M423 mutation were among the Mesolithic inhabitants of the Balkan region. The I-M423 men may have learned agriculture technology from migrant farmers from Turkey, who entered the Balkan region of Europe during the Neolithic. According to Battaglia and others (2009: 827), as agriculture expanded from the Balkans into other areas of Europe, the mutation I-M423 also expanded.

### 4.2.3 Sardinian I-Group.

The Sardinian I-Group is defined by the I2a2-M26 haplogroup. Please refer to Table 5 in the Appendix. This sub-haplogroup is rarely encountered in Europe with the exception of Sardinia, an island in the Mediterranean, as well as in certain isolated areas of the Pyrenees region, a mountain range along the Franco-Spanish border. Based on the presence of the I-M26 mutation in certain populations of the Pyrenees Region of Spain, López-Parra and others take the position that the Sardinian I-Group expanded southwards after the last Ice Age from an Iberian refuge area (2009: 50). The average age of I-M26 variation, about 13,000 years, supports this expansion scenario (Rootsi et al. 2004: 135).

Interestingly, the presence of the I-M26 mutation in Iberia has also been used to support the position that I-M253 also expanded from the same area following the last Ice Age (Francalacci and Sanna 2008: 65).

### 4.2.4 Central European I-Group.

Please refer to Table 6. The Central European I-Group is defined by the I2b1-M223 haplogroup. The frequency of the I-M223 mutation peaks in Germany and the Netherlands, and is present in about 10% of the men in both countries. Rootsi and others suggest that the Central European I-Group co-migrated with the Scandinavian I-Group from an Iberian refuge area following the last Ice Age. A Mesolithic expansion of the I-M223 mutation is supported by data from Underhill and others, who provide an average age about 12,000 years for this mutation (Underhill et al. 2007: 39).

## 4.3 Finno-Baltic N-Group.

The Finno-Baltic N-Group is currently defined by the N1c1-M178 haplogroup. Please refer to Table 7. The N-M178 mutation is found at a heavy frequency in Finland. Moderate frequencies of this mutation are reported for the Baltic region. Russians have an elevated frequency of this mutation. Elsewhere in the Europe the frequency of the N-M178 mutation is low or absent. Underhill and others in their 2001 report connect the evolutionary history of the Finno-Baltic N Group to the main Out-of-Africa migration about 45,000 years ago (2001: 53-55). In their 2007 report, Rootsi and others report additional information about the evolution history of the N-M178 mutation. According to the report (2007: 207), the ancestral N-M231 mutation diverged from NO-M214 in Southeast Asia about 35,000 years ago during the Paleolithic. However, the N-M178 mutation, a descendant of the N-M231 mutation, represents more recent population expansions beginning with the end of the last Ice Age.

The N-M178 mutation was first sequenced by Zerjal and others in 1997. They suggested (1997: 1179) that the M-178 mutation arose in China or Mongolia. Villems and others in 2002, on the other hand, suggested an alternate location, claiming the M178 mutation arose in the Volga Basin based on the elevated presence of this haplogroup in the ethnic Tartars, Udmurt and Chuvashi of this region (2002: 276-277). The study by Tambets and others in 2004 also supports the position taken by Villems and others. In their 2004 report, Tambets and others also used mitochondrial DNA evidence to link origins of the N-M178 mutation with the Volga-Ural region (2004: 676 - 678). Two studies in 2007 redirected the focus of M-178 origins back to Northern Asia. In their 2007 study, Derenko and others sought to clarify the origins of N-M178 mutation by attempting to resolve the controversy surrounding the age of this haplogroup. At the time, the Finno-Baltic N Group was identified in the nomenclature as N3a. Derenko and others stated that previous age determinations for N3a were skewed. The study posited that N3a had two informative sub-haplogroups, N3a1 and N3a2. N3a2 was not very common in Eastern Europeans, whereas N3a1 was far more common (Derenko et al. 2007: 764). Furthermore, N3a1 is much older than N3a2: 9,100 versus 5,000 years. Consequently, according to this study, the Finno-Baltic N group would have arisen in Northern Asia, because N3a1 is older in Southern Siberia than in Eastern Europe: 10,000 versus 8,200 years (Derenko et al. 2007: 768). However, the 2007 report by Derenko and others has become problematic because they failed to provide mutation numbers for N3a1 and N3a2. In 2008 the nomenclature for Y-chromosome haplogroups was refined and reorganized, and without the mutation number, I cannot convert N3a1 and N3a2 into the new nomenclature (Karafet et al. 2008). Once again, nomenclature changes are a serious obstacle for the non-geneticist.

In 2007 Rootsi and others also published a study advocating northern Asia as the source of the Finno-Baltic N Group. In their report they maintained that the N-M178 mutation arose in northern China based on dating results from this region. In their report, they also maintained that the N-M178 mutation appeared about 12,000 years ago, yet also urge caution with this estimate due to the small sample size and possible sampling error (Rootsi et al. 2007: 208). The 2007 study by Rootsi and others suggested that the N-M178 mutation expanded after the last Ice Age, from China in a counterclockwise direction, to Siberia, then along the Arctic Ocean and over the Ural mountains, to northeastern Europe, terminating in Scandinavia (Rootsi et al. 2007: 208).

The report by Mirabal and others in 2009 provided an alternative migration route. Rather than a counterclockwise migration, the N-M178 mutation may have spread clockwise, possibly from northern China, through the Volga basin and along the Caspian Sea, eventually reaching northeastern Europe. From northeastern Europe, the M178 mutation underwent another expansion, eastwards into Siberia and northwards into Scandinavia. This migration scenario is based on the date of M178 mutations found in Russian Slavic populations, dated at roughly 8,000-9,000 years, which drops to 5,500 years among the Komi people of northwestern Siberia (Mirabal et al. 2009: 1271).

Despite the controversy over the direction of migration, population reports have been consistent in maintaining that Scandinavia was at the terminal end of the N-M178 migration. Reports also agree that the N-M178 arrived relatively recently in Scandinavia. Semino and others suggest that men with the M178 mutation arrived in Scandinavia within the last 4,000 years (Semino et al. 2000a: 1158). Passarino and others suggested that men with the M178 mutation migrated to Scandinavia 5,000 years ago (Passarino et al. 2002: 524). Lappalainen and others in their 2006 report also seem to agree with this migration scenario, connecting the arrival of the M178 mutation with the spread of the arrival of the Comb Ceramic Culture in Scandinavia (Lappalainen et al. 2006: 213).

## 4.4. European E-Group.

The evolutionary history of haplogroup E is rather complex and important for understanding prehistoric migration and settlement in sub-Saharan Africa, northern Africa, the Middle East, and finally Europe (Luis et al. 2004: 541). In this dissertation, I will use the label "European E-Group" to describe how haplogroup E contributed the evolutionary history of Europe. The European E-Group is currently defined by the E1b1b1a2-V13 haplogroup. This haplogroup attains a moderate frequency in Greece as well as in several nations in the Balkans. From the Balkans, the frequency of the European E-Group declines both eastwards and northwards (cf. Table 3 in the Appendix). The label "European E Group" allows me to utilize data that only reports for the ancestral markers of the E-V13 mutation, E-M35 and E-M78, which is common. For the Mediterranean Region, I found it necessary to define the European E-Group strictly with the E1b1b1a2-V13 mutation, to distinguish it from other variants of the E-M78 mutation, found in the same regions (E-V12, E-V22 and E-V65), that have a different demographic history. However, for the rest of Europe I sometimes represent the European E-Group with the ancestral E-M78 and E-M35 mutations. I assumed that the E-M78 marker would have been the E-V13 marker if further tested because the E-V12, E-V22 and E-V65 markers are virtually non-existent outside of Mediterranean Europe and Iberia (Cruciani et al. 2007: 1302). I also assumed that E-M35 would have been E-V13 marker, because the E-M81 mutation is virtually absent in Europe except in Iberia and Sicily (Semino et al. 2004: 1025). Thus, I believe that V13 data in the Mediterranean Region and Iberia, as well as E-M78 and E-M35 data elsewhere in Europe, represent the same population expansion. Figure 4.2 (below) may be helpful in understanding these complex phylogenetic relationships.

**Figure 4.2 The Evolutionary History of Haplogroup E.**

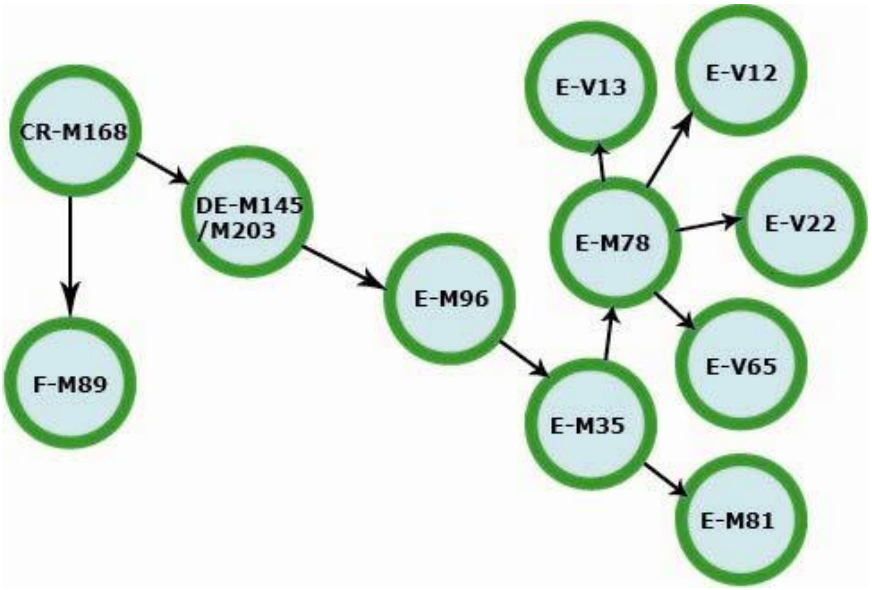


Figure 4.2 (above) provides an evolutionary overview of haplogroup E. Each circle represents a population. Within each circle is the genetic Y-chromosome mutational signature of the population. The arrows represent the emergence of a group from an ancestral population. The reader is asked to locate the population with the CR-M168 mutation. This population represents human origins in Africa. From this group emerged a new population defined by the F-M89 mutation. The F-M89 represents the main out-of Africa migration. As reflected by Figures 4.1 and 4.2, unlike the other Y-chromosome haplogroups discussed in this dissertation, haplogroup E did not participate in the main out-of-Africa migration about 45,000 years ago. Instead, most European men with haplogroup E chromosomes are the descendants of a more recent out of Africa migration that occurred at the beginning of the Mesolithic (Underhill et al. 2001: 50-51). (The reader may wish to locate the "human origins in Africa" group in Figure 4.2 and follow the arrow to the population labeled DE- M145/M203.)

The E-M35 mutation, an ancestral population for the E-V13 mutation, appeared somewhere in east Africa about 25,000 years ago (Cruciani et al. 2004: 1015). The E-M78 mutation, a descendant of E-M35, appeared in northeastern Africa about 17,000 years ago (Cruciani et al. 2007: 1305). The E-M78 mutation eventually dispersed over a wide area, and today is found in 21.5% of Eastern Africans, 18.5% of Northern Africans, 5.8% of Near Easterners, and 7.2% of Europeans (Cruciani et al. 2004: 1015). The E-V13 mutation further defines the evolutionary history of the E-M78 mutation in European men. This mutation is also the most common E-M78 variant in Europe, accounting for about 85% of the E haplogroup variation on this continent (Cruciani 2007: 1307). According to Cruciani and others in 2007, the E-V13 mutation arose in the Middle East about 11,000 years ago (1307). The E-V13 mutation may have evolved in a group of men having the M78 mutation, who migrated out of Africa from a refuge area on the current Sudanese-Egyptian border. According to Battaglia and others (2009: 827), this expansion occurred during the Mesolithic following improved climatic conditions.

In 2000, Semino and others initially reported that haplogroup E, along with haplogroups J and G, were part of the Neolithic expansion from Anatolia (2000a: 1157). However, geneticists now seem to agree that men with the E-V13 mutation entered Europe via the Balkans during the Mesolithic, and thus the E-V13 mutation was part of the pre-Neolithic genetic inventory of this continent (Cruciani et al. 2007: 1307; King et al. 2008: 212; Battaglia et al 2009: 827). Nevertheless, geneticists favor a Neolithic expansion of E-V13 mutation throughout Europe, which implies that men with the V-13 mutation learned how to farm and moved into the Mediterranean Region and Central Europe in search of land (Underhill et al. 2001: 50-51; Semino et al. 2004: 1032; King et al. 2008: 211; Battaglia et al. 2009: 828). Cruciani and others offer a different expansion scenario, taking the position that E-V13 expanded much later, during the Balkan Bronze Age, following rivers from southern Balkans into central Europe (2007: 1308). King and others dispute the Bronze Age expansion, asserting that Cruciani and others incorrectly estimated the age of E-V13 mutation (2008: 211).

As stated earlier, haplogroup E is found over a vast geographical expanse, and consequently tells the story of several population expansions. Also, as stated earlier, E-M78 is the ancestral mutation of E-V13. The ancestral mutation of E-M78 is E-M35. Another descendant of E-M35 is E-M81. The E-M81 mutation is commonly found in northwest Africa, especially prevalent among the Berbers (Cruciani et al. 2004: 1018). Discussions of haplogroup E-M81 variation in Europe often attempt to determine the extent of North African contribution to the gene pool in Iberia, in the various Mediterranean islands, as well as in the Italian and Greek gene pools. For example, based on E-M81 data, Bosch and others (2001: 1027) estimated the North African contribution in the Iberian gene pool to be about 5%. Furthermore, according to Bosch and others (2001: 1020), part of the North African contribution to the Iberian gene pool may have occurred as a result of the Arab seizure of the Iberian Peninsula in the year 711. Besides the Iberian Peninsula, men with the E-M81 mutation also migrated to the Mediterranean island of Sicily. According to Di Gaetano and others (2009: 98) the E-M81 mutation reflects a North African contribution of about 6% to the Sicilian gene pool.

The search for a North African genetic contribution in Europe is also provided by E-V12, E-V22 and E-V65, which are descendants of the E-M78 mutation. According to Cruciani and others, E-V12, E-V22 and E-V65, along with E-M81, represent a North African contribution of 5.6% in the Iberian gene pool, 3.6% in the Italian gene pool, and 6.6% in the Sicilian gene pools. Cruciani and others also assert that E-V12, E-V22 and E-V65 did not enter Europe through the Balkans. Instead, E-V12, E-V22 and E-V65 entered Europe via sea migration over the Mediterranean at a much later period in the prehistory, beginning perhaps 4,000 years ago or sooner (Cruciani et al. 2007: 1307).

The above discussion of haplogroup E variation, and sea versus land migration, demonstrates that haplogroup E data presents differing demographic scenarios in Europe. For four very good reasons, this dissertation will focus only on the E-V13 mutation. First, E-V13 represents about 85% of European E-M78 chromosomes (Cruciani et al. 2007: 1307). Secondly, as explained earlier, the E-M81 mutation is virtually absent in Europe except in Iberia and Sicily. Thirdly, as detailed earlier, like the E-M81 mutation, E-V12, E-V22 and E-V65 are also virtually absent outside the Mediterranean and Iberia. E-V13, on the other hand, is found from the Mediterranean to the Baltic Sea. Finally, E-V13 may have expanded across Europe carried by the first farmers of this continent, a population expansion with far greater pan-European importance than relatively recent sea migrations across the Mediterranean.

### 4.5 Near Eastern J-Group.

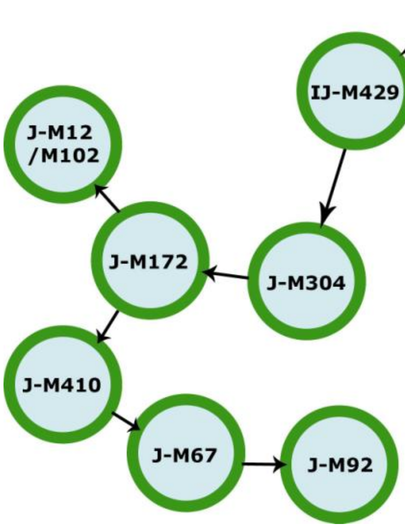
Haplogroup J has a wide distribution, straddling the northern and southern Mediterranean coastline, spreading into the Middle East and terminating in India. The most common variant of haplogroup J in Europe is the J2-M172 haplogroup. This haplogroup also defines the Near Eastern J-Group. The J-M172 mutation attains low to moderate frequencies in Eastern Europe and the Balkans. It is found at moderate levels in the Mediterranean. Elsewhere in Europe, haplogroup J2-M172 represents only a low percentage of the gene pool. In the Middle East and the Caucasus, the J-M172 mutation is found at a moderate frequency. At the terminal eastern end of worldwide haplogroup J distribution, in India, J2-M172 is found at elevated frequencies. For additional information about J-M172 frequencies, please refer to Table 9 in the Appendix.

Please refer to the overview of haplogroup J as provided in Table 4.3 below. Each circle represents a population. Within each circle is the genetic Y-chromosome mutational signature of the population. The arrows represent the emergence of a group from an ancestral population. Haplogroup J is currently defined by the J-M304 mutation. This haplogroup descended from F-M89 and was part of the main out-of Africa migration (Underhill et al. 2001: 53-55). Data from Cinnioğlu and others in their 2004 report suggest that J-M304 separated from haplogroup IJ-M429 somewhere in the Middle East about 20,000 years ago. The same report suggests that roughly 18,000 years ago, J1-M267 and J2-M172 split from J-M304 (2004: 131). J1-M267 peaks in the Arabian Peninsula, decreasing beyond the Middle East and North Africa (El-Sibai et al. 2009: 574). Early debate surrounding the expansion history of the J1-M267 mutation attributed the distribution of this haplogroup to the expansion of Arabian tribes during recorded history (Nebel et al. 2002: 1595). More recent analysis of the J1-M267 haplotype has rejected this position, and maintains that a much earlier post-Ice Age expansion of pastoralists represents a better explanation for the current distribution of this haplogroup (El-Sibai et al. 2009: 574; Tofanelli et al. 2009: 1523-1524).

The highest frequency of the J-M172 mutation is found in Lebanon, decreasing towards the west in North Africa, and east in Arabia and India (El-Sibai et al. 2009: 574). According to Cinnioğlu and others, J-M172 expanded eastwards across Anatolia from the ancient farming settlement at Çatalhöyük in southwestern Turkey (2004: 133). The presence of J2-M172 in Europe is attributed to a further eastward expansion of these early farmers during the European Neolithic, beginning about 8,500 years ago (Semino et al. 2000a: 1157). The M172 mutation also expanded eastward into Iran, Pakistan and India, possibly with the spread of agriculture to this region of the world (Quintana-Murci et al. 2001: 541).

J2-M172 has two major subclades, J2a-M410 and J2b-M12/M102. Data from Cinnioğlu and others suggest that this split occurred roughly 12,000 years ago in the Middle East (Cinnioğlu et al. 2004: 131). Reports concerning haplogroup J2-M172 variation in Europe often attempt to find sub-haplogroups of J2a-M410 and J2b-M12/M102 to further explain local population histories. For example, Di Giacomo and others maintains that J2a2a-M92 represents a Bronze Age expansion in ancient Greece (2004: 367). In another example, King and others argue that J2a1h-M319 represents a Bronze Age expansion to Crete from mainland Greece (King et al. 2008: 210-211). The discussion of local population histories is somewhat confusing in that one is not sure if J2a and J2b are ultimately part of a larger population expansion, the Neolithic expansion of farmers from the Middle East to Europe. I will take the position that J2a and J2b are part of the same population expansion.

**Figure 4.3 The Evolutionary History of Haplogroup J.**



Please refer to Table 10 in the Appendix, which presents, to the extent possible with the available and limited data, the distribution of J2a-M410 and J2b-M12/M102. In Turkey, the purported source of J2 in Europe, J2a represents about 23% of the gene pool, and about 93% of J2 lineages. J2b, on the other hand, represents less than 2% of the gene pool, and about 7% of J2 lineages. However, in the Balkans, J2b represents 5.45% of the gene pool, and an astonishing 60% of all J2 lineages. Due the low presence of J2b in Turkey, King and other suggest that the source of J2b in the Balkans may have arisen further east in Syria (King et al. 2008: 210). Interestingly, J2a represents about 40% of the J2-M172 variation in the Balkans, and about 60% of the variation in Greece. J2b, on the other hand, represents 40% of the J2 variation in Greece. Arguably, based on these statistics, J2a and J2b are almost equally represented in Greece and in the Balkans. However, J2a is the overwhelming J2-M172 variant in Italy, as well as on the islands of Crete and Sicily. However, Capelli and others take the position that haplogroup J variation in Italy, and the J2a sub-haplogroup, also stems from a migration of farmers from Anatolia during the Neolithic (Capelli et al. 2007: 237). Semino and others support this position, claiming that J2a2-M67 and J2a2a-M92 came to Europe from Anatolia as part of the expansion of agriculture from this area (Semino et al. 2004: 1027-1030). Battaglia and others also took a similar position in 2009, maintaining that J2a2a-M92 distribution in the Balkans is consistent with a westward Neolithic expansion of farmers from the Middle East (826.). I realize this is confusing. However, I believe I have justified my decision for lumping all the European variants of the J-M172 mutation in the same population expansion, something I was not able to do for all the variants of the E-M78 mutation.

Another point worthy of clarification is the position, taken by some geneticists, that the J-M172 mutation arrived in Europe by sea while others advocate an overland route via the southern shore of the Black Sea (Semino et al. 2004: 1027-1030; King et al. 2008: 211; Battaglia et al. 2009: 826). I would argue that the picture of the J-M172 expansion into Europe is currently incomplete because little has been reported about haplogroup J diversity in the Ukraine, Romania, and in Bulgaria. As illustrated by Table 9, geneticists report that haplogroup J has contributed between 15 and 20% to the gene pool of Romania and Bulgaria, and the Ukraine has the highest frequency of haplogroup J in Eastern Europe. The lack of information for this area is disturbing because the archaeological record also offers a compelling reason for examination of a northern Black Sea route for haplogroup J. Current archaeological debate surrounding the European Neolithic includes a purported massive flood around the shoreline of the Black Sea roughly at the beginning of the European Neolithic, about nine thousand years ago. An earthen dam at the current Bosporus Straights may have broken, flooding the Black Sea region with water from the Mediterranean Sea. This flood may have driven farmers from the Middle East into Europe. For example, in a 2007 discussion of the Black Sea flood, Douglass W. Bailey, a professor of European Prehistory at the University of Cardiff in Wales, describes the evidence of pre-flood human settlement in southeast Europe as “patchy” at best (2007: 518.). He then describes human settlement post 6,000 BC in the Balkans as “abundant” (2007:521). While Bailey seems to avoid endorsing the flood as a trigger of human migration, he does endorse the need for further inquiry (2007: 527). Here, I endorse further inquiry into the Ukrainian, Bulgarian and Romanian gene pools, maintaining that the exploration of a north Black Sea route for haplogroup J would be an important clarification for geneticists, archaeologists, and perhaps even linguists.

### 4.6 Caucasus G-Group.

Please refer to Table 11 in the Appendix for the distribution of haplogroup G in Europe, the Middle East and South-Central Asia. Haplogroup G-M201 currently defines the Caucasus G-Group. Haplogroup G appears in moderate frequencies in the Caucasus, and elsewhere the frequency is low to elevated. Unlike the other haplogroups discussed so far, haplogroup G fails to establish a clinal pattern across a geographical range, but rather this haplogroup appears intermittently in isolated regions. Thus, elevated levels of haplogroup are often explained as a phenomenon of genetic drift, such as on the island of Sardinia, where the unique I-M26 mutation is also found (Zei et al. 2003: 805).

Similarly, although haplogroup G appears in very low frequencies in mainland Croatia, the frequency is elevated on certain Croatian islands, possibly as the result of genetic drift (Barać et al. 2003: 539-540). The Caucasus also yields an intermittent pattern of haplogroup G variation. For example, haplogroup G is found 74% in the North Ossetian ethnic group, yet is absent in Kazbegi, Lezgi and South Ossetian ethnic groups (Nasidze et al. 2004a: 213).

Semino and others in their 2000a report attributed the presence of haplogroup G in Europe, as well as haplogroups E and J, to a Neolithic expansion of agriculturalists from modern-day Turkey (2000a: 1157). Underhill and others also agreed with this expansion scenario in their 2001 report. In the same report, they maintained that men with the F-M89 mutation represent the main out-of-Africa migration about 45,000 years ago, and that haplogroup G evolved from haplogroup F-M89 somewhere in the Middle East (2001: 53-55). In their 2004 report, Cinnioğlu and others maintained (2004: 133) that haplogroup G was present in Turkey during the Paleolithic, and expanded into the Caucasus, and into Europe, during the Neolithic. However, in their 2008 report on Y-chromosome variation on the island of Sardinia, Contu and others date the G-M201 mutation in Sardinian men to around 20,000 years ago. They asserted that haplogroup G was part of genetic inventory in Paleolithic Europe, and not part of a Neolithic expansion of farmers (2008: 5-6).

In the previous section on haplogroup J (Section 4.5), this dissertation suggested that a further examination of Y-chromosome variation in the Ukraine, Romania and perhaps Bulgaria would reveal a second or alternative route of agricultural expansion into Europe. The presence of haplogroup G in the Ukraine (about 4%) and Romania (about 10%) would also, in my opinion, further support the need for further examination of genetic evidence of the Neolithic migration along the north shore of the Black Sea.

## 4.7 Chapter Conclusion.

In the two-hundred plus population reports detailing the evolutionary history of Europe from a Y-chromosome perspective, the data consist of complex evolutionary relationships identified by a nomenclature system that is constantly revised and refined. I have attempted to simplify the data by identifying ten population expansions that I believe are definitive of the European prehistory and representative of the Y-chromosome data. The population expansions identified in this dissertation are the following:

- Western European R-Group (R1b1b2-M269)
- Eastern European R-Group (R1a1a-M17)
- Scandinavian I-Group (I1-M253)
- Balkan I-Group (I2a2-M423)
- Sardinian I-Group (I2a1-M26)
- Central European I-Group (I2b1-M223)
- European E-Group (E1b1b1a2-V13)

- Near Eastern J-Group (J2-M172)
- Caucasus G-Group (G-M201)
- Finno-Baltic N-Group (N1c1-M178)

These groups arrived in Europe and expanded during different eras of the European prehistory. The four different I-Groups, and perhaps the Caucasus G-Group, represent the descendants of people who arrived in Europe during the Paleolithic, around 30,000 years ago. Following the last Ice Age, about ten to twelve thousand years ago, the I-Groups expanded from refuge areas, occupying areas that were previously uninhabitable. The distribution of the Caucasus G-Group is more complex, perhaps a remnant of Paleolithic Europeans never expanded from the Ice Age refuge areas. The European E-Group represents men from Africa who probably arrived in Europe during the Mesolithic, perhaps ten to twelve thousand years ago. The Near Eastern J-Group represents the Neolithic contribution to the European gene pool, carried farmers who left what is now Turkey. While the Western and Eastern R-Groups may have been in Europe prior to the arrival of agriculture, their current distribution on this continent is thought to have a result of expansion of farming during the Neolithic. Finally, the Finno-Baltic N Group represents men from Siberia who eventually settled in Scandinavia about 4,000 years ago.

Turning now to the next chapter, the dissertation will now focus on language diversity in Europe from a Y-chromosome perspective. Perhaps some may find it surprising that geneticists have used linguistic data to explain genetic data. Perhaps some may also find it surprising that I am attempting to overcome the problems of fragmented reporting and ever-changing nomenclature so linguists can use the genetic data to explain the linguistic data. However, I would argue that cross-disciplinary cooperation between geneticists and linguistics has long-standing historical precedent. The development of genetic theory, the theoretical basis for these population studies, began in 1859 with the publication of *On the Origins of Species* by Charles Darwin, a book that sought to explain genetic variation found in the natural world. In 1863, August Schleicher, a giant in the field of Germanic Linguistics, published an open letter to a professor at a museum in Jena, Germany. In the letter, Schleicher stressed that linguistics and Darwinian theory represented complementary methodologies. One idea, that surfaced repeatedly, is taxonomic relationships, that over time languages and organisms evolve from a common ancestor.

# Chapter Five
# The Correlation Between Linguistic and Genetic Diversity: a Survey of Population Studies.

## 5.0 Chapter Introduction.

The conclusion of the previous chapter referred to August Schleicher and his belief that evolutionary theory and linguistic theory are united by a common methodology. In the twenty-first century, human evolutionary research presents a new opportunity for interdisciplinary cooperation between linguists and geneticists. In 2003, the renowned geneticist Cavalli-Sforza wrote that molecular genetics and linguists, along with archaeology, anthropology, and demography, are “complementary approaches” for reconstructing human evolution (266). Perhaps then it should not be surprising that after mining the population reports discussing Y-chromosome variation, I found a treasure trove of information that could be useful for linguists.

It may appear that this chapter deviates too far from the central goal of this dissertation, which is to discuss contemporary models of Germanic origins. I ask for an indulgence from the reader. Because I believe this dissertation sails into uncharted waters by considering Y-chromosome research, I wrote this chapter to demonstrate to my fellow linguists that molecular genetics is a useful perspective for our discipline. Thus, among my goals in this chapter is to create precedent. Another goal was to present and preserve data that may be useful for other researchers. Surprisingly, I had to cast a net over a broad region, from the Arctic Circle to the equator, and from Iceland to India, in order to obtain a Y-chromosome perspective of Germanic origins.

## 5.1 Africa

In a recent report examining genetic variation in Africa from the perspective of Y-chromosomal, mitochondrial, autosomal and classical markers, two researchers found a remarkable correspondence between linguistic affiliation and genetic diversity in Africa (Campbell and Tishkoff 2010: R168). From a Y-chromosome perspective, Africa provides one of the best examples of a link between genetic and linguistic diversity, the E1b1a-M2 haplogroup, a signature of the Bantu farmers and their language, a dispersal occurring over the past 4,000 years (Richards et al. 2006: 238-339). In the report by Woods and other in 2005, researchers examined mitochondrial DNA and Y-chromosome variation among the Bantus. They suggest that as the Bantus migrated, they replaced existing languages and reduced Y-chromosome diversity because they assimilated Khosian and Pygmy women, while the Khosian and Pygmy men underwent a corresponding loss of reproductive success (2005: 874).

Y-chromosome evidence from speakers of click language may indicate that human beings developed language before the human expansion out of Africa to Asia and beyond.  Clicks, a rare type of consonant found in the human phonetic inventory, are found in some of the languages of sub-Saharan Africa.  A report by Knight and others in 2003 compares Y-chromosome data obtained from the San, a click-speaking people in Botswana, with data from the Hadzabe, a click-speaking people in Tanzania. Both groups share the B2b-M112 haplogroup, one of the oldest Y-chromosome lineages, one that evolved before our species left Africa.  Along with Y-chromosome data, the Knight report also considers maternally inherited mitochondrial DNA data and examines short tandem repeat mutational variation between the two click-speaking populations.  In their report (440-471), the researchers first conclude that clicks are not an independent innovation or a result of language contact among both groups, but rather a relic of their common history.  Based on the genetic evidence, the researchers determined that the San and Hadzabe separated about 40,000 years ago.  Both groups currently live over two thousand kilometers apart from each other, and for this reason, according to the Knight report, language contact remains an unlikely explanation for the presence of click consonants in both groups.  Also, according to the Knight report, since clicks are a rare speech sound, it is unlikely that clicks represent an independent innovation in both groups.  Knight and others then propose that clicks may be among the oldest phonemes in human language.  According to the report, clicks may have also improved hunting success for prehistoric Africans, and thus delivered an evolutionary advantage for these people.

### 5.2 The Role of Gender in Mediating Language Shift.

Another interesting use of genetic data, that has emerged from populations studies focusing on genetic variation in Africa, is the role of gender in mediating language shift in some populations.  In their 2005 report,  Woods and others found a “statistically significant” correlation between linguistic affiliation and Y-chromosome variation (2005: 873-874).  However, the mitochondrial data showed little correlation between genetic variation and language.  Based on the strength of the relationship between Y-chromosome variation and linguistic affiliation, Woods and other concluded that “African languages tend to be passed from father to children.”  Using the Bantu expansion as an example, the study suggests that as this group migrated, men from other populations, such as Khosians and Pygmies, were not assimilated into the Bantu group. However, the Bantu men took Khosian and Pygmy women as wives, who eventually adopted the Bantu language and passed this language to their children.  Thus, the shift to Bantu languages in Africa was mediated by men.

A study from Iran reports linguistic shift mediated by women.  The report examines the Gilaki and Mazandarani, two populations residing in the South Caspian region of Iran.  The Y-chromosome data suggest both groups migrated to their present location from the South Caucasus region.  However, the mitochondrial DNA data suggest a closer relationship between the Gilaki and Mazandarani and their geographical neighbors in Iran.  The linguistic evidence also indicates a closer relationship in that the Gilaki and Mazandarani speak an Indo-European language from Northwestern Branch of the Iranian languages. Nasidze and others suggest, based on the Y-chromosome data, that the Gilaki and Mazandarani have their origins in the Caucasus, and after arriving in Iran, both populations never assimilated with the Iranian men. The mitochondrial DNA evidence suggests, on the other hand, that men in both populations married Iranian women, who in turn mediated a shift in language from Caucasian to Indo-European (Nasidze et al. 2006a: 671).

## 5.3 Afroasiatic.

Y-chromosome variation in North Africa may provide linguists important information surrounding the origins of Afroasiatic languages. Among the modern languages classified as Afroasiatic is Arabic. Haplogroup E represents part of the genetic inventory of Europe during the European Mesolithic about ten to twelve thousand years ago (cf. Section 4.4 for additional details). Arredi and others (2004: 343) suggest haplogroup E-M35 lineages, that entered the Middle East and Europe during the Mesolithic, also back-migrated from the Middle East to North Africa during the Neolithic, about 8,000 years ago or later. In their report, they suggest this back migration may have been farmers who spoke proto-Afroasiatic languages. For the linguist, this raises an interesting question, whether men with the E-V13 mutation (European E-Group) also spoke proto-Afroasiatic when they migrated into Europe during the Mesolithic. Was Proto-Afroasiatic part of the linguistic inventory in prehistoric Europe?

## 5.4 Indo-European Language Origins.

One important question among the linguists surrounds the origin and expansion of Indo-European languages in Europe. In Chapter Two, I detailed two competing models of Indo-European origins in Europe (cf. Section 2.2). The model proposed by the archaeologist Marija Gimbutas attributes the Kurgan expansion from the Russian steppes, during the Bronze Age (about 3,000 years ago) as the mechanism for spreading Indo-European languages. This model of Indo-European origins has been embraced by several geneticists. In their 2000 report, Semino and others, referring to the R-M17 mutation (Eastern European R-Group), wrote the following: “Its [R-M17] spread may have been magnified by the expansion of the Yamnaia culture from the ‘Kurgan culture’ area (present-day southern Ukraine) into Europe and eastward, resulting in the spread of Indo-European languages” (2000a: 1156). Since 2000, this expansion model for R-M17 mutation has been cited numerous times in population reports (Wells et al. 2001: 10248; Zerjal et al. 2002: 477; Peričić et al. 2005a: 509; Lappalainen et al. 2006: 213; Mirabal et al. 2009: 1270). Mirabal and others even claimed in their recent 2010 report that “most investigators agree that the haplogroup [R-M17] arose in the Central Asian steppes and marks the migration of the Kurgan horse culture … thought to have divulged [sic] the Indo-European languages” (2010: 386). Nevertheless, data from India, as well as a 2010 report by Underhill and others, may undermine this position, that the R-M17 mutation spread Indo-European languages.

Focusing now on India, this south-central Asian country is home to over a billion people. This area of the world is not only culturally diverse, but also genetically diverse. India also lies on the eastern periphery of the spread of Indo-European languages. A legacy of this spread is Hindi, an Indo-European language spoken today by 422 million Indians.[3] Because India lies on the eastern periphery of the spread of Indo-European languages, a putative Indo-European homeland could be identified, in my opinion, by a population expansion shared by Europe, the Middle East and India.

In 2001, Quintana-Murci and others published a study examining the source of the R-M17 mutations found in Iran, Pakistan and India. The study concluded that the source of the R-M17 mutations was central Asia, possibly a signature of the Kurgan expansion (2001: 539-541). However, in 2006 Sengupta and others found, based on Y-chromosome data, that central Asia was not the source of the R-M17 mutation in India, thus eliminating the Kurgan expansion as the source of this haplogroup. Rather, the investigators found (2006: 218) that the R-M17 mutation expanded from the Indus Valley, in northwestern India, during the early Holocene (roughly 14,000 years ago), undermining the Kurgan hypothesis both in terms of geography and timing.

The Kurgan theory is also undermined by data provided by Underhill and others in their 2010 report. This report refused to speculate on the source of the R-M17 mutation because additional downstream markers are still needed to make this determination. However, this report distinguishes Asian and Eastern European R-M17 mutations with a downstream haplogroup, R1a1a7-M458, which is virtually absent in Asia, yet the most common R-M17 variant in Central and Eastern Europe (2010: 480). According to Underhill and others, the oldest R-M458 mutations are found in Poland, dating to about 11,000 years ago. European R-M17 mutations without the R-M458 marker also have a similar age (480-481). Consequently, R-M17 variation in Europe, like in India, also undermines the purported Kurgan model of Indo-European expansion from the perspectives of time and location.

The alternative model for explaining the spread of Indo-European languages is the language farming theory proposed by the British archaeologist Colin Renfrew in his 1987 book *Archaeology and Language: the Puzzle of Indo-European Origins* (cf. Section 2.2). Renfrew posits that the first farmers of Europe came from Anatolia (present-day Turkey). According to Renfrew, over a period of about 3,500 years, beginning about 6500 BC, they spread their new technology throughout the continent, beginning in the Balkans, and terminating in Northern Europe. Using this population expansion, in his 1987 book Renfrew proposed that farmers from Anatolia also disseminated the Indo-European language family throughout the European continent. Haplogroup J has been identified as the genetic signature of a westward expansion of agriculture from the Middle East to Europe (cf. Section 4.5) Did an eastward expansion of agriculturalists having the J-M172 mutation also bring Indo-European languages from the Middle East to India?

---

[3] Data from the 2001 Indian census. http://censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/Statement1.htm

In their 2001 report, Quintana-Murci found that J-M172 mutations in India have a source in the Middle East and followed the advance of agriculture from this region (539-541). In 2004, Cordaux and others published a study focusing on Y-chromosome variation among the castes and tribes of India. This is linguistically significant as speakers of Indo-European languages tend to belong to castes, whereas speakers of non-Indo-European languages, such as Dravidian, belong to tribes (e.g. Sahoo 2006: 843). The 2004 Cordaux study found that the second most common Y-chromosome mutation in caste populations is the J-M172 haplogroup (Near Eastern J-Group), with a reported frequency of about twelve percent. For tribes, just three percent of men have this mutation (2004b: 232). A study by Sengupta and others in 2006 also considered J2-M172 variation in India. They report that the J-M172 mutation is present in about nine percent of the men in India, and twelve percent of the men in Pakistan (206). However, the report (215-217) declined to endorse the J-M172 mutation as a marker of agricultural expansion from the Middle East because the date of the J-M172 mutation in India was estimated at 14,000 years, much earlier than the arrival of agriculture in this region. However, the study left also emphasized that additional genetic testing and the identification of additional markers within the J2-M172 haplogroup may eventually demonstrate a Near Eastern and Neolithic origin for the J-M172 mutation in India.

## 5.5 Hungarian.

Hungarians, like Estonians, Finns and Saami, speak a Uralic language. However, other historical reasons account for the presence of Uralic in central Europe. The Magyars, a Uralic-speaking people from Central Asia, invaded and settled in present-day Hungary in the ninth century. While the Magyars contributed to the linguistic legacy of present-day Hungary, their contribution to the gene pool of this area remained unresolved until recently. In their 2000 report, Semino and others considered both mitochondrial DNA and Y-chromosome data, and concluded that the Magyar invasion contributed little to Hungarian gene pool, but rather the Hungarian gene pool is composed of European haplogroups (Semino et al 2000b: 344). However, their report is difficult to decipher as it was published before standardization of the Y-chromosome nomenclature. A more recent population study of the Hungarian gene pool confirmed the European ancestry of this population, finding an absence of Asian Y-chromosome markers, such as haplogroups C, N and O. However, about 7% of Hungarians have the H-M82 mutation, a signature of the Roma (Gypsy) minority in Hungary, a people whose origins are found in India (Völgyi et al. 2008: 384-385).

Y-chromosome data has also been used by geneticists in an attempt to pinpoint the geographic origins of Hungarian. A study by Biró and others in 2009 examined the genetic relatedness between Magyars (Hungarians), and the Madjars, a central Asian population in Kazakhstan, about 2,000 miles away, with a strikingly similar name. According to the Biró study, haplogroup data for both groups is vastly dissimilar. Haplogroup G is present in about 87% of Madjar men, whereas less than 5% of Hungarians have this mutation (2009: 307). Nevertheless, the Biró study concluded (2009: 309), based on short tandem repeat data, that Madjars are genetically closest to

Hungarians than any other group. Based on this conclusion, Biró and others report that they may have found the homeland of the Magyars or Huns. The Biró argument is built on "the statistical significance" of genetic data which I, as a linguist, am not able to really evaluate. However, I do not believe "statistical significance" is necessarily a persuasive argument, and that a glaring deficiency of the Biró report is that they do not discuss the language of the Madjars (Kazakhstan), nor is there any expert opinion as to whether the terms "Magyar" and "Madjar" are indeed etymologically related. The Biró report, in my opinion, demonstrates that the interpretation of genetic data often requires the cooperation of several disciplines, including linguistics.

### 5.6 Slavic and Uralic.

Noel C. Brackney's 2007 book *The Origins of Slavonic, Language Contact and Language Change* represents a recent effort to explain the division of the first Slavic language, Common Slavic, into East, West and South Slavic, which he dates to the sixth century (2007: 18). His approach not only considers linguistic evidence and theory, but also the political and social history of Slavic peoples (2007: 50). Adopting Renfrew's language farming theory, Brackney proposed that the first farmers of Europe were also the ancestors of the Slavic people (2007: 91). He then proposed that they settled north of the Carpathian Mountains in Central and Eastern Europe during the Neolithic. Brackney finally asserts that the Slavic peoples remained in relative isolation in this area until the fifth century, when the socio-political situation climate in Europe changes, creating the opportunity for this people to assert their culture and language onto the historical stage (2007: 91-99).

Brackney's work is cited above because it suggests that the distribution of Slavic languages may not have resulted from a large scale migration similar to the Bantu expansion in Africa (cf. Section 5.1). Rather, Brackney's work leaves open the possibility that language shift explains the current distribution of Slavic, a scenario that is more consistent with the Y-chromosome data. For example, the N-M178 mutation (Finno-Baltic N-Group) is also a common haplogroup among Slavic-speaking men in Belarus (9%), the Ukraine (10%) and the European part of Russia (14%). However, the same mutation is rare in other Slavic-speaking populations (cf. Appendix Table 7). Moreover, the absence or presence of the N-M178 mutation separates the Eastern Slavs from Western and Southern Slavs (Peričić et al. 2005b: 1974). Consequently, the presence of the N-M178 mutation among the Eastern Slavs may provide an example of language shift among Uralic-speaking men, meaning these people stopped speaking Uralic, a non-Indo-European language, and started speaking Russian, Belarusian, or Ukrainian, all of which are Indo-European Slavic languages. Arguably, this shift occurred without a massive infusion of new Y-chromosome haplogroups into the Belarusian, Ukrainian and Russian gene pools. One possible scenario, taken from the perspective of language contact theory, is that Slavic may have been the language of a small population that was ultimately adopted by an unrelated larger group for socio-linguistic reasons, perhaps because Slavic was perceived as a prestige language. Such a possibility was left open by Balanovsky and others in their 2008 report. The researchers claimed that such a scenario is possible because haplogroup N expanded during the Mesolithic, whereas the Slavic tribes expanded only about a 1,000 years ago, thus possibly leaving the gene pool of the groups they encountered unaltered (Balanovsky et al. 2008: 242-246). In my opinion, this scenario is supported by the I-M423 mutation, the genetic signature of the Balkan I-Group, and the R-M17 mutation, the genetic signature of the Eastern European R-Group. Like the N-178 mutation, the R-M17 and I-M423 mutations also depict migration and settlement that predate the expansion of the Slavic tribes. This illustrates the difficulty in measuring the Slavic expansion with genetic data as it occurred fairly recently.

The expansion of Slavic without a corresponding large population expansion is further supported by a 2002 study that examined the Baltic countries, Lithuania, Latvia and Estonia. In all three countries, about one third of the men have the N-M178 mutation (Appendix Table 7). The N-M178 mutation, which I labeled as the Finno-Baltic N-Group, was discussed in Section 4.3 above. This mutation is regarded by many researchers as the signature of a prehistoric migration of Uralic-speaking men from Siberia into the Baltic Region and eventually into Finland and northern Scandinavia (e.g. Depuy et al. 2006: 17; Lappalainen et al. 2006: 213). Lappalainen and others suggests (2006: 213) that the Comb Ware Culture contributed the M-178 mutation and Uralic languages to Finland and Scandinavia about 4,000 to 5,000 years ago.

Latvians and Lithuanians now speak an Indo-European language, whereas Estonians still speak a Uralic language. In 2001, Zerjal and others published a report stating that Estonians were genetically dissimilar to Latvians and Lithuanians, that linguistic differences influenced their genetic differences (1086). A year later, in 2002, Laitinen and others published a report that reached a different conclusion. In the 2002 study researchers concluded (2002: 74-77), based on Y-chromosome data, that Estonians, Latvians and Lithuanians have a common genetic ancestry. Based on the presence of this haplogroup, and short tandem repeat data, the researcher further concluded that Estonians retained the ancestral Uralic language of the Baltic region, whereas language shift occurred among Lithuanians and Latvians.

The 2007 study by Rębała and others attempted to locate the homeland of Slavic-speaking people, obviously an important question for linguists, who have proposed either present-day Poland or central Ukraine. Slavic languages are currently found in eastern and southeastern Europe, and are linguistically classified into three different regional variations. Eastern Slavs consist of Belarusians, Russians, Ukrainians; Western Slavs consist of Poles, Slovaks, Czechs, Lusatians; and Southern Slavs consist of Slovenes, Croats, Bosnians, Montenegrins, Serbs, Macedonians, and Bulgarians. However, according to Rębała and others (2007: 412), Y-chromosome short tandem repeat data reveals a two part genetic division among Slavic-speaking Europeans. Bosnians, Montenegrins, Serbs, Macedonians and Bulgarians comprise one genetic group. Croats and Slovenes from the Southern Slavic group, as well as all Western and Eastern Slavs, comprise the second. Using the same data, Rębała and others traced the origins of Russians, Belarusians, Poles, Slovaks, Croats and Slovenes to the present-day Ukraine, thus favoring this region as the Slavic homeland. Moreover, according to the researchers, limited gene flow between Poles and Belarusians, who border each other, undermines support for a Polish homeland.

## 5.7 The Basques.

The Pyrenees mountain chain along the current French/Spanish border was an important refuge area during the last Ice Age, which peaked 18,000 years ago. During the Holocene, about 10,000 years ago, several different populations migrated from this region, re-colonizing previously uninhabitable areas of the European continent. Among the populations that may have remained in the Pyrenees are the Basques. The Basque language is classified by linguists as an "isolate," meaning the language is not part of a larger language family such as Indo-European or Uralic. Moreover, some linguists regard Basque as the best representative of European language diversity before the arrival of Indo-European languages (e.g. Trask 1996: 191; Vennemann 1994: 263).

In 2005 Alonso and others published a report discussing the evolutionary history of the Basque populations in Spain and France. The report found (2005: 1296) that a majority of the Basque men have the R1b1b2-M269 mutation, a signature of the Western European R-Group. Two informative downstream mutations of R-M269 mutation were also discussed in the Alonso report, R1b1b2c-M153 and R1b1b2d-M169. The report describes both polymorphisms as "putative" Iberian markers. According to the report, the R-M153 marker is found in about 7% of the Basques, and less than 1% of Iberian populations as a whole. The R-M169 mutation was reported in about 2.4% of the Basques and 5.2% of Iberians.

The 2005 Alonso report suggested that the R-M153 mutation is of Basque origins and may have been more widespread in the prehistoric Basque population. Based on this assumption, Alonso and other believe this marker would be the best choice for determining the age of the ancestral Basque population. According to the report, the M153 mutation is approximately 18,000 years old, a date that places the ancestral population of the Basques among the Paleolithic inhabitants of Europe (Alonso et al. 2005: 1298). Two other important conclusions were also reached by investigators. First, based on the R-M153 and R-M169 mutations, the genetic evidence fails to support a long-rage expansion of the Basque people in prehistoric Europe (2005: 1301). Secondly, the Alonso report, contrary to the position taken by Wilson and others in 2001, found no special relationship between Basques, Welsh and Irish populations based on short tandem repeat data (2005: 1297-1298).

In 2009 López-Parra and others published a report that examined Y-chromosome diversity in the entire Pyrenees region. The investigators found (2009: 45-48) a much higher percentage of the R-M169 mutation in the Basques than previously reported by Alonso and others in 2005. According to López-Parra, the R-M169 mutation is present in about 11% of Basques. The report also suggests that this mutation is about 7,400 years old and may have originated in the Pyrenees, and expanded outside the Pyrenees during the Neolithic. López-Parra and others also report that the age of the R-M153 mutation is about 8,500 years old, a date far different than that reported by Alonso and others in 2005. López-Parra and others attribute the dating inconsistencies to different dating techniques.

As stated earlier, Alfonso and others concluded that the Basques had not expanded into northern Europe based on the R-M153 and R-M169 mutations. However, data from López-Parra may support an alternative explanation for the prehistoric expansion of an archaic form of the Basque language. While the López-Parra study does not endorse a prehistoric Basque expansion, their study found that the second most common haplogroup in the Pyrenees is the I-M170 mutation, reported at a frequency of around twelve percent (2009: 48). Unfortunately, the López-Parra report did not further test the I-M170 mutation for the I-M253 (Scandinavian I-Group) and I-M223 (Central European I-Group) mutations (cf. Sections 4.2.1 and 4.2.4 for additional details).

In a 2007 paper, Underhill and others posit (2007: 41) that the I-M170 mutation may indeed be a marker of a prehistoric Basque expansion from the Pyrenees mountains. They support this position citing a 1994 paper by the linguist Theo Vennemann discussing the distribution of Proto-Basque hydronyms across Europe. Hydronymy is a discipline that seeks to determine how bodies of water were named. Underhill and others found a good correlation between the distribution of Proto-Basque hydronyms and the distribution of I-M223 (Central European I-Group) and I-M253 variation in Europe. Finally, mitochondrial DNA variation in Europe also reflects an expansion of human populations from the refuge area along the current French/Spanish border at the end of the last Ice Age. This is based on the frequency distributions of mitochondrial DNA haplogroups H1, H3 and V, as well as the estimated age of these haplogroups (e.g. Achilli 2004: 916).

## 5.8 Tocharian

A 2009 study by Keyser and others is among the few utilizing so-called “ancient” DNA evidence. Investigators are sometimes able to extract DNA from human remains thousands of years old. However, this technique is more successful for obtaining mitochondrial DNA (mtDNA), whereas extracting Y-chromosome data is problematic. A cell has just one copy of the Y-chromosome history, and potentially thousands of copies of the maternally inherited mitochondrial DNA (Pakendorf and Stoneking 2005: 166). Consequently, over time more mtDNA data will survive the effects of decay.

The Kurgan people are part of the Eurasian prehistory and associated by some as the first speakers of Indo-European languages (cf. Section 2.2). Astonishingly, Keyser and others were able to successfully extract both mitochondrial DNA (mtDNA) and Y-chromosome data from cadavers at a Kurgan burial site in Krasnoyarsk region of south-central Siberia. The cadavers were buried 1,600 to 3,800 years ago, but they remained frozen underneath the tundra, preserving the genetic material. Researchers were able to obtain Y-chromosome data from ten cadavers, nine of which contained the R-M17 mutation. Twenty-six cadavers provided mtDNA data. According to the mtDNA data, over time the Western Eurasian female contribution to the gene pool decreased, 90% for Bronze Age versus 67 % for Iron Age, while the Eastern Eurasian female mtDNA contribution increased (2009: 399-404). Based on the Y-chromosomal and mitochondrial data, Keyser made several conclusions about the Kurgan population that provided the burial remains. First, the genetic data suggest that the Kurgan population migrated to Siberia from Eastern Europe. Secondly, men and women migrated together from Eastern Europe to central Siberia. Finally, the population settled in Siberia (2009: 407).

It should be emphasized that studies utilizing ancient DNA are controversial, partly because they utilize a very small number of samples and thus reach conclusions based on insufficient data (e.g. Ammerman et al. 2006: 1875a; Barbujani and Chikhi 2006: 84-85; Torroni et al. 2006: 343). Nevertheless, the direction of the Kurgan migration, from Eastern Europe to Siberia, is significant because discussions of the Kurgan culture mostly posit an expansion form east to west. However, the study by Keyser and others presents data (2009: 408) suggesting a Kurgan migration in the opposite direction. Keyser and others also made another conclusion which may be of significance for linguists. Their data may suggest that an eastward migration of Kurgans might have been the source population for Tocharian (cf. Section 2.1), an extinct Indo-European language found in northwestern China.

## 5.9 The Kalmyk.

A report published in 2005 focuses on the Kalmyk people living north of the Caspian Sea. Researchers attempted to use genetic evidence to confirm whether the Kalmyk people are from Mongolia, as indicated by the historical record and the linguistic evidence. The Kalmyk speak Mongolian. According to the report, the mitochondrial and Y-chromosome data reflect that the Kalmyks migrated in substantial numbers from Mongolia to the lower Volga River. Furthermore, although the Kalmyks have lived in close proximity to ethnic Russians for the last 300 years, Russians have not made a substantial contribution to the Kalmyk gene pool. A possible explanation is cultural factors, including differences in language and religion (Nasidze et al. 2005: 851-852). The Kalmyk are Buddhists and speak a Mongolian language, whereas their neighbors are Slavic-speaking Christians.

## 5.10 The Gagauz.

The Kalmyk provide an example where linguistic identity can result in genetic isolation. In 2006 Nasidze and others published a study that focused on the Gagauz, a linguistic enclave in Moldavia. The Gagauz speak a non-Indo-European language, Turkic, yet are surrounded by speakers of Indo-European, such as Moldavian, Bulgarian, Ukrainian, and Romanian. The 2006 study by Nasidze and others examined both mitochondrial DNA and Y-chromosome data, and found evidence of gene flow from between the Gagauz and their Indo-European neighbors. According to the study, religion may explain why gene flow has occurred between Gagauz and their neighbors, whereas this has not occurred among the Kalmyks. The Kymyks (Section 5.9) remained Buddhists, whereas the Gagauz converted to the Russian Orthodox religion (2006b: 386-387).

## 5.11 The Bakhtiari.

In 2008 Nasidze and others published a study focusing on a linguistic enclave in Iran, the Bakhtiari, speakers of an Indo-European language surrounded by Semitic-speaking Iranian Arabs. This study examined both the mitochondrial and Y-chromosomal data, and concluded (2008: 249) that although substantial gene flow occurred between both groups, Bakhtiari maintained their language rather than shift to the neighboring Semitic language. Thus like the Gagauz, the Bakhtiari also maintained their language while assimilating people from other populations.

## 5.12 Language Shift in Great Britain and Ireland.

The R-M269 mutation, which I labeled the Western European R-Group, potentially reflects the post-Ice Age human settlement of the British Isles and Ireland (Semino et al. 2000a: 1155). Moreover, the R-M269 mutation represents a signature of Celtic-speaking people in Ireland and Britain (Wilson et al. 2001: 5079; Helgason et al. 2000: 714; Hill et al 2000: 351). Wells and others in their 2001 study mention that the Celts arrived relatively recently in Britain and Ireland, about 3,000 years ago. Their study suggests that the Mesolithic settlement of Ireland and Britain, and the associated introduction of the R-M269 mutation, occurred much earlier than the introduction of the Celtic cultural package (2001: 10248). Since Celtic is an Indo-European language, the Celtic cultural package may have involved language shift among the original Paleolithic inhabitants of the Ireland and the British Isles without a replacement of the gene pool in this area of Europe.

The spread of Indo-European languages into the British Isles and Ireland may also have an alternative explanation other than the adoption of a Celtic cultural package. It may be connected to the spread of farming from present-day Turkey and the addition of currently undefined downstream mutations of the R-M269 marker without the Atlantic modal haplotype. As the reader may recall, in the previous chapter I mention a recently published report by Morelli and others, who advocate a dual expansion model of the R-M269 mutation (cf. Section 4.1.1). Morelli and others assert that short tandem repeat data, and more specifically the Atlantic modal haplotype, distinguish R-M269 haplogroups from Iberia from those originating in Turkey (2010: 2). The Atlantic modal haplotype was initially proposed in a report published by Wilson and others in 2001. Wilson and others also took the position (2001: 5079-5018) that the Atlantic modal haplotype is indicative of a common paternal ancestry linking the Basques with the Celtic peoples of Ireland and Britain. Assuming that the Atlantic modal haplotype is indicative of R-M269 mutations from Iberia, data from Capelli and others suggest (2006: 981) that roughly 67% of R-M269 mutation in the United Kingdom and 52% of R-M269 mutations in Ireland are of Iberian origins. The remaining R-M269 mutations, those without the Atlantic modal haplotype, roughly one-third of those in the United Kingdom and one-half of those in Ireland, potentially reflect those that came from an expansion of R-M269 mutations from the Middle East, possibly associated with the spread of agriculture.

The above and perhaps confusing discussion of R-M269 mutations with and without the Atlantic modal haplotype indicates that geneticists have suspected that the R-M269 mutation has downstream mutations that will someday clarify the murky picture of its origins and expansion. Additional downstream markers finally appeared in a 2011 report by Myres and others. The report suggests (2011: 98-99) that the R-M412 marker represents 95% of the R-M269 variation in Europe, and the current distribution of R-M269 variation has resulted from Neolithic or later migrations. On the other hand, Mesolithic migrations (perhaps from Iberia) account for less than 5% of the R-M269 variation in Europe. Hopefully, further resolution of the M-R269 mutation may bring a deeper understanding of Celtic origins and Celtic language shift in Britain and Ireland.

The I-M253 mutation (Scandinavian I-Group) may represent the signature of Anglo-Saxon invasion of Britain. In the fifth century, the Anglo-Saxons left continental Europe and invaded Britain, forcing the Celtic inhabitants to flee to Wales or Brittany. The Anglo-Saxons brought a new language, Old English, an archaic Germanic language and the ancestral language of Modern English. In a report attempting to determine the size of the Anglo-Saxon invasions in Britain, Weale and others (2002: 1012) compared the gene pools of two towns in North Wales with five towns in Central England. As expected, the R-M269 mutation (Western European R-Group) represents most of the Y-chromosome variation for towns in Wales, between 56% and 89%. The frequency of Haplogroup I, however, was only 4% for Wales. The same haplogroup accounted for 27% of the genetic variation in central England, where the Anglo-Saxons settled.

In a detailed Y-chromosome study utilizing over 1,700 samples, published in 2003 by Capelli and others, the I-M170 mutation attains a frequency of 16.6% in Britain and 7.6% in Ireland (2003: 981). As the reader may recall from Section 4.2, the I-M170 mutation has four important subclades, each with a different demographic history: I-M253 (Scandinavian I-Group), I-M423 (Balkan I-Group), I-M223 (Central European I-Group) and I-M26 (Sardinian I-Group). The figures provided by Capelli are for the I-M170 mutation without the I-M26 mutation, but the I-M423, I-M223, and I-M253 mutations were not tested. However, data provided by Underhill in 2007 reflect that about 80% of all the I-M170 variation in the United Kingdom belongs to the I-M253 subclade, whereas the figure for Ireland stands about 55% ( 2007a: 36).

The 2003 study by Capelli and others attempted to identify Celtic, Viking, Danish and Anglo-Saxon components of the British gene pool. The study attempted to utilize a population sample from Schleswig-Holstein in Northern Germany as representative of the Anglo-Saxon invaders. However, this became problematic for the study because Germans from Schleswig-Holstein and their northern neighbors, the Danes, are, according to the study, genetically indistinguishable (2003: 979). Moreover, the 2003 Capelli study examined the Frisian sample utilized by Weale and others in 2002, and found that Frisians, Northern Germans, and Danes are also genetically indistinguishable

(2003: 982). The Weale study from 2002 concluded (2002: 1018) that although a strong genetic barrier separates North Wales and central England, genetic variation between central England and Friesland is virtually absent. In my opinion, the data from Weale and others in 2002, and Capelli and others in 2003, suggest that the presence or absence of haplogroup I-M253 separates the British of Celtic ancestry from those of Anglo-Saxon heritage, or do I dare say, the British of Germanic heritage.

Another question that researchers have attempted to answer, using Y-chromosome data, involves the size of the Anglo-Saxon invasion. Weale and others, based on the genetic evidence, believe that the invasion was "massive" (2002: 1018). In 2006, Thomas and others published a study that examined the Y-chromosome data provided by Weale and others in 2002 and Capelli and other in 2003. According to the 2006 Thomas report, the current Anglo-Saxon genetic component of the British gene pool would have required an invasion of over 500,000 people, a figure not supported by the historical record (2006: 2651). The study proposed an alternative scenario drawing from the former Apartheid system in South Africa as well as the historical record of England. Thomas and others propose that the Anglo-Saxons did not intermarry with the Celtic inhabitants for at least two centuries and had greater reproductive success than the Celtic men. Using computer simulation, the study reports (2006: 2653-2656) that such a social scenario would account for the large Anglo-Saxon presence in the current British gene pool without the need for a massive invasion in the fifth century.

### 5.13 Language Shift in the Caucasus.

The Caucasus region lies between the Black Sea and the Caspian Sea, bordered by Turkey and Iran in the south, and Russia in the north. This area of the world is very complex from the perspective of linguistic diversity. The inhabitants of this region speak languages from the Caucasian, Indo-European and Altaic families. Among those speaking an Indo-European language in the Caucasus are the Armenians. The Azerbaijani are among those speaking a Turkic (Altaic) language. In a 2003 study, Nasidze and others examined mitochondrial and Y-chromosome variation in the Caucasus in an attempt to determine if the genetic evidence would support language shift from Caucasian to Indo-European in Armenia, and Caucasian to Turkic in Azerbaijan. The study found that the Armenians and Azerbaijani are genetically closer to their Caucasian-speaking neighbors than to Indo-European-speaking or Turkic-speaking people outside the region. Based on the genetic data Nasidze and others proposed that language shift occurred in Armenia and Azerbaijan without a large influx of people from outside the region (2003: 259-260). In a more detailed examination of genetic diversity of the Caucasus region, published a year later in 2004, Nasidze and others found significant genetic differentiation among the various populations of the Caucasus. The study maintains language played a minor role in shaping the genetic differences, but rather these differences arose through genetic drift, reflecting small population sizes and geographic isolation (2004a: 216-219).

As a linguist, it would be interesting to examine the direction of lexical and grammatical borrowings among the Armenians, Azerbaijani and other ethnicities in the Caucasus for a more complete picture of language shift and maintenance in this region. Focusing now on an unrelated matter, for the linguist a significant finding stemming from examining genetic diversity in the Caucasus is one that addresses a rumor linking the Basque people of Spain and France to the Caucasus region. Based on the genetic evidence, Nasidze and others rejected the common origins of Basques and Caucasians (2003: 258-259).

### 5.14 Topography as an Explanation of Linguistic and Genetic Diversity.

Arguably, terrain may have controlled the direction of some prehistoric migrations that ultimately brought a new language into a given area. Such a hypothesis can be confirmed by Y-chromosome data and other genetic markers. Iran has two major deserts, the Dasht-e Kavir and Dash-e Lut, as well as the Hindu-Kush mountain range. A study by Regueiro and others in 2006 compared the complex geography of Iran with Y-chromosome variation in this region. According to the study, topography and genetic diversity in Iran are interconnected, reflecting that for thousands of years Iran has been a corridor for human migration between Africa, the Middle East and south Asia (141-142). Among the migrations that traversed the Iranian corridor was one that carried the J-M172 from the Middle East to Pakistan and India (Regueiro et al 2006: 140). In my opinion, this migration from Anatolia may explain the source of Indo-European languages found in India.

The *Linearbandkeramik* (Linear Pottery) Culture brought farming to central Europe during the Neolithic, about 7,000 years ago. In my opinion, the expansion of this group may also explain transmission of Indo-European languages to the Mesolithic people of Scandinavia. According to the archaeological record, the expansion of the *Linearbandkeramik* Culture followed central European river valleys (Scarre 2005b: 407). In their 2010 study, Underhill and others found the R-M458 mutation attains its highest frequency in river basins that were used as an expansion corridor by the *Linearbandkeramik* Culture (2010: 481-482). The R-M458 mutation is a downstream variant of the R-M17 mutation, which I have identified as the Eastern European R-Group.

### 5.15 Conclusion.

Part of the goal of this chapter was to establish precedent, that genetic research represents a useful tool for the linguist. One potential use of genetic data is determining where a language originated. For example, Section 5.1 suggests that humans developed language prior to leaving Africa. Similarly, Tocharian may have originated among the Kurgan people of Eastern Europe (Section 5.8). Genetic data are also useful for determining if a shift in language was mediated by a population expansion. For example, Section 5.1 reflects that an expansion of Bantu farmer also produced an expansion of the

Bantu language. The expansion of the Germanic tribes into Great Britain produced a shift away from Celtic (Section 5.13). However, the adoption of Uralic in Hungary (Section 5.5), Slavic in Eastern Europe (Section 5.6) and Armenian in the Caucasus (Section 5.13) was not preceded by a large population expansion. Genetic studies also can be useful in determining whether a shift in language was mediated by gender. For example, in Africa language shift was mediated by men (Section 5.2). Population studies also question how resistant people are to language shift in the face of cultural exchange, such as marriage outside the culture (Sections 5.10 and 5.11). Additionally, genetic research represents a tool for studying the relationship between topography and linguistic diversity (Section 5.14). For example, the rivers of Europe may have fostered the expansion of Indo-European languages. However, I believe the most significant conclusion that can be drawn from the data in Chapters Four and Five is the occasional correlation between the genetic and linguistic diversity. This correlation seems especially robust for Proto-Basque and the Scandinavian I-Group (cf Sections 4.2.1 and 5.7) Germanic languages and the Scandinavian I-Group (cf. Sections 4.2.1 and 5.12), the Western R-Group and Celtic (Sections 4.1.1 and 5.12), the Finno-Baltic N-Group and Uralic languages (Sections 4.3 and 5.6), the Near Eastern J-Group and Proto-Indo-European languages (Sections 4.5 and 5.4), and the European E-Group and Afroasiatic languages (Sections 4.4 and 5.3).

# Chapter Six
# Evaluating Contemporary Models of Germanic Origins

## 6.0 Chapter Introduction.

Using genetic data, primarily Y-chromosome data, this chapter evaluates four contemporary models of Germanic origins. These models of Germanic origins posit that Proto-Indo-European, Proto-Uralic, Proto-Afroasiatic, and Proto-Basque played a role in the evolution of the first Germanic languages, or Proto-Germanic. Surprisingly, the data collected in Chapters Four and Five clarify the prehistoric migration and settlement of people who spoke Proto-Indo-European, Proto-Uralic, Proto-Afroasiatic, and Proto-Basque. Thus, I would argue the scholar could evaluate contemporary models of Germanic origins by determining if people speaking these languages had, in fact, migrated to the Germanic homeland.

As noted earlier in chapter two (cf. Section 2.4), linguists uniformly place the putative Germanic Homeland in northern Germany, Denmark and Southern Sweden, as proposed by Gustav Kossinna in 1896. Within this area, the ideal population study would examine genetic variation found in modern-day Schleswig Holstein, Denmark and Skäne, the southernmost county in Sweden. However, at the present time the most representative data I can provide for the putative Germanic homeland are those for Denmark. Please refer to Table 6.1 below. The data for Denmark are less than ideal in that they are compiled from several different studies rather than from a single study. Consequently, it would be difficult to report absolute frequency of the haplogroups that comprise the Danish gene pool. Nevertheless, the data clearly suggest that the Danish gene pool was shaped by several different prehistoric population expansions.

I would like to emphasize that the goal of this chapter is not to pass judgment on contemporary models of Germanic origins. Instead, my goal is to establish precedent, to show that genetic data is one of several useful tools for the linguist. In other words, if the contemporary models of Germanic origins were on trial, this chapter simply represents a competency hearing to determine if population genetics should be allowed to testify. The reason for limiting the scope of my inquiry into contemporary models of Germanic origins is that language origins represent a complex research direction ultimately requiring some degree of consensus among several disciplines, including linguistics, anthropology, archaeology and population genetics, and to a lesser extent geology, climatology, ethnic studies, and demographics.

**Table 6.1 Summary of Y-chromosome Data for Denmark.**

**Western European R-Group (R1b1b2-M269)**

| Nomenclature Used | Frequency | Source |
|---|---|---|
| Haplogroup 1 | 41.7% | Helgason et al. 2000 |
| Haplogroup 1 | 50.0% | Rosser et al. 2000 |
| Haplogroup 1 | 57.1% | Scozzari et al. 2000 |
| R1xR1a1/<br>AMH +1 | 39.0% | Capelli et al. 2003 |
| R1b | 36.1% | Tambets et al. 2004 |
| R1b1b2-M269 | 42.9% | Balaresque et al. 2010 |
| R1b1b-M269 | 36.3% | Myres et al. 2011 |

**Eastern European R-Group (R1a1a-M17)**

| Nomenclature Used | Frequency | Source |
|---|---|---|
| Haplogroup 3 | 16.7% | Helgason et al. 2000 |
| Haplogroup 3 | 7.0% | Rosser et al. 2000 |
| Haplogroup 3 | 5.7% | Scozzari et al. 2001 |
| R1a1/3.65+1 | 12.0% | Capelli et al. 2003 |
| R1a | 16.5% | Tambets et al. 2004 |

**Scandinavian I-Group (I1-M253)**

| Nomenclature Used | Frequency | Source |
|---|---|---|
| I1-M253 | 32.8% | Underhill et al. 2007 |

**Central European I-Group (I2b1-M223)**

| Nomenclature Used | Frequency | Source |
|---|---|---|
| I2b1-M223 | 4.9% | Underhill et al. 2007 |

**European E-Group (E1b1b1a2-V13)**

| Nomenclature Used | Frequency | Source |
|---|---|---|
| E3b | 3.0% | Capelli et al. 2003 |
| E-V13 | 2.9% | Cruciani et al. 2007 |

**Finno-Baltic N-Group**

| Nomenclature Used | Frequency | Source |
|---|---|---|
| Haplogroup 16 | 0.0% | Helgason et al. 2000 |
| Haplogroup 16 | 2.0% | Rosser et al. 2000 |
| Haplogroup 16 | 2.9% | Scozzari et al. 2001 |
| N3 | 0.5% | Tambets et al. 2004 |

**Near Eastern J-Group (J2-M172)**

| Nomenclature Used | Frequency | Source |
|---|---|---|
| Haplogroup 9 | 7.0% | Rosser et al. 2000 |

## 6.1 Wiik’s Uralic Substratum Model.

When two or more languages converge, the role played by a language can be defined by the terms substratum, superstratum and adstratum. The term substratum defines the less dominant language, whereas the term superstratum defines the more dominant language.  A co-equal relationship among converging languages is referred to as adstratum.  To illustrate the effect of substratal and superstratal influence, Theo Vennemann (2000: 238-239)  cites an example from the history of the English language.  Vennemann explains that Celtic provides a possible substratum influence in the development of the English language, whereas Norman French provides an example of superstratum influence.  According to Vennemann, substratum influence generally changes the structure and less the lexicon of the dominant language.  Conversely, superstratum influence changes the lexicon and less the structure of a less dominant language.

The Uralic substrate model, published in 2003 by Kalevi Wiik, proposes that Germanic languages diverged from the Uralic language family.  What is particularly unusual about this paper is a note on the first page, apparently from the editors of the journal, claiming that Wiik's view of Germanic origins runs contrary to conventional linguistics and genetic evidence (2003: 43).  The genetic evidence cited by the editors stems from the classical markers published in 1994 by Cavalli-Sforza and others (cf. Section 3.4).  Nevertheless, the editors present the paper because they believe that Wiik has a novel approach and they believe the paper helps to resolve the problem of ethnicity in Mesolithic Europe.

Wiik advocates (67-73) a Germanic homeland located in Northern Germany, Denmark and Southern Sweden.  He maintains that the Mesolithic inhabitants of this region, the Ertebølle Culture, spoke a Uralic language.  Uralic languages are not part of the Indo-European language family, but rather a separate language family extending across northern Eurasia (cf. further Austerlitz 1990: 569-576).  Wiik argues that language shift occurred in the Germanic homeland, that over the course of thousands of years, the people of this region shifted from a Uralic language to an Indo-European language, which later evolved into Germanic.  In his paper, Wiik asserts that the shift from Uralic to Indo-European followed the shift from hunter-gathering food production to agricultural food production.  This borrows from the language-farming theory, which posits that the expansion of Indo-European languages across Europe followed the expansion of agriculture on the continent (cf. Sections 2.2 and 5.4).  In his paper Wiik provides twelve linguistic examples as evidence of a Uralic substratum in the emergence of Proto-Germanic.  He argues, for example, that Grimm's Law occurred as the result of imperfect second language learning of Indo-European by speakers of Uralic.

In modern-day Scandinavia, Finns and Saami speak a Uralic language, whereas ethnic Danes, Swedes and Norwegians speak a Germanic language, part of the Indo-European language family.  In my opinion, this pattern of language variation was shaped by prehistoric population expansions that began with the human re-settlement of Scandinavia shortly after the last Ice Age had ended, about 10,000 years ago.  For Denmark, the genetic signature of this settlement is the I-M253 mutation (Rootsi et al. 2004:129), which I call the Scandinavian I-Group (cf. Section 4.2.1).  The frequency of this mutation in the contemporary Danish gene pool reflects that roughly one-third of all Danes are potential descendants of the founding population in Scandinavia about ten thousand years ago.  The remaining Y-chromosome haplogroups in the Danish gene pool reflect that about two-thirds of all Danes are potential descendants of Neolithic or post-Neolithic population expansions, possibly beginning 7,000 year ago. The J-M172 mutation or Near Eastern J-Group (cf. Section 4.5) represents a very small (4%) Neolithic component of the contemporary Danish gene pool (cf. Table 6.1). Based on the recent discovery of additional downstream markers, the R-M269 (Western European R-Group) and the R-M17 (Eastern European R-Group) mutations probably also represent a Neolithic component in the Danish gene pool (cf. Sections 4.1.1 and 4.1.2).

The genetic evidence tends to undermine the position taken by Wiik. From a Y-chromosome perspective, the Germanic homeland was never inhabited by large numbers of Uralic speakers. The N-M178 mutation or Finno-Baltic N-Group is regarded by many researchers as the signature of a prehistoric migration of a Uralic-speaking population from Siberia into the Baltic Region and eventually into Finland and northern Scandinavia (cf. Section 4.3). However, studies of Y-chromosome variation in Denmark report frequencies of the N-M178 mutation from zero to three percent (see Table 6.1). Moreover, people with the N-M178 mutation were not among the founding populations of Scandinavia during the Mesolithic, about 10,000 years ago. Instead, Uralic speakers arrived in Scandinavia at a much later time in the prehistory, about around 5000 BC (cf. Section 4.3).

Genetic evidence also suggests that the presence of Uralic in Scandinavia may represent a shift from a non-Uralic language to a Uralic language. Studies have found that among the Finns and Saami, the two main Uralic-speaking populations of Scandinavia, the N-M178 mutation is high, about 60% of the entire Finnish gene pool (e.g. Lappalainen et al. 2006: 209) and around 47% among the Saami (e.g. Tambets et al. 2004: 671). However, the I-M253 mutation (Scandinavian I-Group) is also found among both populations, around 28% in Finland (e.g. Lappalainen et al. 2006: 208) and 29% among the Saami (Rootsi et al. 2004:130). Finns and Saami also lack mutations commonly found among the Uralic-speaking people of Siberia, haplogroups N2, C and Q (Lappalainen et al. 2006: 213; Tambets et al. 2004: 671). Among the Saami, the European ancestry of women is even stronger. Almost 90% of the Saami mitochondrial DNA has its origins among the founding female populations of Scandinavia, who like the men, also arrived in this area during the Mesolithic, following a population expansion that originated along the current French/Spanish border (Achilli 2005: 885). Ethnic Finns, from the perspective of mitochondrial DNA, are also of European ancestry. According Lahermo and others (1996: 1319), based on mitochondrial and nuclear DNA, Finns are genetically indistinguishable from other European populations. Thus, in my opinion, the genetic data suggest that the founding populations of Scandinavia adopted Uralic after contact and admixture with men from Asia carrying the N-M178. This direction of language shift is essentially the opposite of that proposed by Wiik.

It is important to emphasize that population genetics cannot provide the full picture of prehistoric language interaction in Scandinavia. For example, the literature often reports that the Finnish language contains numerous loanwords from the Proto-Germanic languages (e.g. Waterman 1976: 22). A list of Germanic and Finnic cognates is provided below in Table 6.2. The study of Finnish borrowings serves two purposes.

First, since the Germanic and Finnish tribes had prehistoric contact, and since Finnish has changed relatively little over the two past millennia, the Germanic borrowings in Finnish are thought to provide a well-preserved image of early Germanic phonology and morphology (e.g. Loikala 1977: 229-230). Secondly, Finnish borrowings are used to supplement the archaeological record in assessing the extent of interaction between Uralic and Germanic tribes. For example, in an article published in 1977, Hans Fromm argues that loanwords may point to the presence of the Germanic tribes in central Sweden during the Bronze Age, roughly 3,000 years ago.

**Table 6.2 Germanic and Finnic Cognates.**

| **Germanic Cognate** | **Finnic Cognate** |
|---|---|
| OLD HIGH GERMAN *feld* ‘field’ | FINNISH *pelto* ‘field’ |
| OLD NORSE *hringr* ‘ring’ | FINNISH *rengas* ‘ring’ |
| OLD NORSE *konungr* ‘king’ | FINNISH *kuningas* ‘king’ |
| GOTHIC *lamb* ‘lamb’ | FINNISH *lammas* ‘lamb’ |
| GOTHIC *mulda* ‘earth’ | FINNISH *multa* ‘earth’ |
| GOTHIC *wein* ‘wine’ | FINNISH *viina* ‘alcohol’ |
| GOTHIC *mēki* ‘sword’ | FINNISH *miekka* ‘sword ’ |
| GOTHIC *aiþei* ‘mother’ | FINNISH *äiti* ‘mother’ |
| GOTHIC *skauns* ‘beautiful’ | FINNISH *kaunis* ‘beautiful’ |
| GOTHIC *gulþ* ‘gold’ | FINNISH *kulta* ‘gold’ |
| GOTHIC *paida* ‘tunic’ | FINNISH *paita* ‘shirt’ |

**Source: Loikala 1977: 227-240.**

## 6.2 Anthony’s Kurgan Model.

In an article published in 2008, David Anthony (2008: 5-10) endorses Gimbutas’ Kurgan theory, arguing that the Indo-European homeland is located on the Russian steppes, north of the Black and Caspian Seas. He frames his argument by maintaining that the homeland is either Anatolia, as posited by Renfrew (1987) or the Russian steppes as posited by Gimbutas (1997). (cf. Section 2.2 for a more detailed discussion of both theories). Using wheeled vehicle vocabulary, Anthony pinpoints the age of Indo-European languages to about 4000 BC. He asserts that Indo-European languages share a common vocabulary for wheeled vehicles, and this technology was invented no later than 6,000 years ago. Anthony then presents a time depth argument to challenge Renfrew's theory, maintaining the spread of farming in Europe could not have spread Indo-European languages because the agriculture expansion occurred about 2,500 years prior to the invention of wheeled vehicles.

Anthony maintains (2008: 38-43) that the emergence of Germanic resulted from an expansion of the Usatovo culture that began around 3300 BC from an area near the Black Sea. The Usatovo culture, according to Anthony, is the Pre-Germanic descendants of Kurgan culture. He asserts this expansion continued along the Dniester River and into Poland. From Poland, the Pre-Germanic Usatovo culture expanded into northern Europe following the Corded Ware cultural expansion. In his paper, Anthony emphasizes (2008: 21-27) a socio-linguistic dimension to account for the success of Proto-Indo-European languages; he maintains that this language was one of prestige and status. Using archaeological remains and linguistic reconstruction, Anthony provides examples to support this position, arguing that the Indo-European had horses, wheeled vehicles, and culture that cemented relationships with other cultures through a system of patron-client relationships and gift-giving.

In a previous section of this dissertation (cf. 5.4**)**, I reported that several geneticists have endorsed the R-M17 mutation (Eastern European R-Group) as a marker of the Kurgan and Indo-European language expansion. However, this position is undermined by two population reports that posit two different and unrelated R-M17 expansions, one from the Indus Valley in India and another from Eastern Europe. Both reports also assert that these expansions preceded any Kurgan expansion by thousands of years. Furthermore, the current distribution of R-M17 variation in Europe, based on the recent discovery of a downstream haplogroup, the R-M458 mutation, follows the initial expansion of agricultural technology on this continent rather than a Kurgan expansion from the Russian steppes. While the R-M17 mutation no longer seems to be associated with an expansion of the Kurgan culture, Underhill and others (2010: 481) hinted that the discovery of additional markers may one day provide a clearer picture of Bronze Age population expansions, which fit the time frame of Anthony's model. An alternative approach may examine haplotype frequencies (short tandem repeat data) to support a Kurgan expansion.

What is especially compelling about Anthony's position is his assertion that the spread of Indo-European followed the Dniester River. I previously mention (cf. Section 5.14) that genetic research sometimes supports topography as a mediator of language variation. Anthony's paper is also compelling in that he asserts (2008: 42) the importance of socio-linguistics in understanding the success of prehistoric Indo-European languages in Europe. Indo-European may have been the language of feasting. I will expand on this idea further in Section 6.4.

### 6.3 Renfrew's Language-Farming Model.

Colin Renfrew, in a 2003 paper, reworked his language-farming theory (cf. Section 2.2) of Indo-European origins using his interpretation of the archaeological record. His primary goal was to offer a plausible explanation for differing verbal morphologies that appeared in early attested Indo-European languages, monothematic versus polythematic verbal stems. He also offered an explanation for the *satem*/*centum* classification of Indo-European languages, which divides early attested Indo-European languages according to lexical equivalent for 'one-hundred.' His strategy (2003: 29-35) was to separate Proto-Indo-European into diachronic variants, beginning with Archaic Proto-Indo-European, progressing to Balkan Proto-Indo-European, and ending with Late Proto-Indo-European and social upheaval in the Balkans. For the purposes of this dissertation, the important assertion stemming from his paper is one that advocates the emergence of Proto-Germanic from the oldest variant of Proto-Indo-European. According to Renfrew, Ancient Proto-Indo-European was spoken by the *Linearbandkeramik* culture, the initial expansion of agriculture in Europe, a migration that terminated at the southern border of the Germanic homeland at around 5500 to 5000 BC (40-42).

The Near Eastern J-Group, as well as the Western and Eastern R-Groups, represent potential contributors to the Danish gene pool during the Neolithic, and by extension, were potentially the first speakers of Indo-European languages in this area. About seven percent of Danes have the J-M172 mutation, the genetic signature of the Near Eastern J-Group and the initial spread of Indo-European language across Eurasia (Sections 4.5 and 5.4). Genetic sampling in Denmark also reports a frequency of about 36% for the R-M269 mutation, the signature of the Western R-Group and the Neolithic expansion of Celtic, an Indo-European language (Sections 4.1.1 and 5.12). Around 10% to 15% of Danes have the R-M17 mutation, the signature of the Eastern R-Group. (cf. Table 6.1). This population was part of the Mesolithic genetic inventory in Eastern Europe, and expanded after acquiring agricultural technology (Sections 4.1.2 and 5.4). Perhaps this population also acquired an Indo-European language after adopting agriculture.

The genetic evidence clearly suggests that one approach to understanding language shift among the Mesolithic peoples of the Germanic homeland is to examine how they changed subsistence strategy, the transition from hunter-gathering to agriculture. Such an examination depends heavily on an interpretation of the archaeological record. From the archaeological record we know that the *Linearbandkeramik* Culture reached the southern border of the Germanic homeland by 5500 BC (cf. Hartz et al 2007: 570). However, archaeologist are generally in agreement that the transition to agriculture in Scandinavia was slow, that the inhabitants of the Germanic homeland may have resisted the adoption of this new technology for 2,000 years (e.g. Price 2003: 280). Debate surrounding this issue has presented three models for explaining the ultimate transition to agriculture: human migration, a food shortage, or socio-economic change (Fischer 2002: 343). Anders Fischer, a Danish archaeologist, rejects both the food shortage and migration models, maintaining these explanations of

Neolithic origins in Denmark are contrary to the archaeological evidence. Fischer (2002: 350) believes that leading figures of the Ertebølle culture began small scale agriculture and animal husbandry around 4000 BC for social functions, like feasting, and that hunter-gathering remained the chief sustenance strategy for another four hundred years. According to Fischer (372-373, 380), farm products were initially introduced in the Ertebølle Culture in order to acquire prestige. Farm products provided an opportunity for these people to acquire “exotic” food products such as meat from cattle, pigs, sheep and goats, and perhaps even grain to produce beer. He maintains (377) that the emergence of a new form of pottery called funnel beakers may have been a ceremonial vessel for beer consumption. This is significant in that these ceremonial vessels, often referred to as *Trichterbecher*, are generally considered as a marker of the transition to agriculture in Denmark, hence the term *Trichterbecherkultur* or Funnel Beaker Culture (cf. Hartz et al. 2007: 585-586). Further support for the feasting hypothesis comes from skeletal remains. By measuring stable carbon and nitrogen isotopes found in human remains from the Mesolithic and Neolithic, researchers (cf. Richards et al. 2003: 293-293) concluded that the people of prehistoric Denmark made a dietary change around 4000 BC. Instead of obtaining food from the sea, they obtained food from agriculture. Archaeological evidence (Nelson 2005: 11-13) also supports beer production in prehistoric northern Europe, based primarily on grain and other plant residue found in pottery.

Fischer’s explanation of the transition to agriculture in Denmark may illustrate how researchers could unite the approaches taken by Renfrew and Anthony. From a Y-chromosome perspective, Renfrew identified the source of the Indo-European expansion within Europe, modern-day Turkey, as well as the timing of this expansion, the Neolithic. However, Anthony’s approach is compelling in that the researcher must also consider the socio-linguistic dimension. Particularly compelling is Anthony’s suggestion (2008: 41-42) that Proto-Indo-European may have been the languages of feasting. He essentially arrived at the same conclusion as that made by Anders Fischer. Both Anthony and Fischer offer insight into the success of Proto-Indo-European in prehistoric Europe. The success of this language may have been mediated by food and feasting. Taking this a step further, it would conveniently explain the eventual language shift among prehistoric inhabitants of the Germanic homeland, from a non-Indo-European to an Indo-European language. This, in turn might explain the non-Indo-European elements of attested Germanic languages, that these elements are remnants of a non-Indo-European language once spoken in the Germanic homeland, before human populations in this area shifted to an Indo-European language that they considered more prestigious. However, language shift is just one of three outcomes that may occur when languages converge (cf. Section 2.3.2). The possibility of language maintenance or even the formation of a creole must also be weighed.

According to Fischer (2002: 380), the descendants of the Ertebølle Culture eventually adopted agricultural as the chief means of food production because it was “economically advantageous.” This would explain why Indo-European was a prestige language, that speaking this language bestowed some type of economic advantage. However, Fischer’s model raises a troubling question from the standpoint of genetics. Fischer seems to reject migration as mediating the transition to agriculture in Denmark, yet the genetic evidence seems to reach a different conclusion. As previously indicated, two-thirds of the Danish gene pool potentially reflects Neolithic ancestry. However, from a linguistic standpoint, Fischer's model is supported by the observation that a good portion of Germanic agricultural terminology has a non-Indo-European origin (cf. Waterman 1976: 36).

### 6.4 Vennemann's Language Convergence Model.

In 2000 Theo Vennemann published an article taking the position that the emergence of Proto-Germanic represents the convergence of Proto-Basque, Proto Indo-European and Proto-Afroasiatic. For an overview of Vennemann's model, please refer to Figure 6.1 below. I have slightly revised Vennemann's model of Germanic origins by using more standard descriptions of the languages. I use Proto-Basque for "Alteuropäisch (Vaskonisch) and Proto-Afroasiatic for "Atlantisch (Semitidisch). Vennemann argues that the convergence of Indo-European and Proto-Basque initially occurred and this was followed by a convergence of Pre-Germanic and Proto-Afroasiatic. The up-arrow is used to indicate that Proto-Basque had a substratum influence, and the down-arrow signals the superstratum influence of Proto-Afroasiatic. Vennemann uses primarily linguistic evidence to support his model. For example, citing the words shilling (twenty pence) and score (twenty of something), Vennemann (254-255) asserts that the system of counting by twenty is a remnant of Proto-Basque influence in Germanic. Vennemann (255-257) cites ablaut as an example of Proto-Afroasiatic influence in the formation of Germanic. The term ablaut refers to a change in vowel quality to signal tense distinctions in the strong class of Germanic verbs, e.g. Eng. *drink*, *drank*, *drunk*.

**Figure 6.1 Overview of Vennemann's Language Convergence Model.**

Proto-Afroasiatic
↓
Archaic Indo-European → Western Indo-European → Germanic
(Pre-Germanic)
↑
Proto-Basque

**Source: Vennemann 2000: 261.**

Turning now to the possible influence of Proto-Basque in the emergence of Germanic, as noted in Section 5.6 above, the Basque language is often cited among linguists as the best representation of European languages before the arrival of Indo-

European speakers. The Basque homeland straddles the current French/Spanish border. This is significant because during the last Ice Age, human populations in Europe congregated in this area to survive the cold and ice. After the last Ice Age, about 10,000 years ago, human populations expanded from the current French/Spanish border and migrated northwards. Among the populations expanding from this refuge area were those having the I-M253 mutation (Scandinavian I-Group). The expansion of the I-M253 mutation ultimately terminated in Scandinavia (cf. Section 4.2.1 above).

As outlined above, from a Y-chromosome perspective, the genetic evidence points to the I-M253 mutation as the genetic signature of the founding population of the Germanic homeland. Since this founding population migrated from the homeland of the Basque people, the first inhabitants of the Germanic homeland may have spoken a non-Indo-European language, something akin to Proto-Basque. This is supported by a similar distribution pattern of Proto-Basque hydronyms and I-M253 mutations in Europe (cf. Section 5.7). Thus, the genetic evidence offers a plausible explanation for the influence of Proto-Basque in the evolution of Germanic, as asserted by Vennemann (2000). However, from the perspective of language contact theory, genetics and archaeology, Vennemann's model may have to be revised. As the current model stands, Vennemann posits that Indo-European speakers maintained their language. However, from the perspective of population genetics, archaeology and language contact theory, in the Germanic homeland speakers of Proto-Basque may have shifted to the Indo-European language of farmers (cf. Section 6.2) starting around 4000 BC.

Turning now to the second part of Vennemann's convergence model, from the perspective of linguistics, the convergence of Pre-Germanic and Proto-Afroasiatic may seem rather doubtful. Languages classified as Afroasiatic are mostly found in the Near East and North Africa and include Arabic, Hebrew, Egyptian and Berber. (cf. Hetzron 1990: 647-653 for more information) However, I would like to suggest that the E-V13 mutation (European E-Group) may be the genetic signature of Proto-Afroasiatic speakers during the European Mesolithic. According to Arredi (2004: 343), a back-migration of the E-M35 mutation into Africa from the Middle East may have introduced Afroasiatic languages to North Africa. By extension, I suggest that E-V13 populations entering Europe from the Near East may have also spoken something akin to Proto-Afroasiatic. (cf. Sections 4.4 and 5.3 above for additional details) Although the E-V13 mutation contributed little to current Danish gene pool, three percent at best (cf. Table 6.1), the influence of Afroasiatic populations might be greater than implied by this small frequency figure. Archaeological evidence points to prehistoric commerce between southeastern Europe and the Ertebølle culture around 4000 BC. During this time in the prehistory, the Ertebølle Culture may have imported copper axes from what is now modern-day Serbia (cf. Klassen 2002: 312). In this region of Europe, the frequency of the E-V13 mutation is around twenty percent (Appendix Table 8). Based on the strength of the E-V13 mutation in Serbia, the traders of copper axes in prehistoric Denmark may have spoken Proto-Afroasiatic. Thus, a convergence of Pre-Germanic and Proto-Afroasiatic is plausible, especially since copper axes must have been a prized commodity in Mesolithic Denmark.

## 6.5 Chapter Conclusion.

Perhaps what is striking about the four contemporary models of Germanic origins (6.1 Uralic Substratum, 6.2 Kurgan Expansion, 6.3 Language-Farming, 6.4 Convergence) is that they seem to gravitate away from the traditional *Stammbaum* model of Germanic origins, that Proto-Germanic separated along with Proto-Balto-Slavic from a parent Proto-Indo-European language and developed its unique innovations in isolation. Rather, the four models I introduced in this section seem to be more receptive to language contact theory. However, I am not prepared to endorse any of the four models of Germanic origins discussed in this chapter. Again, my goal is not to render judgment on the four contemporary models of Germanic origins; rather, my goal is to establish precedent. I want to demonstrate that population genetics is a valuable tool for evaluating contemporary models of Germanic origins, and that this tool complements the traditional linguistic and archaeological approaches. In chapter seven I conclude that population genetics is indeed a valuable tool for linguists.

# Chapter Seven
# Dissertation Conclusion

In Chapter One, I set forth the central idea or theme of this dissertation: Y-chromosome data are a useful tool for evaluating contemporary models of Germanic origins. The Second Chapter in this dissertation provides an overview of where the search for Germanic origins now stands. The traditional tools for exploring the origins of Germanic languages have been linguistics and archaeology. From a linguistic perspective, *Stammbaum* theory (cf. Section 2.3.1) and language contact theory (cf. Section 2.3.2) provide two alternative approaches for explaining the origins of Germanic languages. In Chapter Two, I also introduce population genetics and Y-chromosome data as a new tool for exploring Germanic origins (cf. Section 2.6). However, three barriers currently prevent researchers from utilizing this new approach to language origins. First, the methodology behind population genetics requires additional explication so that a wider audience can evaluate the usefulness of this tool. Secondly, the reporting of Y-chromosome data has used a nomenclature system that has undergone standardization and refinement over the last eleven years. Finally, the reporting of Y-chromosome data is extremely fragmented, contained in hundreds of population reports.

In Chapter Three, I explain the methodology behind population genetics (cf. Sections 3.1 - 3.3 for additional details). This dissertation focuses primarily on a research direction in population genetics that gathers and interprets data from the non-recombining region of the human Y-chromosome. In the course of mammalian evolution, a large section of the Y-chromosome was damaged. Consequently, much of the Y-chromosome now avoids a "reshuffling" of genetic material known as recombination. The absence of recombination means that much of the Y-chromosome is inherited largely unaltered from generation to the next. Nevertheless, the Y-chromosome can vary from one population to the next. The source of this variation is mutations called single nucleotide polymorphisms. These Y-chromosome mutations record prehistoric migration and settlement because prehistoric populations tended to have their own genetic signature due to drift. By examining the frequency of Y-chromosome variation over a given geographic distance, geneticists can decipher the origin and direction of a prehistoric migration. Geneticists can also estimate when a single nucleotide polymorphism arose by counting short tandem repeats, another type of genetic mutation. In Chapter Three (cf. Section 3.4) I also explain why this dissertation focuses on Y-chromosome data, and to a lesser extent, mitochondrial DNA data. The reason for focusing on these molecular markers stems partly from the amount of literature that has been published, and partly because the data for both markers can be organized into easy-to-understand phylogenetic trees.

In Chapter Four I untangle the confusing nomenclature used by geneticists to describe Y-chromosome mutations. Research exploring Y-chromosome variation seeks to identify mutations that are often described as single nucleotide polymorphisms or haplogroups. These mutations form clades that are organized into a tree-like phylogenetic hierarchy that begins with an ancestral haplogroup or clade, which ultimately divides and expands into twenty major Y-chromosome haplogroups or clades. Each clade, in turn, has a number of subclades or sub-haplogroups. The nomenclature used to describe the Y-chromosome cladistic relationships, or haplogroups, was first standardized in 2002, and the nomenclature and cladistic relationships continue to be updated and refined. Nevertheless, for the purposes of this dissertation, the data continues to identify ten population expansions that are representative of the European prehistory. Based on the available genetic data for prehistoric population expansions, it appears that the roots of Germanic languages extend into the Mesolithic (cf. Section 4.2.1 for a discussion of the Scandinavian I-Group). Later population expansions into Scandinavia, the Western European R-Group (cf. Section 4.1.1) and Eastern R-Group (cf. Section 4.1.2), as well as the Near Eastern J-Group (cf. Section 4.5), also made a contribution to the evolution of Germanic. Finally, expansion of the Finno-Baltic N-Group (cf. Section 4.3) and the European E-Group (cf. Section 4.4) may have also made a contribution to the evolution of Germanic languages.

Chapter Five attempts to establish precedent, that genetic data are a useful tool for the linguist. The chapter presents a survey of population studies that explore the correlation between linguistic and genetic diversity. Indeed, genetic data are useful for determining if a population expansion mediated a shift in language. For example, the shift to Slavic and Hungarian was not preceded by a large population expansion, whereas a large population expansion preceded the spread of Bantu languages in Africa (cf. Sections 5.2, 5.5 and 5.5). Nevertheless, the most striking conclusion is that a language may have its own genetic signature. This seems to be the case for Proto-Indo-European, Proto-Basque, Proto-Afroasiatic, Proto-Germanic, Proto-Celtic and Proto-Uralic (Section 5.15).

Chapter Six provides a forum for returning to my thesis, that genetic data are a useful tool for evaluating contemporary models of Germanic origins. The discussion of contemporary models in Chapter Six is driven by the data rendered in Chapters Four and Five. Through the discussion, the usefulness of Y-chromosome data for evaluating contemporary models of Germanic origins becomes obvious when one considers that contemporary models are more receptive to language contact theory. Prior to the availability of genetic data, evidence for language contact induced change was limited to the historical record. Population genetics is able to overcome this inherent weakness of language contact theory. With this paradigm, the linguist has a reliable tool for evaluating possible prehistoric language convergence. According to the available genetic data, when Early Germanic finally appeared in the historical record, about 2,000 years ago, this language group had probably undergone thousands of years of linguistic evolution, ultimately becoming the product of numerous language convergences. Further research in the area of Germanic origins is warranted considering the roles played by Proto-Basque, Proto-Indo-European, Proto-Celtic, Proto-Uralic and Proto-Afroasiatic in the evolution of Early Germanic.

# Bibliography

Adams, Susan M. et al. 2008. “The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula.” *American Journal of Human Genetics* 83: 725-736.

Achilli, Alessandro et al. 2005. “Saami and Berbers – an expected mitochondrial DNA link.” *Am. J. Hum. Genet*. 76: 883-886.

Achilli, Alessandro et al. 2004. “The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool.” *Am. J. Hum. Genet*. 75: 910-918.

Ali, Sher and Seyed Ehtesham Hasnain 2002. “Molecular dissection of the Y-chromosome.” *Gene* 283: 1-10.

Alonso, Santos et al., 2005. “The place of the Basques in the European Y-Chromosome diversity landscape.” *European Journal of Human Genetics* 13: 1293-1302.

Alvarez, Luis et al. 2009. “Y-chromosome Variation in South Iberia: Insights into the North African Contribution.” *American Journal of Human Biology* 21: 407-409.

Al-Zahery, N. et al. 2003. “Y-chromosome and mtDNA polymorphisms in Iraq, a crossroads of the early human dispersal and of post-Neolithic migrations.” *Molecular Phylogenetics and Evolution* 28: 458-472.

Ammerman et al. 2006. “Comment of ‘ancient DNA from the first European farmers in 7500-year-old Neolithic sites.’” *Science* 312: 1875a.

Ammerman, A.J. and L.L Cavalli-Sforza 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton: Princeton University Press.

Anthony, David 2008. “A new approach to language and archaeology: the Usatovo culture and the separation of Pre-Germanic. *The Journal of Indo-European Studies* 36 (1/2): 1-51.

Anthony, David W. 2007(a). *The Horse, the Wheel, and Language, How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton, N.J.; Woodstock: Princeton University Press.

Anthony, David W. 2007(b). “Pontic-Caspian Mesolithic and Early Mesolithic Societies at the time of the Black Sea flood: a small audience and small effects.” In: *The Black Sea Flood Question: Changes in Coastline, Climate, and Human Settlement*. Eds. Valentina Yanko-Hombach et al. Dordrecht: Springer. 345-370.

Arnold, Bettina 2000. “A transatlantic perspective on German archaeology.” In: *Gesellschaften und Staaten im Epochenwandel* Band 7. Archaeology, Ideology and Society: the German Experience. Ed. Heinrich Härke. Frankfurt am Main; New York: Peter Lang. 398-422.

Arnold, Bettina 1990. “The past as propaganda: totalitarian archaeology in Nazi Germany.” *Antiquity* 64: 464-478.

Arredi, Barbara et al. 2004. “A predominately Neolithic origin for Y-chromosomal DNA variation in North Africa.” *Am. J. Hum. Genet*. 75: 338-345.

Arvidsson, Stefan 2006. *Aryan Idols: Indo-European Mythology as Ideology and Science*. Chicago: University of Chicago Press.

Athey, T. Whit and Kenneth Nordvedt 2005. “Resolving the placement of haplogroup I-M223 in the Y-chromosome phylogenetic tree.” *Journal of Genetic Genealogy* 1: 54-55.

Austerlitz, Robert 1990. “Uralic Languages.” In: *The World’s Major Languages*. Ed. Bernard Comrie. New York: Oxford University Press. 569-576.

Ayling, L.J. and D.K. Griffin 2003. “The evolution of the sex chromosomes.” *Cytogenetic Genome Res*. 99: 125-140.

Bailey, Douglass W. 2007. “Holocene changes in the level of the Black Sea: consequences at a human scale.” In: *The Black Sea Flood Question: Changes in Coastline, Climate, and Human Settlement*. Eds. Valentina Yanko-Hombach et al. Dordrecht: Springer. 515-536.

Balanovsky, Oleg et al. 2008. “Two sources of the Russian patrilineal heritage in their Eurasian context.” *The American Journal of Human Genetics* 82: 236-250.

Balaresque, Patricia et al. 2010. “A predominantly Neolithic origins for European paternal lineages.” *PLoS Biology,* 8 (1): e1000285.

Barać, Lovorka et al. 2003 “Y chromosome heritage of Croatian population and its island isolates.” *European Journal of Human Genetics* 11: 535-542.

Barbujani, L. and L. Chikhi. 2006. “DNAs from the European Neolithic.” *Heredity* 97: 84-85.

Barbujani, Guido and David B. Goldstein 2004. “Africans and Asians abroad: genetic diversity in Europe.” *Annu. Rev. Genomics Hum. Genet*. 5: 119-50.

Barbujani, Guido and Giorgio Bertorelle 2001. “Genetics and the population history of Europe.” *PNAS* 98(1): 22-25.

Battaglia, Vincenza et al. 2009. “Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe.” *European Journal of Human Genetics* 17: 820-830.

Beekes, Robert S.P. 1995. *Comparative Indo-European Linguistics*. Amsterdam; Philadelphia: John Benjamins Publishing.

Behar, Doron M. 2004. "Contrasting patterns of Y-chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. *Hum. Genet*. 114: 354-365.

Behar, Doron M. et al. 2003. "Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries." *Am. J. Hum. Genet*. 73: 768-779.

Beleza, Sandra et al. 2006. "Micro-phylogenetic and demographic history of Portuguese male lineages." *Annals of Human Genetics* 70: 181-194.

Belle, Elise M.S. et al. 2006. "Origins and evolution of the Europeans' genome: evidence from multiple microsatellite loci." *Proceedings Biological Sciences* 273 (1594): 1595-1602.

Belledi, Michele et al. 2000. "Maternal and paternal lineages in Albania and the genetic structure of Indo-European populations." *European Journal of Human Genetics* 8: 480-486.

Bellwood, Peter 2005. *First Farmers, the Origins of Agricultural Societies*. Malden, MA: Blackwell Publishing.

Berben, Paul 1975. *Dachau 1933-1945: The Official History*. London: Norfolk Press.

Bertranpetit, Juame 2000. "Genome, diversity, and origins: the Y-chromosome as a story teller." *PNAS* 97(13): 6927-6929.

Bíró, A.Z. et al. 2009. "A Y-chromosomal comparison of the Madjars (Kazakhstan) and the Magyars (Hungary). *American Journal of Physical Anthropology* 139: 305-310.

Blankholm, Hans Peter 2008. "Southern Scandinavia." In: *Mesolithic Europe*. Geoff Bailey and Penny Spikins (editors). New York: Cambridge University Press. 107-131.

Bogucki, Peter 2000. "How agriculture came to Europe." In: *Europe's First Farmers*. T. Douglas Price (editor). Cambridge, U.K.; New York: Cambridge University Press. 197-218.

Børglum, Anders D. et al. 2007. "No signature of Y chromosomal resemblance between possible descendants of the Cimbri in Denmark and Northern Italy." *American Journal of Physical Anthropology* 132: 278-284.

Bollmus, Reinhard 1970. *Das Amt Rosenberg und seine Gegner: zum Machtkampf im nationalsozialistischen Herrschaftssystem.* Studien zur Zeitgeschichte. Herausgeben vom Institut für Zeitgeschichte. Stuttgart: Deutsches Verlags-Anstalt GmbH.

Bosch, E. et al. 2006. “Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns.” *Annals of Human Genetics* 70: 459-487.

Bosch, Elena et al. 2003. “High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosome analysis.” *Hum. Genet.* 112: 353-363.

Bosch, Elena et al. 2001. “High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula.” *Am. J. Hum. Genet*. 68: 1019-1029.

Brackney, Noel C. 2007. *The Origins of Slavonic: Language Contact and Language Change*. Munich: LINCOM Europa.

Bradley, Daniel G. 2000. “Mitochondrial DNA diversity and origins of domestic livestock.” In: *Archaeogenetics: DNA and the Population Prehistory of Europe*. Eds. Colin Renfrew and Katie Boyle. Cambridge: McDonald Institute for Archaeological Research. 315-320.

Bramanti et al. 2009. “Genetic discontinuity between local hunter-gathers and central Europe’s first farmers.” *Science* 326: 137-140.

Brion, Maria et al. 2004. “Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis.” *Gene* 329: 17-25.

Brion, M. et al. 2003. “Insights into Iberian population origins through the construction of highly informative Y-chromosome haplotypes using biallelic markers, STRs, and the MSY1 minisatellite.” *American Journal of Physical Anthropology* 122: 147-161.

Broch, Olaf 1927. “Russennorsk.” *Archiv für slavische Philologie* 41: 209-262

Cadenas, Alicia M. et al. 2008. “Y-chromosome diversity characterizes the Gulf of Oman.” *European Journal of Human Genetics* 16: 374-386.

Campbell, Lyle and William J. Poser 2008. *Language Classification: History and Method*. Cambridge, UK; New York: Cambridge University Press.

Campbell, Lyle 2004. *Historical Linguistics*, 2nd Edition. Edinburgh: Edinburgh University Press.

Campbell, Michael C. and Sarah A. Tishkoff. 2010. “The evolution of human genetic and phenotypic variation in Africa.” *Current Biology* 20: R166-R173.

Campanile, Enrico. 1998. “The Indo-Europeans: origins and culture.” In: *The Indo-European Languages*. Eds. Anna Giacalone Ramat and Paolo Ramat. London; New York: Routledge Publishing. 1-24.

Cann, Rebecca L. and Alan C. Wilson 2003. “The recent African Genesis of humans.” *Scientific American Special Edition* 13: 54-61.

Capelli, Cristian et al. 2007. “Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic-Neolithic encounter.” *Molecular Phylogenetics and Evolution* 44: 228-239.

Capelli, C. et al. 2006. “Population structure in the Mediterranean Basin: a Y chromosome perspective.” *Annals of Human Genetics* 70: 207-225.

Capelli, Cristian et al. 2003. “A Y chromosome census of the British Isles.” *Current Biology* 13: 979-984.

Carmeli, Daphna Birenbaum 2004. “Prevalence of Jews as subjects in genetic research: figures, explanation, and potential implications.” *American Journal of Medical Genetics* 130A: 76-83.

Caratti, S. et al. 2009. “Subtyping of Y-chromosomal haplogroup E-M78 (E1b1b1a) by SNP assay and its forensic application.” *Int. J. Legal Med.* 123: 357-360.

Cavalli-Sforza, L. Luca and Marcus W. Feldman 2003. “The application of molecular genetic approaches to the study of human evolution.” *Nature Genetics Supplement* 33: 266-275.

Cavalli-Sforza, Luigi Luca 2000. *Genes, Peoples, and Languages*. New York: North Point Press.

Cavalli-Sforza et al. 1994. *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.

Cecil, Robert 1972. *The Myth of the Master Race: Alfred Rosenberg and Nazi Ideology*. London: B.T. Batsford Ltd.

Chaix, Raphaëlle et al. 2008. “Genetic traces of east-to-west human expansion waves in Eurasia.” *American Journal of Physical Anthropology* 136: 309-317.

Chikhi, Lounès et al. 2002. “Y genetic data support the Neolithic demic diffusion model.” *PNAS* 99 (17): 11008-11013.

Chiaroni, Jacques et al. 2009. “Y-chromosome diversity, human expansion, drift, and cultural evolution.” *PNAS* 106 (48) 20174-20179.

Chiaroni, Jacques et al. 2009. “Correlation of annual precipitation with human Y-chromosome diversity and the emergence of Neolithic agricultural and pastoral economies in the Fertile Crescent.” *Antiquity* 82: 281-289.

Cinnioğlu, Cengiz et al. 2004. “Excavating Y-chromosome haplotype strata in Anatolia.” *Hum. Genet.* 114: 127-148.

Clackson, James 2007. *Indo-European Linguistics, an Introduction*. Cambridge; New York: Cambridge University Press.

Comrie, Bernard 2002. “Farming dispersal in Europe and the spread of Indo-European language family.” In: *Examining the Farming/Language Dispersal Hypothesis*. Eds. Peter Bellwood and Colin Renfrew. Cambridge: McDonald Institute for Archaeological Research. 409-419.

Contu, Daniela et al. 2008. “Y-chromosome based evidence for pre-Neolithic origin of the genetically homogeneous but diverse Sardinian population: inference for association scans.” *PLoS One* 3(1): e1430.

Cordaux, Richard et al. 2004(a). “The northeast Indian Passageway: a barrier or corridor for human migrations?” *Molecular Biology and Evolution* 21: 1525-1533.

Cordaux, Richard et al. 2004(b). “Independent origins of Indian caste and tribal paternal lineages.” *Current Biology* 14: 231-235.

Crawford, Michael H. 2007. “Foundations of anthropological genetics.” In: *Anthropological Genetics*. Ed. Michael H. Crawford. Cambridge; New York: Cambridge University Press. 1-16.

Cruciani, Fulvio et al. 2011. “A revised root for the human Y-chromosomal phylogenetic tree: the origins of patrilineal diversity in Africa. *The American Journal of Human Genetics* 88: 814-818.

Cruciani, Fulvio et al. 2007. “Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosome haplogroups E-M78 and J-M12.” *Mol. Biol. Evol.* 24(6): 1300-1311.

Cruciani, Fulvio et al. 2006. “Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): a posteriori evaluation of a microsatellite-network-based approach through six new biallelic markers.” *Hum. Mutat*. 27: 831-832.

Cruciani, Fulvio. 2004. “Phylogenetic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa.” *Am. J. Hum. Genet*. 74: 1014-1022.

Cruciani, Fulvio et al. 2002 “A back migration from Asia to sub-Saharan Africa supported by high-resolution analysis of human Y-chromosome haplotypes.” *Am. J. Hum. Genet.* 70: 1197-1214.

Currat, Mathias and Laurent Excoffier 2005. “The effect of the Neolithic expansion on European molecular diversity.” *Proceedings: Biological Sciences* 272: 679-688.

Derenko, Miroslava et al. 2007. “Y-chromosome haplogroup N dispersal from South Siberia to Europe.” *J. Hum. Genet*. 52: 763-770.

Derenko, Miroslava 2006. “Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altain-Sayan regions.” *Hum. Genet*. 118: 591-604.

Dergachev, Valentin and Pavel M. Dolukhanov 2007. "The Neolithization of the North Pontic Area and the Balkans in the context of the Black Sea floods." In: *The Black Sea Flood Question: Changes in Coastline, Climate, and Human Settlement*. Eds. Valentina Yanko-Hombach et al. Dordrecht: Springer. 489-514.

Diego, Vicente García de (ed.) 1985. *Diccionario Etimológico: Español e. Hispánico*. Madrid: Editorial S.A.E.T.A.

Di Gaetano, Cornelia et al. 2009. "Differential Greek and northern African migration to Sicily are supported by genetic evidence from the Y-chromosome." *European Journal of Human Genetics* 17: 91-99.

Di Giacomo, F. et al. 2004. "Y-chromosomal haplogroup J as a signature of the post-Neolithic colonization of Europe." *Hum. Genet.* 115: 357-371.

Di Giacomo, F. et al. 2003. "Clinal patters of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. *Molecular Phylogenetics and Evolution* 28: 387-395.

Dupanloup, Isabelle et al. 2004. "Estimating the impact of prehistoric admixture on the genome of Europeans." *Molecular Biology and Evolution*, 21: 1361-1372.

Dupuy, Berit Myhre et al. 2006. "Geographical heterogeneity of Y chromosomal lineage in Norway." *Forensic Science International* 164: 10-19.

Edmonds, Christopher A. et al. 2004. "Mutations arising in the wave front of an expanding population." *PNAS* 101(4): 975-979.

El-Sibai, Mirvat et al. 2009. "Geographical structure of the Y-Chromosomal genetic landscape of the Levant: a coastal-inland contrast." *Annals of Human Genetics* 73: 568-581.

Fechner, Angela et al. 2008. "Boundaries and clines in the West Eurasian Y-chromosome landscape: insights from the European part of Russia." *American Journal of Physical Anthropology*, 137: 41-47.

Ferguson-Smith, M.A. 2009. "It is 50 years since the discovery of the male determining role of the Y-chromosome!" *Sex. Dev.* 3: 233-236.

Fischer, Anders 2002. "Food for feasting? An evaluation of explanations for the neolithisation of Denmark and southern Sweden." In: *The Neolithisation of Denmark: 150 Years of Debate*. Eds. Anders Fischer and Kristian Kristiansen. Sheffield, England: J.R. Collins Publications. 343-393.

Flores, Carlos et al. 2005. "Isolates in a corridor of migrations: a high-resolution analysis of Y-chromosome variation." *J. Hum. Genet.* 50: 435-441.

Flores, Carlos et al. 2004. “Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography.” *European Journal of Human Genetics* 12: 855-863.

Flores, C. et al. 2003. “A predominant European ancestry of paternal lineages from Canary Islanders.” *Annals of Human Genetics* 67: 138-152.

Fortson, Benjamin W. 2010. *Indo-European Language and Culture* 2nd edition. Chichester, U.K.; Malden, MA: Blackwell Publishing.

Francalacci, Paolo et al. 2010. “The history and geography of the Y chromosome SNPs in Europe: an update.” *Journal of Anthropological Sciences* 88: 207-214.

Francalacci Paolo and Daria Sanna 2008. “History and geography of human Y-chromosome in Europe: a SNP perspective.” *Journal of Anthropological Sciences* 86: 59-89.

Francalacci, P. et al. 2003. “Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability.” *American Journal of Physical Anthropology* 121: 270-279.

Frayer, D.W. 1987. “The supra-acetabular fossa and grove: a skeletal marker for Northwest European Mesolithic populations.” *Human Evolution* 3(3): 163-176.

Fromm, Hans 1997. “Germanen in bronzezeitlichen Mittelschweden?” *Finnisch-Ugrische Forschungen* 54: 127-150.

Gagnon, Alain et al. 2006. “Transmission of migratory propensity increases genetic divergence between populations. *American Journal of Physical Anthropology* 129: 630-636.

Gemmell, Neil J. and Frank Y. T. Sin 2002. “Mitochondrial mutations may drive Y chromosome evolution.” *Bio Essays* 24: 275-279.

Gérard, Nathalie et al. 2006. “North African Berber and Arab influences in the Western Mediterranean revealed by Y-chromosome DNA haplotypes.” *Human Biology* 78 (3): 307-316.

Ghiani, Maria et al. 2009. “Significant genetic differentiation within the population of the island of Corsica (France) revealed by chromosome analysis.” *Genetic Testing and Molecular Biomarkers* 13(6): 773-777.

Gimbutas, Marija 1997a. “The Fall and Transformation of Old Europe: Recapitulation 1993.” In: *The Kurgan Culture and the Indo-Europeanization of Europe: Selected Articles from 1952-1993*. Eds. Marija Gimbutas; Miriam Robbins Dexter and Karlene Jones-Bley. Washington D.C.: Institute for Study of Man. 351-372.

Gimbutas, Marija 1997b. “The Three Waves of the Kurgan People into Old Europe, 4500-2500 B.C.” In: *The Kurgan Culture and the Indo-Europeanization of Europe: Selected Articles from 1952-1993*. Eds. Marija Gimbutas; Miriam Robbins Dexter and Karlene Jones-Bley. Washington D.C.: Institute for Study of Man. 240-268.

Gonçalves, Rita et al. 2005. “Y-chromosome lineages from Portugal, Madeira and Açores record elements of Sephardim and Berber ancestry.” *Annals of Human Genetics* 69: 443-454.

Gresham, David et al. 2001. “Origins and divergence of the Roma (Gypsies).” *Am. J. Hum. Genet*. 69: 1314-1331.

Greif, Thomas 2000. *Der SS-Standort Waischenfeld: Hilfswerklager und Ahnenerbe*. Erlangen: Palm & Enke.

Grimm, Jacob 1822. *Deutsche Grammatik* erster Teil, zweite Aufgabe. Göttingen: Deutsche Buchhandlung.

Grünert, Heinz 2002. *Gustaf Kossinna 1858-1931, vom Germanisten zum Prähistoriker: ein Wissenschaftler im Kaiserreich und in der Weimarer Republik*. Rahden, [Germany]: Verlag Marie Leidorf.

Gusmão, Al. et al. 2008. “A perspective on the history of the Iberian gypsies provided by phylogenetic analysis of Y-chromosome lineages.” *Annals of Human Genetics* 72: 215-227.

Haak, Wolfgang et al. 2010. “Ancient DNA from European early Neolithic farmers reveals their Near Eastern affinities.” *PLoS Biology* 8 (11): e1000536.

Haak, Wolfgang et al. 2005. “Ancient DNA from the first European farmers in 7500-year-old Neolithic sites.” *Science* 310: 1016-1018.

Hammer, Michael F. et al. 2003. “Human population structure and its effects on sampling Y chromosome sequence variation.” *Genetics* 164: 1495-1509.

Hammer, Michael F. and Stephen L. Zegura 2002. “The human Y chromosome haplogroup tree.” *Annu. Rev. Anthropol*. 31: 303-321.

Hammer, Michael F. 2001. “Hierarchical patterns of global human Y-chromosome diversity.” *Mol. Biol. Evol.* 18(7): 1189-1203.

Hartz, Sönke et al. 2007. “From Fish and seal to sheep and cattle: new research into the process of neolithisation in northern Germany.” In: *Going over the Mesolithic-Neolithic Transition in North-West Europe*. Eds. Alasdair Whittle and Vicki Cummings. Oxford: Oxford University Press. 567-594.

Haßman, Henning. 2000. “Archaeology in the ‘Third Reich.’” In: *Gesellschaften und Staaten im Epochenwandel* Band 7. Ed. Heinrich Härke. Frankfurt am Main; New York: Peter Lang. 67-142.

Hawkins; John A. “Germanic Languages.” 1990 In: *The World’s Major Languages*. Ed. Bernard Comrie. New York: Oxford University Press. 68-76.

Heine, Bernd and Tania Kuteva. 2006. *The Changing Languages of Europe*. Oxford; New York: Oxford University Press.

Helgason, Agnar et al. 2000 “Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland.” *Am. J. Hum. Genet.* 67: 697-717.

Henriksen, Carol and Johan van der Auwera 1994. “The Germanic Languages.” In: *The Germanic Languages*. Eds. Ekkehard König and Johan van der Auwera. London; New York: Routledge. 1-18.

Hetzron, Robert 1990. “Afroasiatic Languages.” In: *The World’s Major Languages*. Ed. Bernard Comrie. New York: Oxford University Press. 647-653

Hewitt, Godfrey 2000. “The genetic legacy of the Quaternary ice ages.” *Nature* 405: 907-913.

Hill, Emmeline W. et al. 2000. “Y-chromosome variation and Irish origins.” *Nature* 404: 351-352.

Hirschenfeld, Ludwik and Hanka Hirschenfeld 1919. “Serological differences between the blood of different races.” *The Lancet* 194 (5016): 675-679.

Hutton, Christopher 1999. *Linguistics and the Third Reich: Mother-Tongue Fascism, Race and the Science of Language*. London; New York: Routledge.

Iverson Gregory K. and Joseph Salmons 2003. “Laryngeal Enhancement in Early Germanic.” *Phonology* 20: 43-72.

Jobling, Mark A. and Chris Tyler-Smith 2003. “The human Y chromosome: an evolutionary marker comes of age.” *Nature Reviews, Genetics* 4: 598-612.

Jorde, L.B. et al. 2000 “The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data.” *Am. J. Hum. Genet*. 66: 979-988.

Jorgensen, Tove H. et al. 2004 “The origin of the isolated population of the Faroe Islands investigated using Y chromosome markers.” *Hum. Genet.* 115: 19-28.

Jurić, Ivan 2005. “On the Y chromosome haplotype of the first farmers in the historical territory of Croatia and the directions of agricultural diffusion in Europe.” *Agriculturae Conspectus Scientificus* 70(4): 121-126.

Jurmain, Robert et al. (eds.) 2005. *Introduction to Physical Anthropology* 10th edition. Belmont, CA: Thomas Wadsworth.

Kalaydjieva, Luba et al. 2001. “Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages.” *European Journal of Human Genetics* 9: 97-104.

Karachanak, Sena et al. 2009. “Y-chromosomal haplogroups in Bulgarians.” *Compte rendus de l’Académie bulgare des Sciences*. 62(30): 393-400.

Karafet, Tatiana et al. 2008. “New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree.” *Genome Res.* 18: 830-838.

Karafet, Tatiana M. et al. 2002. “High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life.” *Human Biology* 74( 6): 761-789.

Karlsson, Andreas O. et al. 2006. “Y-chromosome diversity in Sweden - a long-time perspective.” *European Journal of Human Genetics* 14: 963-970.

Kasperavičiūtė, D. et al. 2004. “Y-chromosome and mitochondrial DNA variation in Lithuanians.” *Annals of Human Genetics* 68: 438-452.

Kater, Michael H. 1974. *Das “Ahnenerbe” der SS 1935 - 1945. Ein Beitrag zur Kulturpolitik des Dritten Reiches*. Stuttgart: Deutsche Verlags-Anstalt.

Kayser, Manfred et al. 2005. “Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis.” *Hum. Genet.* 117: 428-443.

Kayser, Manfred et al. 2004. “A comprehensive survey of human Y-chromosomal micro-satellites.” *Am. J. Hum. Genet*. 74: 1183-1197.

Keyser, Christine et al. 2009. “Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum. Genet.* 126: 395-410.

Kharkov, V. N. et al. 2007. “Gene pool differences between Northern and Southern Altaians inferred from the data on Y-chromosomal haplogroups.” *Russian Journal of Genetics*, 43(5): 675-687.

Kharkov, V. N. et al. 2005. “Frequency of Y chromosome binary haplogroups in Belarusians.” *Russian Journal of Genetics,* 41(8): 1132-1136.

Kharkov, V. N. et al. 2004. “Gene pool structure of Eastern Ukrainians as inferred from the chromosome haplogroups.” *Russian Journal of Genetics*, 40 (3): 326-331.

Khrunin, A.V. et al. 2005. “Polymorphism of Y-chromosomal microsatellites in Russian populations from the Northern and Southern Russia as exemplified by the populations of Kursk and Arkhangelsk Oblast.” *Russian Journal of Genetics* 41(8): 1125-1131.

Khusnutdinova, E.K. et al. 2008. “Genetic landscape of the Central Asia and Volga-Ural region.” In: *Biosphere Origin and Evolution*. Eds. N. Dobretsov et al. London; New York: Springer. 373-381.

King, R.J. et al. 2008. “Differential Y-chromosome Anatolian influences on the Greek and Cretan Neolithic.” *Annals of Human Genetics* 72: 205-214.

King, Roy and Peter A. Underhill 2002. “Congruent distribution of Neolithic painted pottery and ceramic figurines with Y-chromosome lineages.” *Antiquity* 76: 707-714.

Kivisild, Toomas et al. 2006. “The world mtDNA phylogeny.” In: *Nucleic Acids and Molecular Biology*, 18. Eds. Hans-Jürgen Bandelt et al. Editors. Berlin: Springer Verlag. 149-179.

Kivisild, T. et al. 2003. “The genetic heritage of the earliest settlers persists both in Indian tribal and caste population.” *Am. J. Hum. Genet.* 72: 313-332.

Klassen, Lutz 2002. “The Ertebølle Culture and the Neolithic continental Europe: traces of contact and interaction.” In: *The Neolithisation of Denmark: 150 Years of Debate*. Eds. Anders Fischer and Kristian Kristiansen. Sheffield, England: J.R. Collis Publications. 305-317.

Klopfstein, Seraina 2006. “The fate of mutations surfing on the wave of a range expansion.” *Mol. Biol. Evol*. 23(3): 482-490.

Kniff, Peter de 2000. “Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome.” *Am. J. Hum. Genet.* 67: 1055-1061.

Knight, Alec et al. 2003. “African Y chromosome and mtDNA divergence provides insight into the history of the click languages.” *Current Biology* 13: 464-473.

König, Werner 2007. *dtv-Atlas Deutsche Sprache*, 16th edition. München [Munich]: Deutscher Taschenbuch Verlag.

Kossinna, Gustaf 1896. “Die vorgeschichtliche Ausbreitung der Germanen in Deutschland.” *Zeitschrift des Vereins für Volkskunde*. 6: 1-14.

Kravchenko, S.A. et al. 2002. “Polymorphism of STR loci of the Y chromosome in three populations of Eastern Slavs from Belarus, Russia and Ukraine. *Russian Journal of Genetics,* 39(1): 80-86.

Kristiansen, Kristian. 2002. “The birth of ecological archaeology in Denmark: history and research environments 1850-2000.” In: *The Neolithisation of Denmark: 150 Years of Debate*. Eds. Anders Fischer and Kristian Kristiansen. Sheffield, England: J.R. Collis Publications. 11-31.

Lahermo et al., 1996. “The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA.” *Am. J. Hum. Genet*. 58: 1309-1322.

Lahn, Bruce T. et al. 2001. “The human Y chromosome in light of evolution.” *Nature Reviews Genetics* 2: 207-216.

Laitinen, Virpi et al. 2002. “Y-chromosomal diversity suggests that Baltic males share common Finno-Ugric-speaking forefathers.” *Human Heredity* 53: 68-78.

Lappalainen, T. et al. 2009. “Population structure in contemporary Sweden - a Y-chromosomal and mitochondrial DNA analysis.” *Annals of Human Genetics* 73: 61-73.

Lappalainen, T. et al. 2008. “Migration waves to the Baltic Sea region.” *Annals of Human Genetics* 72: 337-348.

Lappalainen, Tuuli et al. 2006. “Regional differences among the Finns: a Y-chromosome perspective.” *Gene* 376: 207-215.

Lao, Oscar et al. 2008. “Correlation between genetic and geographic structures in Europe.” *Current Biology* 18: 1241-1248.

Loikala, Paula. 1977. “The oldest Germanic loanword in Finnish and their contribution to German philology.” *Studi italiani di linguistica teorica ed applicata*. 6(1-2): 223-250.

López-Parra, A.M. et al. 2009. “In search of the pre- and post-Neolithic genetic substrates in Iberia: evidence from Y-chromosome in Pyrenean populations.” *Annals of Human Genetics* 73: 42-53.

Luca, F. et al. 2007. “Y-chromosomal variation in the Czech Republic.” *American Journal of Physical Anthropology* 132: 132-139.

Lucotte, Gérard et al. 2003. “Y-chromosome DNA haplotype XI in Eastern Europe.” *Human Biology*: 75(3): 405-410.

Lucotte, Gérard et al. 2002. “Haplotypes du chromosome Y en Corse.” *C.R. Biologies* 325: 191-196.

Luis, J.R. et al. 2004. “The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations.” *Am. J. Hum. Genet*. 74: 532-544.

Lynnerup, Niels and Søren Nørby 2004. “The Greenland Norse: bones, graves, computers, and DNA.” *Polar Record* 40 (213): 107-111.

Maca-Meyer, N. et al. 2003. “Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain).” *Annals of Human Genetics* 67: 329-339.

Mailhammer, Robert 2007. The Germanic Strong Verbs: Foundations and development of a New System. Berlin; New York: Mouton de Gruyter.

Malaspina, P. et al. 2001. “A multistep process for the dispersal of Y chromosomal lineage in the Mediterranean area.” *Ann. Hum. Genet*. 65: 339-349.

Malaspina, P. et al. 2000. “Patterns of male-specific inter-population divergence in Europe, West Asia and North Africa.” *Ann. Hum. Genet*. 64: 395-412.

Mallory, J.P. 1989. *In Search of the Indo-Europeans: Language, Archaeology and Myth*. New York: Thames and Hudson.

Malyarchuk, B.A. and M.V. Derenko 2008. “Gene pool structure of Russian populations from the European part of Russia inferred from the data on Y-chromosome haplogroup distributions.” *Russian Journal of Genetics* 44(2): 187-192.

Malyarchuk, Boris et al. 2004. “Differentiation of mitochondrial DNA and Y chromosomes in Russian populations.” *Human Biology* 76(6): 877-900.

Marjanovic, D. et al. 2005. “The peopling of modern Bosnia-Herzegovina: Y-chromosome haplogroups in the three main ethnic groups.” *Annals of Human Genetics* 69: 757-763.

Martinez, Laisel et al. 2007. “Paleolithic Y-haplogroup heritage predominates in Cretan highland plateau.” *European Journal of Human Genetics* 15: 485-493.

McCann, W.J. 1990. “Volk und Germanentum: the presentation of the past in Nazi Germany.” In: *The Politics of the Past*. Eds. Peter Gathercole and David Lowenthal. London; Boston: Unwin Hyman. 74-84.

McEvoy, Brian et al. 2006. “The scale and nature of Viking settlement in Ireland from Y-chromosome admixture analysis.” *European Journal of Human Genetics* 14, 1288-1294.

Melchior, L. 2008. “Rare mtDNA haplogroups and genetics differences in rich and poor Danish Iron-Age Villages.” *American Journal of Physical Anthropology*. 135: 206-215.

Mielke, James H. and Alan G. Fix. 2007. “The confluence of anthropological genetics and anthropological demography.” In: *Anthropological Genetics*. Cambridge University Press, 112-140.

Mirabal, Sheyla et al. 2010. “Human Y-chromosome short tandem repeats: a tale of acculturation and migrations as mechanisms for the diffusion of agriculture in the Balkans Peninsula.” *American Journal of Physical Anthropology* 142: 380-390.

Mirabal, Sheyla et al. 2009. “Y-chromosome distribution within the geo-linguistic landscape of northwestern Russia.” *European Journal of Human Genetics* 17: 1260-1273.

Moore, Andrew M.T. 2003. “The Abu Hureyra project, investigating the beginning of farming in Western Asia.” In: *The Widening Harvest: the Neolithic Transition in Europe: Looking Back, Looking Forward*. Eds. Albert J. Ammerman and Paolo Biagi. Boston: Archaeological Institute of America. 59-74.

Moore, Laoise et al. 2006. “A Y-chromosome signature of hegemony in Gaelic Ireland.” *The American Journal of Human Genetics* 78: 334-338.

Morelli, Laura et al. 2010 “A comparison of Y-chromosome variation in Sardinia and Anatolia is more consistent with culture rather than demic diffusion of agriculture.” *PLoS One* 5(4): e10419.

Murray, Robert W. 2001. “Historical linguistics: the study of language change.” In: *Contemporary Linguistics, an Introduction*. Eds. William O’Grady and John Archibald. Boston; New York: St. Martin’s Press. 289-346.

Myres et al. 2011. “A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe*”* *European Journal of Human Genetics* 19: 95-101.

Nasidze, I. et al. 2008. “Close genetic relationship between Semitic-speaking and Indo-European-speaking groups in Iran.” *Annals of Human Genetics* 72: 241-252.

Nasidze, Ivan et al. 2006(a). “Concomitant replacement of language and mtDNA in South Caspian populations of Iran.” *Current Biology* 16: 668-673.

Nasidze, Ivan et al. 2006(b). “The Gagauz, a linguistic enclave, are not a genetic isolate.” *Annals of Human Genetics* 71: 379-389.

Nasidze, Ivan et al. 2005. “Genetic evidence for the Mongolian ancestry of Kalmyks.” *American Journal of Physical Anthropology* 128: 846-854.

Nasidze, I. et al. 2004(a) “Mitochondrial and Y-chromosome variation in the Caucasus.” *Annals of Human Genetics* 68: 205-221.

Nasidze, I. et al. 2004(b) “Genetic evidence concerning the origins of South and North Ossetians.” *Annals of Human Genetics* 68: 588-599.

Nasidze, Ivan et al. 2003. “Testing hypotheses of language replacement in the Caucasus: evidence from the Y-Chromosome.” *Hum. Genet*. 112: 255-261.

Nebel, Almut. 2005. “Y chromosome evidence for a founder effect in Ashkenazi Jews.” *European Journal of Human Genetics* 13: 388-391.

Nebel, Almut et al. 2002. “Genetic evidence for the expansion of Arabian tribes into the southern Levant and North Africa.” *Am. J. Hum. Genet*. 70: 1594-1596.

Nelson, Max 2005*. The Barbarian’s Beverage: a History of Beer in Ancient Europe.* London; New York: Routledge, 2005.

Nichols, Johanna. 2007. “Language dispersals from the Black Sea region.” In: *The Black Sea Flood Question: Changes in Coastline, Climate, and Human Settlement*. Eds. Valentina Yanko-Hombach et al. Dordrecht: Springer. 775-796.

Nielsen, Hans Frede 1989. *The Germanic Languages: Origins and Early Dialectal Interrelations*. Tuscaloosa: University of Alabama Press.

Novelletto, Andrea 2007. "Y-chromosome variation in Europe: continental and local processes in the formation of the extant gene pool." *Annals of Human Biology,* 34(2): 139-172.

Oota, Hiroki et al. 2005. "Recent origin and cultural reversion of a hunter-gather group." *PLoS Biology* 3(3): e71.

Otto, Karl Heinz 1988. "Zur Forschungsgeschichte - allgemeine Übersicht." In: *Die Germanen, Geschichte und Kultur der germanischen Stämme in Mitteleuropa*, Band 1, ausgearbeitet von einem Autorenkollektiv unter Leitung von Bruno Krüger. Berlin: Akademie-Verlag. 13-36.

Owens, Francis 1960. *The Germanic People - Their Origin, Expansion and Culture*. New York: Bookman Associates.

Özdoğan, Mehmet. 2007. "Coastal changes of the Black Sea and Sea of Marmara in archaeological perspective." In: *The Black Sea Flood Question: Changes in Coastline, Climate, and Human Settlement*. Eds. Valentina Yanko-Hombach et al. Dordrecht: Springer. 651-669.

Pakendorf, Brigitte and Mark Stoneking 2005. "Mitochondrial DNA and human evolution." *Annu. Rev. Genomics Hum. Genet*. 6: 165-183.

Pala, Maria et al. 2009. "Mitochondrial haplogroup U5b3: a distant echo of the Epipaleolithic in Italy and the legacy of the early Sardinians." *The American Journal of Human Genetics* 84: 814-821.

Palo, Jukka U. 2009. "Genetic markers and population history: Finland revisited." *European Journal of Human Genetics* 17: 1336-1346.

Passarino, Giuseppe et al. 2002. "Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms." *European Journal of Genetics* 10: 521-529.

Passarino, Giuseppe et al. 2001. "The 49a,f haplotype 11 is a new marker of the EU 19 lineage that traces migrations from northern regions of the Black Sea." *Human Immunology* 62: 922-932.

Pedersen, Holger 1967. *The Discovery of Language: Linguistic Science in the Nineteenth Century*. Translated by John Webster Spargo. Bloomington: Indiana University Press.

Peña, José A. et al. 2006. "Gene flow in the Iberian peninsula determined from Y-chromosome STR loci." *American Journal of Human Biology* 18: 532-539.

Pereira, Luísa et al. 2005. "High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium." *Genome Research* 15: 19-24.

Pereira, Luísa et al. 2001. "Y-chromosome mismatch distributions in Europe." *Mol. Biol. Evol.* 19(7): 1259-1271.

Peričić, Marijana et al. 2005(a). "Review of Croatian genetic heritage as revealed by mitochondrial DNA and Y-chromosomal lineages." *Croat. Med. J.* 46(4): 502-513.

Peričić, Marijana et al. 2005(b) "High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations." *Mol. Biol. Evol*. 22(10): 1964-1975.

Peričić, Marijana et al. 2005(c) "Population genetics of 8 Y chromosome STR loci in Macedonians and Macedonian Romani (Gypsy)." *Forensic Science International* 154: 257-261.

Pimenoff, Ville N. 2008. "Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers." *European Journal of Human Genetics* 16: 1254-1264.

Price, T. Douglas 2003. "The arrival of agriculture in Europe as seen from the north." In: *The Widening Harvest: the Neolithic Transition in Europe: Looking Back, Looking Forward*. Eds. Albert J. Ammerman and Paolo Biagi. Boston: Archaeological Institute of America. 273-296.

Price, T. Douglas and Anne Birgitte Gebauer 1995. *Last Hunters, First Farmers: New Perspectives on the Prehistoric Transition to Agriculture*. Santa Fe, New Mexico: School of American Research Press.

Pringle, Heather 2006. *The Master Plan: Himmler's Scholars and the Holocaust*. New York: Hyperion.

Prokosch, E. 1939. *A Comparative Germanic Grammar*. Philadelphia: Linguistic Society of America.

Pyles, Thomas and John Algeo 1993. *The Origins and Development of the English Language* 4th edition. Boston: Heinle Publishing.

Quintana-Murci, Luís et al. 2001. "Y-chromosome lineages trace diffusion of people and languages in southwestern Asia." *Am. J. Hum. Genet*. 68: 537-542.

Race, Ethnicity, and Genetics Working Group 2005. "The use of racial, ethnic, and ancestral categories in human genetics research." *Am. J. Hum. Genet*. 7: 519-532.

Raitio, Mirja et al. 2001. "Y-chromosomal SNPs in Finno-Ugric speaking populations analyzed by minisequencing on microarrays." *Genome Res*. 11: 471-482.

Ramachandran, Sohini et al. 2005. "Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa." *PNAS* 102 (44): 15942-15947.

Rauch, Irmengard 2003. *The Gothic Language*. New York: Peter Lang Publishing.

Rębała, Krzysztof et al. 2007. “Y-STR variation among Slavs: evidence for the Slavic homeland in the middle Dnieper basin.” *J. Hum. Genet.* 52: 406-414.

Regueiro, M. et al. 2006. “Iran: tricontinental nexus of Y-chromosome driven migration. *Human Heredity* 61: 132-143.

Renfrew, Colin 2003. “Time depth, convergence theory, and innovation in proto-Indo-European: ‘Old Europe’ as a PIE linguistic area.” In: *Languages in Prehistoric Europe*. Eds. Alfred Bammesberger and Theo Vennemann. Heidelberg: Winter. 11-48.

Renfrew, Colin 1987. *Archaeology and Language: the Puzzle of Indo-European Origins*. Cambridge; New York: Cambridge University Press.

Richards, Martin et al. 2006 “A model for the dispersal of modern humans out of Africa.” In: *Nucleic Acids and Molecular Biology* 18. Eds. Hans-Jürgen Bandelt et al. Berlin: Springer Verlag. 225-265.

Richards, Martin. 2003. “The Neolithic invasion of Europe.” *Annual Review of Anthropology* 32: 135-162.

Richards, Martin et al. 2002. “In search of geographical patterns in European mitochondrial DNA.” *Am. J. Hum. Genet* 71: 1168-1174.

Richards, Martin et al. 2000. “Tracing European founder lineages in the Near Eastern mtDNA pool.” *Am. J. Hum. Genet.* 67: 1251-1276.

Richards, M.B. et al. 1998. “Phylogeography of mitochondrial DNA in western Europe.” *Ann. Hum. Genet*. 62: 241-260.

Richards, Michael P. et al. 2003. “Mesolithic and Neolithic subsistence in Denmark: new stable isotope data.” *Current Anthropology* 44: 288-295.

Ringe, Don 2006. *From Proto-Indo-European to Proto-Germanic*. Oxford; New York: Oxford University Press.

Roewer, Lutz 2005. “Signature of recent historical events in the European Y-chromosomal STR haplotype distribution.” *Hum. Genet* 116: 279-291.

Rootsi, Siiri et al. 2007. “A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe.” *European Journal of Human Genetics* 15: 204-211.

Rootsi, Siri et al. 2004. “Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe.” *Am. J. Hum. Genet* 75: 128-137.

Rossem, Cefas van and Hein van der Voort, eds. 1996. *Die Creol Taal: 250 Years of Negerhollands Texts*. Amsterdam: Amsterdam University Press.

Rosser, Zoë H. et al. 2000. “Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than language.” *Am. J. Hum. Genet*. 67: 1526-1543.

Rudbeck, Lars et al. 2005. “mtDNA analysis of human remains from an early Danish Christian cemetery.” *American Journal of Physical Anthropology,* 128: 424-429.

Rubicz, Rohina C. et al. 2007 “Molecular markers in anthropological genetic studies.” In: *Anthropological Genetics*. Ed. Michael H. Crawford. Cambridge; New York: Cambridge University Press. 141-186.

Ryan, William B.F. et al. 1997(a). “An abrupt drowning of the Black Sea shelf.” *Marine Geology* 138: 119-126.

Ryan, William B.F. et al. 1997(b). “An abrupt drowning of the Black Sea shelf at 7.5 KRY.” *Geo-Eco-Marina* 2: 115-125.

Ryan, William B.F. et al. 2003. “Catastrophic Flooding of the Black Sea.” *Annu. Rev. Earth Planet Sci.* 31:525-554.

Sahoo, Sanghamitra et al. 2006. “A prehistory of Indian Y-chromosomes: evaluating demic diffusion scenarios.” *PNAS* 103(3): 843-848.

Sajantila, Antti and Svante Pääbo 1995. “Language Replacement in Scandinavia.” *Nature Genetics* 11: 359-360.

Salmela, Elina 2008. “Genome-wide analysis of single nucleotide polymorphisms uncovers population structure in Northern Europe.” *PloS One* 3(10): e3519

Sampietro, M.L. et al. 2007. “Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe.” *Proc. R. Soc. B.* 274: 2161-2167.

Sanchez, Juan J. et al. 2003. “Multiplex PCR and minisequencing of SNPs - a model with 35 Y chromosome SNPs.” *Forensic Science International* 137: 74-84.

Santos-Lopez, Simone S. et al. “A worldwide phylogeny for the human X chromosome.” *PloS One* 2(6): e557.

Scarre, Christopher. 2005(a) “The world transformed: from foragers and farmers to states and empires.” In: *The Human Past, World Prehistory and the Development of the Human Societies*. Ed. Christopher Scarre. London: Thames and Hudson. 176-199.

Scarre, Christopher. 2005(b) “Holocene Europe.” In: *The Human Past, World Prehistory and the Development of the Human Societies*. Ed. Christopher Scarre. London: Thames and Hudson. 392-431.

Schaffner, Stephen F. 2004. “The X chromosome in population genetics.” *Nature Reviews Genetics* 5: 43-51.

Schirmer, Alfred and Walther Mitzka 1969. *Deutsche Wortkunde* 6th edition. Berlin: Walter de Gruyter.

Schleicher, August 1863. *Die Darwinsche Theories und die Sprachwissenschaft: Offenes Sendsschreiben an Herrn Dr. Ernst Häcket, a. o. Professor der Zoologie und Director des zoologischen Museums and Universität Jena*. Weimar: Hermann Böhlau.

Schleicher, August 1862. *Compendium der Vergleichenden Grammatik der Indogermanischen Sprachen* Band 2. Weimar: Hermann Böhlau.

Schleicher, August 1861. *Compendium der Vergleichenden Grammatik der Indogermanischen Sprachen* Band 1. Weimar: Hermann Böhlau, Weimar.

Schmidt, Ulf 2004. *Justice at Nuremberg: Leo Alexander and the Nazi Doctors' Trial***.** New York: Palgrave MacMillan.

Schmidt, Johannes 1872. *Die Verwandtschaftsverhältnisse der Indogermanischen Sprachen*. Weimar: Hermann Böhlau.

Schultze, Erdmute 1986. "Die Sprachentwicklung bis zur Herausbildung des althochdeutschen Wortschatzes." In*: Die Germanen, Geschichte und Kultur der germanischen Stämme in Mitteleuropa* Band 2. Ausgearbeitet von einem Autorenkollektiv unter Leitung von Bruno Krüger. Berlin: Akademie-Verlag. 315-335.

Scozzari, Rosaria et al. 2001. "Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region." *Human Immunology* 62: 871-884.

Semino, Ornella et al 2004. "Origin, Diffusion, and Differentiation of Y-Chromosome Haplogroups E and J: inferences on the Neolithization of Europe and later migratory events in the Mediterranean area." *Am. J. Hum. Genet*. 74: 1023-1034.

Semino, Ornella et al. 2000(a) "The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective." *Science* 290: 1155-1159.

Semino, Ornella et al. 2000(b) "MtDNA and Y-chromosome polymorphisms in Hungary: inferences from Paleolithic, Neolithic, and Uralic influences on the modern Hungarian gene pool." *European Journal of Human Genetics* 8: 339-346.

Sengupta, Sanghamitra et al. 2006. "Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveals minor genetic influence of central Asian pastoralists." *The American Journal of Human Genetics* 78: 201-221.

Seyer, Rosemarie 1988. "Antike Nachrichten." In: *Die Germanen, Geschichte und Kultur der germanischen Stämme in Mitteleuropa*. Band 1. Ausgearbeitet von einem Autorenkollektiv unter Leitung von Bruno Krüger. Berlin: Akademie-Verlag. 37-55.

Shen, Peidong et al. 2000. "Population genetic implications from sequence variation in four Y chromosome genes." *PNAS* 97(13): 7354-7359.

Simoni, Lucia et al. 2000. “Geographic patterns of mtDNA diversity in Europe.” *Am. J. Hum. Genet.* 66: 262-278.

Sims, Lynn M. et al. 2009. “Improved resolution haplogroup G phylogeny in the Y chromosome, revealed by a set of newly characterized SNP’s. *PLoS One* 4(6): e5792.

Sjödin, Per and Oliver François (2011). “Wave-of-Advance Models of the Diffusion of the Y Chromosome Haplogroup R1b1b2 in Europe.” *PLoS One* 6(6): e21592.

Smolla, Günter 1980. “Das Kossinna-Syndrom.” *Fundberichte aus Hessen* 19/20: 1-9.

Soares, Pedro et al 2010. “The archaeogenetics of Europe.” *Current Biology* 20: R174-R183.

Soares, Pedro et al. 2009. “Correcting for purifying selection: an improved human mitochondrial molecular clock.” *The American Journal of Human Genetics* 84: 740-759.

Stefan, Mihaela et al. 2001. “Y chromosome analysis reveals a sharp genetic boundary in the Carpathian region.” *European Journal of Human Genetics* 9: 27-33.

Tambets, Kristina et al. 2004. “The western and eastern roots of the Saami - the story of genetic ‘outliers’ told by the mitochondrial DNA and Y Chromosome.” *Am. J. Hum. Genet.* 74: 661-682.

Templeton, Alan R. 2002. “Out of Africa again and again.” *Nature* 416: 45-51.

Tetushkin, E. Ya. 2001. “Genetics and the origins of the human races.” *Russian Journal of Genetics* 37( 8): 853-867.

Thomas, Mark G. et al. 2006 “Evidence for an apartheid-like social structure in early Anglo-Saxon England.” *Proc. R. Soc. B.* 273: 2651-2657.

Thomason, Sarah G. 2001. *Language Contact: an Introduction*. Washington, DC: Georgetown University Press.

Thomason, Sarah Grey and Terrence Kaufman 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.

Thomson, Russell et al. 2000. “Recent common ancestry of Y chromosomes: evidence from DNA sequence data.” *PNAS* 97(13): 7360-7365.

Todd, Malcolm. 1992. *The Early Germans*. Malden, MA: Blackwell Publishing.

Tofanelli, Sergio et al. 2009. “J1-M267 Y lineage marks climate-driven pre-historical human displacements.” *European Journal of Human Genetics* 17: 1520-1524.

Torroni, Antonio. 2006 “Harvesting the fruit of the human DNA tree.” *Trends in Genetics* 22(6): 339-345.

Torroni, Antonio et al. 2001. “A signal, from human mtDNA, of post-glacial recolonization in Europe.” *Am. J. Hum. Genet.* 69: 844-852.

Trask, R.L. 1996. *Historical Linguistics*. London: Arnold.

Trubetzkoy, N.S. 1939 “Gedanken über das Indogermanischenproblem.” *Acta Linguistica* (1): 81-89.

Trubetzkoy, N.S. 1929. “Proposition 16.” *Actes du Premier Congrès International de Linguiste á la Haye. Du 10-15 Avril 1928.* Leiden: Sijthoff. 18.

Underhill, Peter A. et al. 2010. “Separating the post-Glacial coancestry of European and Asia Y chromosomes within haplogroup R1a.” *European Journal of Human Genetics* 18: 479-484.

Underhill, Peter A. et al. 2007. “New phylogenetic relationships for Y-chromosome haplogroup I.” In: *Rethinking the Human Revolution*. Eds. P. Mellars et al. Cambridge: McDonald Institute for Archaeological Research. 33- 42.

Underhill, Peter A. and Toomas Kivisild 2007. “Use of Y-Chromosome and Mitochondrial DNA population structure in tracing human migrations.” *Annu. Rev. Genet*. 41: 539-564.

Underhill, P.A. et al. 2001. “The phylogeography of Y chromosome binary haplogroups and the origins of modern human populations.” *Ann. Hum. Genet* 65: 43-62.

Underhill, P.A. et al. 2000. “Y chromosome sequence variation and the history of human populations.” *Nature Genetics* 26: 358-361.

Veit, Ulrich 2002. “From nationalism to nazism: Gustaf Kossinna and his concept of a national archaeology.” In: *Gesellschaften und Staaten im Epochenwandel* Band 7. Archaeology, Ideology and Society: the German Experience. Ed. Heinrich Härke. Frankfurt am Main; New York: Peter Lang. 41-66.

Vennemann, Theo 2003. *Europa Vasconica, Europa Semitica*. Berlin; New York: Mouton de Gruyter.

Vennemann, Theo 2000. “Zur Entstehung des Germanischen.” *Sprachwissenschaft* 25(3): 233-269.

Vennemann, Theo 1994. “Linguistic reconstruction in the context of European prehistory. ” *Transactions of the Philological Society*, 92(2): 215-284.

Villems, Richard et al. 2002. “Archaeogenetics of Finno-Ugric speaking populations.” In: *The Roots of Peoples and Languages of Northern Eurasia IV.* Ed. Kyösti Julku. Turku: Societas Historiae Fenno-Ugricae 271-284.

Völgyi, Antonia et al. 2008. "Haplogroup distribution of Hungarian population and the largest minority group." *Forensic Science International Genetics Supplement Series* 1: 383-385.

Walsh, Bruce 2001. "Estimating the time to the most recent common ancestor for Y chromosome or mitochondrial DNA for a pair of individuals. *Genetics* 158: 897-912.

Waterman, John T. 1976. *A History of the German Language: with Special Reference to the Cultural and Social Forces that Shaped the Standard Literary Language.* Seattle: University of Washington Press.

Waterman, John T. 1970. *Perspectives in Linguistics*. Chicago: University of Chicago Press.

Watkins, Trevor 2005. "From foragers to complex societies in Southwest Asia." In: *The Human Past, World Prehistory and the Development of the Human Societies*. Ed. Christopher Scarre. London: Thames and Hudson. 200-233.

Weale, Michael E. et al. 2002. "Y chromosome evidence for Anglo-Saxon mass migration." *Mol. Biol. Evol.* 19(7): 1008-1021.

Weale, Michael E. et al. 2001 "Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group." *Hum. Genet*. 109: 659-674.

Weindling, Paul Julian 2004. *Nazi Medicine and the Nuremberg Trials: From Medical War Crimes to Informed Consent*. Houndmills, Basingstoke, Hampshire; New York: Palgrave MacMillan.

Wells, R. Spencer et al. 2001. "The Eurasian heartland: a continental perspective on Y-chromosome diversity." *PNAS* 98(18): 10244-10249.

Wilder, Jason A. et al. 2004. "Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males." *Nature Genetics* 36(10): 1122-1125.

Wiik, Kavevi 2008. "Where did European men come from?" *Journal of Genetic Genealogy* 4: 35-85.

Wiik, Kalevi 2003. "Finnic-Type Pronunciation in the Germanic languages." *The Mankind Quarterly* 44(1): 43-90.

Wiik, Kalevi 1997. "The Uralic and Finno-Ugric substratum in Proto-Germanic." *Linguistica Uralica,* 33(4): 258-280.

Wilkins, Jon F. 2006. "Unraveling male and female histories from genetic data." *Current Opinion in Genetics & Development* 16: 611-617.

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

Winford, Donald 2003. *An Introduction to Contact Linguistics*. Malden, MA: Blackwell.

Wolfram, Saine 2000. "'Vorsprung durch Technik' or 'Kossinna Syndrom.'" In: *Gesellschaften und Staaten im Epochenwandel* Band 7. Archaeology, Ideology and Society: the German Experience. Ed. Heinrich Härke. Frankfurt am Main; New York: Peter Lang. 183-204.

Wood, Elizabeth T. 2005. "Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biases demographic processes." *European Journal of Human Genetics* 13: 867-876.

Woźniak, Marcin et al. 2010. "Similarities and distinctions in Y chromosome gene pool of Western Slavs." *American Journal of Physical Anthropology,* 142: 540-548.

Xiao, Feng-Xia et al. 2004. "Human X-chromosomal lineages in Europe reveal Middle Eastern and Asiatic contacts." *European Journal of Human Genetics* 12: 301-311.

The Y Chromosome Consortium 2002. "A nomenclature system for the tree of human Y-chromosome binary haplogroups." *Genome Research* 12: 339-348.

Yotova, Vania et al. 2007. "Tracing genetic history of modern humans using X-chromosome lineages." *Hum. Genet.* 122: 431-443.

Zalloua, Pierre A. et al. 2008(a). "Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean." *The American Journal of Human Genetics* 83: 633-642.

Zalloua, Pierre A. et al. 2008(b). "Y-chromosome diversity in Lebanon is structured by recent historical events." *The American Journal of Human Genetics* 82: 873-882.

Zei, Gianna et al. 2003. "From surnames to the history of Y chromosomes: the Sardinian population as a paradigm." *European Journal of Human Genetics* 11: 802-807.

Zerjal, Tatiana et al. 2002. "A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia." *Am. J. Hum. Genet.* 71: 466-482.

Zerjal, Tatiana et al. 2001. "Geographical, linguistic, and cultural influence on genetic diversity: Y-chromosome distributions in Northern European populations." *Mol. Biol. Evol.* 18(6): 1077-1087.

Zerjal, Tatiana et al. 1997. "Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* 60: 1174-1183.

Zhivotovsky, Lev. A. and Peter A. Underhill 2004. "On the evolutionary mutation rate at Y-chromosome STRs: comments on paper by Di Giacomo et al." *Hum. Genet.* 117: 529-532.

Zhivotovsky, Lev. A. et al. 2004. "The effective mutation rate at Y-chromosome short tandem repeats, with application to human population-divergence time." *Am. J. Hum. Genet.* 74: 50-61.

## Appendix Table 1: Western European R-Group (R1b1b2-M269)

### Western Europe

| **Population Studied** | **Nomenclature used for Western European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Basque Region of Europe | Haplogroup 1 | 73.0% | Rosser et al. 2000 |
| Basque Region of France | Eu 18 | 86.4% | Semino et al. 2000(a) |
| Basque Region of Spain | Eu 18 | 88.9% | Semino et al. 2000(a) |
| Basque Region of France | R1b1b2-M269 | 61.0% | Balaresque et al. 2010 |
| Basque Region of Spain | R1b1b2-M269 | 87.1% | Balaresque et al. 2010 |
| Belgium | Haplogroup 1 | 63.0% | Rosser et al. 2000 |
| France | Haplogroup 1 | 50.0% | Rosser et al. 2000 |
| France | Haplogroup 1 | 54.8% | Scozzari et al. 2001 |
| France | Eu 18 | 52.2% | Semino et al. 2000(a) |
| France | R1b | 59.0% | Tambets et al. 2004 |
| France | R1b1b2-M269 | 74.0% | Balaresque et al. 2010 |
| Iceland | Haplogroup 1 | 41.4% | Helgason et al. 2000 |
| Iceland | Haplogroup 1 | 46.0% | Rosser et al. 2000 |
| Ireland | Haplogroup 1 | 81.5% | Helgason et al. 2000 |
| Ireland | Haplogroup 1 | 81.0% | Rosser et al. 2000 |
| Ireland | R1xR1a1/<br>AMH +1 | 87.4% | Capelli et al. 2003 |
| Ireland | R1b3 | 85.4% | Moore et al. 2006 |

## Appendix Table 1: Western European R-Group (R1b1b2-M269)

| | | | |
|---|---|---|---|
| Netherlands | Haplogroup 1 | 43.0% | Rosser et al. 2000 |
| Netherlands | Eu 18 | 70.4% | Semino et al. 2000(a) |
| Netherlands | R1b1b2-M269 | 42.0% | Balaresque et al. 2010 |
| Portugal (southern) | R1* | 50.5% | Flores et al. (2004) |
| Portugal (southern) | Haplogroup 1 | 56.0% | Rosser et al. 2000 |
| Portugal (northern) | Haplogroup 1 | 62.0% | Rosser et al. 2000 |
| Portugal | R1b3* (xR1b3f) | 57.7% | Beleza et al. 2006 |
| Portugal (southern) | R1b1b2-M269 | 46.2% | Balaresque et al. 2010 |
| Spain | Haplogroup 1 | 68.0% | Rosser et al. 2000 |
| Spain (Catalonia) | Eu 18 | 79.2% | Semino et al. 2000(a) |
| Spain | Haplogroup 1 | 55.6% | Scozzari et al. 2001 |
| Spain | R1b1b2-M269 | 55.6% | López-Parra et al. 2009 |
| Spain | R1b1b2-M269 | 69.0% | Balaresque et al. 2010 |
| Iberia | R1 | 52.1% | Flores et al. (2004) |
| Iberia | R1b3 | 55.0% | Adams et al. 2008 |
| United Kingdom (British) | Haplogroup 1 | 68.8% | Helgason et al. 2000 |
| United Kingdom (Scotland) | Haplogroup 1 | 77.1% | Helgason et al. 2000 |
| United Kingdom (Western Scotland) | Haplogroup 1 | 72.0% | Rosser et al. 2000 |
| United Kingdom (Scotland) | Haplogroup 1 | 79.0% | Rosser et al. 2000 |
| United Kingdom (Cornwall) | Haplogroup 1 | 82.0% | Rosser et al. 2000 |
| United Kingdom (East Anglia) | Haplogroup 1 | 56.0% | Rosser et al. 2000 |
| United Kingdom | M173 | 72.0% | Wells et al. 2001 |

# Appendix Table 1: Western European R-Group (R1b1b2-M269)

| | | | |
|---|---|---|---|
| United Kingdom | R1xR1a1/<br>AMH +1 | 70.4% | Capelli et al. 2003 |
| United Kingdom (Cornwall) | R1b1b2-M269 | 64.0% | Balaresque et al. 2010 |
| United Kingdom (Leicestershire) | R1b1b2-M269 | 43.0% | Balaresque et al. 2010 |
| United Kingdom (Wales) | R1b1b2-M269 | 92.3% | Balaresque et al. 2010 |

**Mediterranean**

| **Population Studied** | **Nomenclature used for Western European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Corsica | Haplogroup 1 | 48.9% | Scozzari et al. 2001 |
| Crete | R1b3-M269 | 14.9% | Martinez et al. 2007 |
| Crete | R1b3-M269 | 17.0% | King et al. 2008 |
| Cyprus | Haplogroup 1 | 9.0% | Rosser et al. 2000 |
| Cyprus | R1(xR1a1) | 9.2% | Capelli et al. 2006 |
| Cyprus | R1(xR1a1) | 10.2% | El-Sabai et al. 2009 |
| Greece | Haplogroup 1 | 14.3% | Helgason et al. 2000 |
| Greece | Haplogroup 1 | 11.0% | Rosser et al. 2000 |
| Greece | Eu 18 | 27.6% | Semino et al. 2000(a) |
| Greece | P (xR1a) | 12.8% | Di Giacomo et al. 2003 |
| Greece | Haplotype XV | 3.8% | Lucotte et al. 2003 |
| Greece | R1b-PN25 | 12.2% | Bosch et al. 2006 |
| Greece | R1b3-M269 | 13.0% | King et al. 2008 |
| Italy | Haplogroup 1 | 35.8% | Helgason et al. 2000 |

## Appendix Table 1: Western European R-Group (R1b1b2-M269)

| | | | |
|---|---|---|---|
| Italy | Haplogroup 1 | 44.0% | Rosser et al. 2000 |
| Italy (north - central) | Eu 18 | 62.0% | Semino et al. 2000(a) |
| Italy (Calabria) | Eu 18 | 32.4% | Semino et al. 2000(a) |
| Italy | Haplogroup 1 | 34.2% | Scozzari et al. 2001 |
| Italy | P (xR1a) | 36.4% | Di Giacomo et al. 2003 |
| Italy (southern) | R1(xR1a1) | 25.0% | Capelli et al. 2006 |
| Italy | R1 (xR1a1) | 40.0% | Capelli et al. 2007 |
| Northeast Italy | R1b1b2-M269 | 60.8% | Balaresque et al. 2010 |
| Northwest Italy | R1b1b2-M269 | 45.0% | Balaresque et al. 2010 |
| Malta | R1(xR1a1) | 32.2 | Capelli et al. 2006 |
| Sardinia | Haplogroup 1 | 30.0% | Rosser et al. 2000 |
| Sardinia | Eu 18 | 22.1% | Semino et al. 2000(a) |
| Sardinia | Haplogroup 1 | 20.2 | Scozzari et al. 2001 |
| Sardinia | R-M269 | 20.8% | Zei et al. 2003 |
| Sardina | R1(xR1a1) | 21.0% | Capelli et al. 2006 |
| Sardinia | M269 | 17.0% | Contu et al. 2008 |
| Sicily | Haplogroup 1 | 30.0% | Scozzari et al. 2001 |
| Eastern Sicily | R1(xR1a1) | 19.5% | Capelli et al. 2006 |
| Southwest Sicily | R1(xR1a1) | 29.1% | Capelli et al. 2006 |
| Northwest Sicily | R1(xR1a1) | 25.7% | Capelli et al. 2006 |
| Sicily | R1b1c-M269 | 24.6% | Di Gaetano et al. 2009 |

## Appendix Table 1: Western European R-Group (R1b1b2-M269)

### Central Europe

| **Population Studied** | **Nomenclature used for Western European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Czech Republic | Haplogroup 1 | 19.0% | Rosser et al. 2000 |
| Czech Republic and Slovakia | Eu 18 | 35.6% | Semino et al. 2000(a) |
| Czech Republic | Haplotype XV | 27.9% | Lucotte et al. 2003 |
| Czech Republic | P-DYS257 (xR1a) | 28.0% | Luca et al. 2007 |
| Germany | Haplogroup 1 | 46.9% | Helgason et al. 2000 |
| Germany (Bavaria) | Haplogroup 1 | 48.0% | Rosser et al. 2000 |
| Germany | Haplogroup 1 | 40.0% | Rosser et al. 2000 |
| Germany | Eu 18 | 50.0% | Semino et al. 2000(a) |
| Germany | R1(xRa1)<br>(M173) | 38.9% | Kayser et al. 2005 |
| Germany (Bavaria) | R1b1b2-M269 | 32.3% | Balaresque et al. 2010 |
| Hungary | Haplogroup 1 | 30.0% | Rosser et al. 2000 |
| Hungary | EU18 | 13.3% | Semino et al. 2000(a) |
| Hungary | Haplotype XV | 9.3% | Lucotte et al. |
| Hungary | R1b | 20.4% | Tambets et al. 2004 |
| Hungary | R1b-P25 | 16.0% | Völgyi et al. 2008 |
| Slovakia | Haplogroup 1 | 17.0% | Rosser et al. 2000 |
| Slovakia | Haplotype XV | 0.4% | Lucotte et al. 2003 |

# Appendix Table 1: Western European R-Group (R1b1b2-M269)

## Balkans

| **Population Studied** | **Nomenclature used for Western European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Albania | EU18 | 17.6% | Semino et al. 2000(a) |
| Albania | Haplotype XV | 14.7% | Lucotte et al. 2003 |
| Albania | R1b-M173 | 21.1% | Peričić et al. 2005(b) |
| Albania | R1b-PN15 | 13.3% | Bosch et al. 2006 |
| Bosnia-Herzegovina | R1b | 3.9% | Marjanovic et al. 2005 |
| Bosnia-Herzegovina (Bosnia) | R1b-M173 | 15.7% | Peričić et al. 2005(b) |
| Bosnia-Herzegovina (Herzegovina) | R1b-M173 | 1.4% | Peričić et al. 2005(b) |
| Croatia | Eu 18 | 10.3% | Semino et al. 2000(a) |
| Croatia | R1b | 7.9% | Barać et al. 2003 |
| Croatia | R1b | 7.5% | Peričić et al. 2005(a) |
| Croatia | R1b-M173 | 15.7% | Peričić et al. 2005(b) |
| Macedonia | Eu 18 | 10.0% | Semino et al. 2000(a) |
| Macedonia | R1b-M173 | 10.6% | Peričić et al. 2005(b) |
| Macedonia | R1b-PN25 | 13.5% | Bosch et al. 2006 |
| Serbia | R1b-M173 | 3.6% | Peričić et al. 2005(b) |
| Serbia | R1b1b2-M269 | 10.0% | Balaresque et al. 2010 |
| Slovenia | Haplogroup 1 | 21.0% | Rosser et al. 2000 |
| Slovenia | R1b1b2-M269 | 20.6% | Balaresque et al. 2010 |

## Appendix Table 1: Western European R-Group (R1b1b2-M269)

| | | | |
|---|---|---|---|
| Yugoslavia | Haplogroup 1 | 11.0% | Rosser et al. 2000 |
| Yugoslavia | Haplotype XV | 10.0% | Lucotte et al. 2003 |

### Middle East

| **Population Studied** | **Nomenclature used for Western European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Iran (Tehran) | M173 | 4.0% | Wells et al. 2003 |
| Iran (northern) | R1b1a-M269 | 15.2% | Regueiro et al. 2006 |
| Iran (southern) | R1b1a-M269 | 6.0% | Regueiro et al. 2006 |
| Iraq | R-M269 | 10.8% | Al-Zahery et al. 2003 |
| Jordan | R-M173 | 7.9% | Flores et al. 2005 |
| Jordan | R1(xR1a1) | 9.0% | El-Sabai et al. 2009 |
| Kuwait | R1(xR1a1) | 9.5% | El-Sabai et al. 2009 |
| Lebanon | Eu 18 | 6.4% | Semino et al. 2000(a) |
| Lebanon | M173 | 6.0% | Wells et al. 2001 |
| Lebanon | R1b | 7.9% | Zalloua et al. 2008(b) |
| Lebanon | R1(xR1a1) | 7.9% | El-Sabai et al. 2009 |
| Qatar | R1b1a-M269 | 1.4% | Cadenas et al. 2008 |
| Qatar | R1(xR1a1) | 1.4% | El-Sabai et al. 2009 |
| Syria | Eu 18 | 15.0% | Semino et al. 2000(a) |
| Syria | R1(xR1a1) | 4.5% | El-Sabai et al. 2009 |
| Turkey | Haplogroup 1 | 20.0% | Rosser et al. 2000 |
| Turkey | Eu 18 | 6.6% | Semino et al. 2000(a) |
| Turkey (Istanbul) | Haplotype XV | 7.8% | Lucotte et al. 2003 |
| Turkey (Ankara) | Haplotype XV | 2.6% | Lucotte et al. 2003 |

## Appendix Table 1: Western European R-Group (R1b1b2-M269)

| | | | |
|---|---|---|---|
| Turkey | R1b3-M269 | 14.7% | Cinnioğlu et al. 2004 |
| United Arab Emirites | R1b1a-M269 | 3.7% | Cadenas et al. 2008 |
| United Arab Emirites | R1(xR1a1) | 4.3% | El-Sabai et al. 2009 |

### South Central Asia

| **Population Studied** | **Nomenclature used for Western European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| India | R1b | 7.0% | Kivisbild et al. 2003 |
| India | R1b3-M269 | 0.6% | Sengupta et al. 2006 |
| Pakistan | R1b3-M269 | 2.8% | Sengupta et al. 2006 |

### Northern Europe

| **Population Studied** | **Nomenclature used for Western European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Denmark | Haplogroup 1 | 41.7% | Helgason et al. 2000 |
| Denmark | Haplogroup 1 | 50.0% | Rosser et al. 2000 |
| Denmark | Haplogroup 1 | 57.1% | Scozzari et al. 2000 |
| Denmark/Schleswig Holstein | R1xR1a1/ AMH +1 | 39.0% | Capelli et al. 2003 |
| Denmark | R1b | 36.1% | Tambets et al. 2004 |
| Denmark | R1b1b2-M269 | 42.9% | Balaresque et al. 2010 |
| Denmark | R1b1b-M269 | 36.3% | Myres et al. 2011 |
| Finland | Haplogroup 1 | 2.0% | Rosser et al. 2000 |

## Appendix Table 1: Western European R-Group (R1b1b2-M269)

| | | | |
|---|---|---|---|
| Finland | Haplogroup 1 | 0.0% | Zerjal et al. 2001 |
| Finland | R1b | 3.7% | Lappalainen et al. 2006 |
| Finland (eastern) | R1b | 2.6% | Lappalainen et al. 2008 |
| Finland (western) | R1b | 5.2% | Lappalainen et al. 2008 |
| Norway | Haplogroup 1 | 25.9% | Helgason et al. 2008 |
| Norway | Haplogroup 1 | 29.0% | Rosser et al. 2000 |
| Norway | Haplogroup 1 | 29.0% | Zerjal et al. 2001 |
| Norway | Eu 18 | 27.8% | Passarino et al. 2002 |
| Norway | R1xR1a1/ AMH +1 | 30.0% | Capelli et al. 2003 |
| Norway | P(xR1a) | 31.3% | Dupuy et al. 2006 |
| Saami | Haplogroup 1 | 6.0% | Rosser et al. 2000 |
| Saami | Eu 18 | 8.3% | Semino et al. 2000(a) |
| Saami | Haplogroup 1 | 6.0% | Zerjal et al. 2001 |
| Saami | R1b | 3.9% | Tambets et al. 2004 |
| Saami | R1b3 | 7.9% | Karlsson et al. 2006 |
| Sweden | Haplogroup 1 | 20.0% | Helgason et al. 2000 |
| Sweden (Gotland) | Haplogroup 1 | 17.0% | Rosser et al. 2000 |
| Sweden (northern) | Haplogroup 1 | 23.0% | Rosser et al. 2000 |
| Sweden (Gotland) | Haplogroup 1 | 17.0% | Zerjal et al. 2001 |
| Sweden | Haplogroup 1 | 23.0% | Zerjal et al. 2001 |
| Sweden | R1b | 22.0% | Tambets et al. 2004 |
| Sweden | R1b3 | 23.6% | Karlsson et al. 2006 |
| Sweden | R1b | 13.1% | Lappalainen et al. 2008 |
| Sweden | R1b1b-M29 | 20.9% | Myres et al. 2011 |

## Appendix Table 1: Western European R-Group (R1b1b2-M269)

### Eastern Europe

| Population Studied | Nomenclature used for Western European R | Percentage Found | Reference |
|---|---|---|---|
| Belarus | Haplogroup 1 | 10.0% | Rosser et al. 2000 |
| Belarus | R1b3 | 4.4% | Kharkov et al. 2005 |
| Bulgaria | Haplogroup 1 | 17% | Rosser et al. 2000 |
| Bulgaria | Haplotype XV | 12.9% | Lucotte et al. 2003 |
| Bulgaria | R-M269 | 11.0% | Karachanak et al. 2009 |
| Moldavia | R1-M173 | 13.0% | Nasidze et al 2006(a) |
| Poland | Haplogroup 1 | 18.0% | Rosser et al. 2000 |
| Poland | Eu 18 | 16.4% | Semino et al. 2000(a) |
| Poland | Haplogroup 1 | 19.4% | Scozzari et al. 2001 |
| Poland | Haplotype XV | 16.7% | Lucotte et al. 2003 |
| Poland | R1b | 13.4% | Tambets et al. 2004 |
| Poland | R1(xRa1)<br>(M173) | 11.6% | Kayser et al. 2005 |
| Poland | R1b1b2-M269 | 11.6% | Balaresque et al. 2010 |
| Romania | Haplogroup 1 | 18.0% | Rosser et al. 2000 |
| Romania | Haplogroup 1 | 16.3% | Stefan et al. 2001 |
| Romania | Haplotype XV | 23.0% | Lucotte et al. 2003 |
| Romania | R1b-PN25 | 11.9% | Bosch et al. 2006 |
| Russia | Haplogroup 1 | 26.7% | Helgason et al. 2000 |
| Russia | Haplogroup 1 | 7.0% | Rosser et al. 2000 |
| Russia | Haplogroup 1 | 0.0% | Scozzari et al. 2001 |
| Russia | M173 | 3 – 7% | Wells et al. 2001 |

## Appendix Table 1: Western European R-Group (R1b1b2-M269)

| | | | |
|---|---|---|---|
| Russia | Haplotype XV | 7.3% | Lucotte et al. 2003 |
| Russia | P | 9.2% | Malyarchuk et al. 2004 |
| Russia (northern) | R1b3-M269 | 5.4% | Balanovsky et al. 2008 |
| Russia (central) | R1b3-M269 | 7.1% | Balanovsky et al. 2008 |
| Russia (southern) | R1b3-M269 | 8.8% | Balanovsky et al. 2008 |
| Russia (Europe) | R1 (M173) | 5.1% | Fechner et al. 2008 |
| Russia (northwest) | R1b1b2-M269 | 1.7% | Mirabal et al. 2009 |
| Ukraine | Haplogroup 1 | 4.0% | Rosser et al. 2000 |
| Ukraine | Eu 18 | 2.0% | Semino et al. 2000(a) |
| Ukraine | Haplotype XV | 5.0% | Lucotte et al. 2003 |
| Ukraine (eastern) | P | 9.6% | Kharkov et al. 2004 |

### Baltic Region

| **Population Studied** | **Nomenclature used for Western European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Estonia | Haplogroup 1 | 9.0% | Rosser et al. 2000 |
| Estonia | Haplogroup 1 | 1.4% | Scozzari et al. 2001 |
| Estonia | Haplogroup 1 | 5.0% | Zerjal et al. 2001 |
| Estonia | HG 1 | 5.1% | Laitinen et al. 2002 |
| Estonia | R1b | 9.1% | Tambets et al. 2004 |
| Estonia | R1b | 4.2% | Lappalainen et al. 2008 |
| Latvia | Haplogroup 1 | 15.0% | Rosser et al. 2000 |
| Latvia | Haplogroup 1 | 15.0% | Zerjal et al. 2001 |
| Latvia | HG 1 | 9.6% | Laitinen et al. 2002 |
| Latvia | R1b | 9.3% | Tambets et al. 2004 |

## Appendix Table 1: Western European R-Group (R1b1b2-M269)

| | | | |
|---|---|---|---|
| Latvia | R1b | 9.7% | Lappalainen et al. 2008 |
| Lithuania | Haplogroup 1 | 5.0% | Rosser et al. 2000 |
| Lithuania | Haplogroup 1 | 5.0% | Zerjal et al. 2001 |
| Lithuania | HG 1 | 3.5% | Laitinen et al. 2002 |
| Lithuania | P(xR1a) | 5.1% | Kasperavičiūtė et al. 2004 |
| Lithuania | R1b | 4.9% | Lappalainen et al. 2008 |
| Russia (Karelia) | R1b | 0.8% | Lappalainen et al. 2008 |

**Caucasus**

| **Population Studied** | **Nomenclature used for Western European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Armenia | Haplogroup 1 | 25.0% | Rosser et al. 2000 |
| Armenia | M173 | 36.0% | Wells et al. 2001 |
| Armenia | R1* | 19.0% | Nasidze et al. 2003 |
| Georgia | Haplogroup 1 | 19.0% | Rosser et al. 2000 |
| Georgia | Eu 18 | 14.3% | Semino et al. 2000(a) |

## Appendix Table 2: Eastern European R-Group (R1a1a-M17)

### Eastern Europe

| Population (s) Studied | Nomenclature used for Eastern European R | Percentage Found | Reference |
|---|---|---|---|
| Belarus | Haplogroup 3 | 39.0% | Rosser et al. 2000 |
| Belarus | R1a1 | 45.6% | Kharkov et al. 2005 |
| Bulgaria | Haplogroup 3 | 12.0% | Rosser et al. 2000 |
| Bulgaria | Haplotype XI | 9.7% | Lucotte et al 2003 |
| Bulgaria | R-M17 | 17.3% | Karachanak et al. 2009 |
| Moldavia | R1a1-M17 | 28.3% | Nasidze et al 2006(a) |
| Poland | Haplogroup 3 | 54.0% | Rosser et al. 2000 |
| Poland | Eu 19 | 56.4% | Semino et al. 2000(a) |
| Poland | Eu 19 | 59.7% | Passarino et al. 2001 |
| Poland | Haplogroup 3 | 41.7% | Scozzari et al. 2001 |
| Poland | Haplotype XI | 38.9% | Lucotte et al 2003 |
| Poland | R1a | 55.9% | Tambets et al. 2004 |
| Poland | R1a1* (M17) | 57.0% | Kayser et al. 2005 |
| Romania | Haplogroup 3 | 20.0% | Rosser et al. 2000 |
| Romania | Haplogroup 3 | 23.2% | Stefan et al. 2001 |
| Romania | Haplotype XI | 25.6% | Lucotte et al 2003 |
| Romania | R1a1 | 7.5% | Bosch 2006 |
| Russia | Haplogroup 3 | 43.3% | Helgason et al. 2000 |
| Russia | Haplogroup 3 | 47.0% | Rosser et al. 2000 |
| Russia | Haplogroup 3 | 26.9% | Scozzari et al. 2001 |
| Russia | M17 | 11.0 – 47.0% | Wells et al. 2001 |

## Appendix Table 2: Eastern European R-Group (R1a1a-M17)

| | | | |
|---|---|---|---|
| Russia | Haplotype XI | 43.9% | Lucotte et al 2003 |
| Russia | R1a | 47.3% | Malyarchuk et al. |
| Russia (northern) | R1a-SRY$_{1532}$ | 34.2% | Balanovsky et al. 2008 |
| Russia (central) | R1a-SRY$_{1532}$ | 46.5% | Balanovsky et al. 2008 |
| Russia (southern) | R1a-SRY$_{1532}$ | 55.4% | Balanovsky et al. 2008 |
| Russia (Europe) | R1a1 (M17) | 47.3% | Fechner et al. 2008 |
| Russia (northwest) | R1a1-M198 | 35.6% | Mirabal et al. 2009 |
| Ukraine | Haplogroup 3 | 30.0% | Rosser et al. 2000 |
| Ukraine | Eu 19 | 54.0% | Semino et al. 2000(a) |
| Ukraine | Eu 19 | 50.0% | Passarino et al. 2001 |
| Ukraine | Haplotype XI | 44.0% | Lucotte et al 2003 |
| Ukraine | R1a | 43.6% | Kharkov et al. 2004 |

### Central Europe

| **Population (s) Studied** | **Nomenclature used for Eastern European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Czech Republic | Haplogroup 3 | 38.0% | Rosser et al. 2000 |
| Czech Republic and Slovakia | Eu 19 | 26.7% | Semino et al. 2000(a) |
| Czechoslovakia | Eu 19 | 32.9% | Passarino et al. 2001 |
| Czech Republic | Haplotype XI | 39.3% | Lucotte et al 2003 |
| Czech Republic | R1a-SRY$_{10831}$ | 34.2% | Luca et al. 2007 |
| Hungary | Haplogroup 3 | 22.0% | Rosser et al. 2000 |
| Hungary | Eu 19 | 60.0% | Semino et al. 2000(a) |
| Hungary | Haplotype XI | 40.7% | Lucotte et al 2003 |

## Appendix Table 2: Eastern European R-Group (R1a1a-M17)

| | | | |
|---|---|---|---|
| Hungary | R1a | 20.4% | Tambets et al. 2004 |
| Hungary | R1a1-M198 | 22.7% | Völgyi et al. 2008 |
| Germany | Haplogroup 3 | 9.4% | Helgason et al. 2000 |
| Germany (Bavaria) | Haplogroup 3 | 15.0% | Rosser et al. 2000 |
| Germany | Haplogroup 3 | 30.0% | Rosser et al. 2000 |
| Germany | Eu 19 | 6.2% | Semino et al. 2000(a) |
| Germany | R1a1* (M17) | 17.9% | Kayser et al. 2005 |
| Slovakia | Haplogroup 3 | 47.0% | Rosser et al. 2000 |
| Slovakia | Haplotype XI | 38.0% | Lucotte et al 2003 |

**Balkans**

| **Population (s) Studied** | **Nomenclature used for Eastern European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Albania | EU19 | 9.8% | Semino et al. 2000(a) |
| Albania | EU19 | 12.6% | Passarino et al. 2001 |
| Albania | Haplotype XI | 8.8% | Lucotte et al 2003 |
| Albania | R1a1-M17 | 13.3% | Bosch et al. 2006 |
| Bosnia - Herzegovina | R1a1 | 13.7% | Marjanovic et al. 2005 |
| Bosnia - Herzegovina (Bosnia) | R1a1 | 24.6% | Peričić et al. 2005(b) |
| Bosnia - Herzegovina (Herzegovina) | R1a1 | 12.1% | Peričić et al. 2005(b) |
| Croatia | Eu 19 | 29.3% | Semino et al. 2000(a) |
| Croatia | Eu 19 | 23.0% | Passarino et al. 2001 |

## Appendix Table 2: Eastern European R-Group (R1a1a-M17)

| | | | |
|---|---|---|---|
| Croatia | R1a | 34.0% | Barać et al. 2003 |
| Croatia | R1a1-SRY$_{1532}$ | 25.0% | Peričić et al. 2005(a) |
| Croatia | R1a | 34.3% | Peričić et al. 2005(b) |
| Macedonia | Eu 19 | 35.0% | Semino et al. 2000(a) |
| Macedonia | R1a | 15.2% | Peričić et al. 2005(b) |
| Macedonia | R1a1-M17 | 13.5% | Bosch et al. 2006 |
| Serbia | R1a | 15.9% | Peričić et al. 2005(b) |
| Slovenia | Haplogroup 3 | 37.0% | Rosser et al. 2000 |
| Yugoslavia | Haplogroup 3 | 16.0% | Rosser et al. 2000 |
| Yugoslavia | Haplotype XI | 18.3% | Lucotte et al 2003 |

### Western Europe

| **Population (s) Studied** | **Nomenclature used for Eastern European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Basque | Haplogroup 3 | 0.0% | Rosser et al. 2000 |
| Basque Region of Spain | Eu 19 | 0.0% | Semino et al. 2000(a) |
| Basque Region of France | Eu 19 | 0.0% | Semino et al. 2000(a) |
| Basque Region (Spain) | Eu 19 | 0.0% | Passarino et al. 2001 |
| Basque Region (France) | Eu 19 | 0.0% | Passarino et al. 2001 |
| Belgium | Haplogroup 3 | 4.0% | Rosser et al. 2000 |
| France | Haplogroup 3 | 5.0% | Rosser et al. 2000 |
| France | Eu 19 | 0.0% | Semino et al. 2000(a) |

## Appendix Table 2: Eastern European R-Group (R1a1a-M17)

| | | | |
|---|---|---|---|
| France | Haplogroup 3 | 2.7% | Scozzari et al. 2001 |
| Iceland | Haplogroup 3 | 23.8% | Helgason et al. 2000 |
| Iceland | Haplogroup 3 | 21.0% | Rosser et al. 2000 |
| Ireland | Haplogroup 3 | 0.5% | Helgason et al. 2000 |
| Ireland | Haplogroup 3 | 1.0% | Rosser et al. 2000 |
| Ireland | R1a1 | 1.7% | Capelli et al. 2003 |
| Netherlands | Haplogroup 3 | 13.0% | Rosser et al. 2000 |
| Netherlands | Eu 19 | 3.7% | Semino et al. 2000(a) |
| Netherlands | Eu 19 | 2.9% | Passarino et al. 2001 |
| Portugal (southern) | Haplogroup 3 | 2.0% | Rosser et al. 2000 |
| Portugal (northern) | Haplogroup 3 | 0.0% | Rosser et al. 2000 |
| Portugal (northern) | R1a | 0.0% | Flores et al. 2004 |
| Portugal | R1a* | 2.0% | Beleza et al. 2006 |
| Spain | Haplogroup 3 | 2.0% | Rosser et al. 2000 |
| Spain (Andalusia) | Eu 19 | 0.0% | Semino et al. 2000(a) |
| Spain (Calabria) | Eu 19 | 0.0% | Semino et al. 2000(a) |
| Spain (Andalucia) | Eu 19 | 3.3% | Passarino et al. 2001 |
| Spain (Catalonia) | Eu 19 | 0.0% | Passarino et al. 2001 |
| Spain | Haplogroup 3 | 5.3% | Scozzari et al. 2001 |
| Spain | R1a | 1.7% | Flores et al. 2004 |
| Spain and Portugal (Iberia) | R1a1 | 1.0% | Adams et al. 2008 |
| Spain | R1a* -<br>SRY $_{10831.2}$ | 0.0% | López-Parra et al. 2009 |
| United Kingdom (British) | Haplogroup 3 | 9.4% | Helgason et al. 2000 |

## Appendix Table 2: Eastern European R-Group (R1a1a-M17)

| | | | |
|---|---|---|---|
| United Kingdom (Scotland) | Haplogroup 3 | 6.6% | Helgason et al. 2000 |
| United Kingdom (Western Scotland) | Haplogroup 3 | 7.0% | Rosser et al. 2000 |
| United Kingdom (Scotland) | Haplogroup 3 | 7.0% | Rosser et al. 2000 |
| United Kingdom (Cornwall) | Haplogroup 3 | 0.0% | Rosser et al. 2000 |
| United Kingdom (East Anglia) | Haplogroup 3 | 9.0% | Rosser et al. 2000 |
| United Kingdom | R1a1/3.65+1 | 7.6% | Capelli et al. 2003 |

**Northern Europe**

| **Population (s) Studied** | **Nomenclature used for Eastern European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Denmark | Haplogroup 3 | 16.7% | Helgason et al. 2000 |
| Denmark | Haplogroup 3 | 7.0% | Rosser et al. 2000 |
| Denmark | Haplogroup 3 | 5.7% | Scozzari et al. 2001 |
| Denmark/Schleswig Holstein | R1a1/3.65+1 | 12.0% | Capelli et al. 2003 |
| Denmark | R1a | 16.5% | Tambets et al. 2004 |
| Finland | Haplogroup 3 | 10.0% | Rosser et al. 2000 |
| Finland | Haplogroup 3 | 8.0% | Zerjal et al. 2001 |
| Finland | R1a1 | 7.1% | Lappalainen et al. 2006 |
| Finland (eastern) | R1a1 | 5.9% | Lappalainen et al. 2008 |
| Finland (western) | R1a1 | 8.7% | Lappalainen et al. 2008 |
| Norway | Haplogroup 3 | 17.9% | Helgason et al. 2000 |

# Appendix Table 2: Eastern European R-Group (R1a1a-M17)

| | | | |
|---|---|---|---|
| Norway | Haplogroup 3 | 31.0% | Rosser et al. 2000 |
| Norway | Haplogroup 3 | 31.0% | Zerjal et al. 2001 |
| Norway | R1a1/3.65+1 | 34.0% | Capelli et al. 2003 |
| Norway | R1a | 26.3% | Dupuy et al. 2006 |
| Saami | Haplogroup 3 | 21.0% | Rosser et al. 2000 |
| Saami | Eu 19 | 8.3% | Semino et al. 2000(a) |
| Saami | R1a1 | 15.8% | Karlsson et al. 2006 |
| Sweden | Haplogroup 3 | 17.3% | Helgason et al. 2000 |
| Sweden (northern) | Haplogroup 3 | 19.0% | Rosser et al. 2000 |
| Sweden (Gotland) | Haplogroup 3 | 16.0% | Rosser et al. 2000 |
| Sweden (Gotland) | Haplogroup 3 | 16.0% | Zerjal et al. 2001 |
| Sweden | Haplogroup 3 | 19.0% | Zerjal et al. 2001 |
| Sweden | R1a | 18.4% | Tambets et al. 2004 |
| Sweden | R1a1 | 11.8% | Karlsson et al. 2006 |
| Sweden | R1a1 | 24.4% | Lappalainen et al. 2008 |

## Baltic Region

| **Population (s) Studied** | **Nomenclature used for Eastern European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Estonia | Haplogroup 3 | 27.0% | Rosser et al. 2000 |
| Estonia | Haplogroup 3 | 36.5% | Scozzari et al. 2001 |
| Estonia | Haplogroup 3 | 25.0% | Zerjal et al. 2001 |
| Estonia | HG 3 | 37.3% | Laitinen et al. 2002 |
| Estonia | R1a | 33.5% | Tambets et al. 2004 |
| Estonia | R1a1 | 37.3% | Lappalainen et al. 2008 |

## Appendix Table 2: Eastern European R-Group (R1a1a-M17)

| | | | |
|---|---|---|---|
| Latvia | Haplogroup 3 | 41.0% | Rosser et al. 2000 |
| Latvia | Haplogroup 3 | 41.0% | Zerjal et al. 2001 |
| Latvia | HG 3 | 39.5% | Laitinen et al. 2002 |
| Latvia | R1a | 38.4% | Tambets et al. 2004 |
| Latvia | R1a1 | 38.9% | Lappalainen et al. 2008 |
| Lithuania | Haplogroup 3 | 34.0% | Rosser et al. 2000 |
| Lithuania | Haplogroup 3 | 34.0% | Zerjal et al. 2001 |
| Lithuania | HG 3 | 36.0% | Laitinen et al. 2002 |
| Lithuania | R1a | 44.9% | Kasperavičiūtė et al. 2004 |
| Lithuania | R1a1 | 34.1% | Lappalainen et al. 2008 |
| Russia (Karelia) | R1a1 | 25.0% | Lappalainen et al. 2008 |

### Mediterranean

| **Population (s) Studied** | **Nomenclature used for Eastern European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Corsica | Haplogroup 3 | 0.0% | Scozzari et al. 2001 |
| Crete | R1a1-M198 | 10.7% | Martinez et al. 2007 |
| Crete | R1a1-M17 | 17.0% | King et al. 2008 |
| Cyprus | Haplogroup 3 | 2.0% | Rosser et al. 2000 |
| Cyprus | R1a1 | 3.1% | Capelli et al. 2006 |
| Greece | Haplogroup 3 | 4.8% | Helgason et al. 2000 |
| Greece | Haplogroup 3 | 8.0% | Rosser et al. 2000 |
| Greece | Eu 19 | 11.8% | Semino et al. 2000(a) |
| Greece | Eu 19 | 17.4% | Passarino et al. 2001 |

## Appendix Table 2: Eastern European R-Group (R1a1a-M17)

| | | | |
|---|---|---|---|
| Greece | R1a | 9.8% | Di Giacomo et al. 2003 |
| Greece | Haplotype XI | 15.1% | Lucotte et al 2003 |
| Greece | R1a1-M17 | 22.0% | Bosch et al. 2006 |
| Greece | R1a1-M17 | 11.0% | King et al. 2008 |
| Italy | Haplogroup 3 | 2.7% | Helgason et al. 2000 |
| Italy | Haplogroup 3 | 2.0% | Rosser et al. 2000 |
| Italy (north - central ) | Eu 19 | 4.0% | Semino et al. 2000(a) |
| Italy (northern) | Eu 19 | 4.5% | Passarino et al. 2001 |
| Italy (central) | Eu 19 | 0.9% | Passarino et al. 2001 |
| Italy | Haplogroup 3 | 3.9% | Scozzari et al. 2001 |
| Italy | R1a | 3.4% | Di Giacomo et al. 2003 |
| Italy (southern) | R1a1 | 2.9% | Capelli et al. 2006 |
| Malta | R1a1 | 3.3% | Capelli et al. 2006 |
| Sardinia | Haplogroup 3 | 0.0% | Rosser et al. 2000 |
| Sardinia | Eu 19 | 0.0% | Semino et al. 2000(a) |
| Sardinia | Eu 19 | 1.4% | Passarino et al. 2001 |
| Sardinia | Haplogroup 3 | 1.5% | Scozzari et al. 2001 |
| Sardinia | R1a1 | 0.0% | Capelli et al. 2006 |
| Sardinia | M17 | 1.4% | Contu et al. 2008 |
| Sicily | Eu 19 | 0.0% | Passarino et al. 2001 |
| Sicily | Haplogroup 3 | 3.8% | Scozzari et al. 2001 |
| Sicily (eastern) | R1a1 | 2.3% | Capelli et al. 2006 |
| Sicily (southwestern) | R1a1 | 1.8% | Capelli et al. 2006 |
| Sicily | R1a1 | 2.9% | Capelli et al. 2006 |

# Appendix Table 2: Eastern European R-Group (R1a1a-M17)

| | | | |
|---|---|---|---|
| (northwestern) | | | |
| Sicily | R1a1-M17 | 5.5% | Di Gaetano et al. 2009 |
| Syria | Eu 19 | 10.0% | Semino et al. 2000(a) |

**Middle East**

| **Population (s) Studied** | **Nomenclature used for Eastern European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Iran | HG 3 | 29.0% | Quitana-Murci et al. 2001 |
| Iran (Tehran) | M17 | 4.0% | Wells et al. 2001 |
| Iran (northern) | R1a1-M198 | 3.0% | Regueiro et al. 2006 |
| Iran (southern) | R1a1-M198 | 15.4% | Regueiro et al. 2006 |
| Iraq | R-M17 | 6.5% | Al-Zahery et al. 2003 |
| Jordan | R-M17 | 2.0% | Flores et al. 2005 |
| Lebanon | Eu 19 | 9.7% | Semino et al. 2000(a) |
| Lebanon | R1a1 | 2.5% | Zalloua 2008(b) |
| Qatar | R1a1-M17 | 6.9% | Cadenas et al. 2008 |
| Syria | Eu 19 | 10.0% | Semino et al. 2000(a) |
| Turkey | Haplogroup 3 | 5.0% | Rosser et al. 2000 |
| Turkey | Eu 19 | 6.6% | Semino et al. 2000(a) |
| Turkey | Eu 19 | 11.7% | Passarino et al. 2001 |
| Turkey (Istanbul) | Haplotype XI | 17.2% | Lucotte et al 2003 |
| Turkey (Ankara) | Haplotype XI | 10.3% | Lucotte et al 2003 |

## Appendix Table 2: Eastern European R-Group (R1a1a-M17)

| | | | |
|---|---|---|---|
| Turkey | R1a1 | 6.9% | Cinnioğlu et al. 2004 |
| United Arab Emirites | R1a1-M17 | 7.3% | Cadenas et al. 2008 |

**South Central Asia**

| **Population (s) Studied** | **Nomenclature used for Eastern European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| India (southern) | M17 | 4.0 – 39.0% | Wells et al. 2001 |
| India | R1a-M17 | 27.0% | Kivisbild et al. 2003 |
| India | R1a1-M017 | 15.8% | Sengupta et al. 2006 |
| Pakistan | HG 3 | 4.5% | Quintana-Murci et al. 2001 |
| Pakistan | R1a1-M017 | 24.4% | Sengupta et al. 2006 |

**Caucasus**

| **Population (s) Studied** | **Nomenclature used for Eastern European R** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Armenia | Haplogroup 3 | 6.0% | Rosser et al. 2000 |
| Armenia | M17 | 9.0% | Wells et al. 2001 |
| Armenia | R1a1* | 6.0% | Nasidze et al. 2003 |
| Georgia | Haplogroup 3 | 6.0% | Rosser et al. 2000 |
| Georgia | Eu 19 | 7.9% | Semino et al. 2000(a) |
| Georgia | Eu 19 | 11.4% | Passarino et al. 2001 |
| Georgia | M17 | 8.0% | Wells et al. 2001 |

## Appendix Table 2: Eastern European R-Group (R1a1a-M17)

| | | | |
|---|---|---|---|
| Georgia | R1a1* | 10.0% | Nasidze et al. 2003 |

# Appendix Table 3: Scandinavian I-Group (I1-M253)

## Northern Europe

| Population Studied | Nomenclature used for Scandinavian I Group | Percentage Found | Reference |
|---|---|---|---|
| Denmark | I1-M253 | 32.8% | Underhill et al. 2007 |
| Finland | I1a | 28.0% | Lappalainen et al. 2006 |
| Finland (eastern) | I1a | 19.0% | Lappalainen et al. 2008 |
| Finland (western) | I1a | 40.0% | Lappalainen et al. 2008 |
| Norway | I1a-M253 | 38.9% | Rootsi et al. 2004 |
| Saami | I1a-M253 | 28.6% | Rootsi et al. 2004 |
| Saami (Sweden) | I1a-M253 | 31.6% | Karlsson et al. 2006 |
| Sweden (southern) | I1a-M253 | 35.7% | Rootsi et al. 2004 |
| Sweden (northern) | I1a-M253 | 26.3% | Rootsi et al. 2004 |
| Sweden | I1a-M253 | 37.0% | Karlsson et al. 2006 |
| Sweden | I1a | 35.6% | Lappalainen et al. 2008 |

## Central Europe

| Population Studied | Nomenclature used for Scandinavian I Group | Percentage Found | Reference |
|---|---|---|---|
| Austria | I1-M253 | 2.3% | Underhill et al. 2007 |
| Czech Republic and Slovakia | I1a-M253 | 4.5% | Rootsi et al. 2004 |
| Czech Republic | I1a-M253 | 5.1% | Luca et al. 2007 |
| Czech Republic | I1-M253 | 8.5% | Underhill et al. 2007 |
| Hungary | I1a-M253 | 9.9% | Rootsi et al. 2004 |
| Hungary | I1a-M253 | 8.4% | Vögyi et al. 2008 |

| | | | |
|---|---|---|---|
| Germany | I1a-M253 | 25.0% | Rootsi et al. 2004 |
| Germany | I1-M253 | 15.2% | Underhill et al. 2007 |
| Switzerland | I1a-M253 | 5.6% | Rootsi et al. 2004 |

## Western Europe

| **Population Studied** | **Nomenclature used for Scandinavian I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| France (southern) | I1a-M253 | 5.3% | Rootsi et al. 2004 |
| France (Normandy) | I1a-M253 | 11.9% | Rootsi et al. 2004 |
| France | I1-M253 | 10.0% | Underhill et al. 2007 |
| Ireland | I1-M253 | 6.0% | Underhill et al. 2007 |
| Netherlands | I1a-M253 | 16.7% | Rootsi et al. 2004 |
| Netherlands | I1-M253 | 14.0% | Underhill et al. 2007 |
| Portugal | I1a-M253 | 1.3% | Rootsi et al. 2004 |
| Spain (Catalan ) | I1a-M253 | 3.1% | Rootsi et al. 2004 |
| United Kingdom (England) | I1-M253 | 15.4% | Underhill et al. 2007 |

## Baltic Region

| **Population Studied** | **Nomenclature used for Scandinavian I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Estonia | I1a-M253 | 14.8% | Rootsi et al. 2004 |
| Estonia | I1a | 11.9% | Lappalainen et al. 2008 |
| Latvia | I1a-M253 | 4.7% | Rootsi et al. 2004 |
| Latvia | I1a | 3.5% | Lappalainen et al. 2008 |

## Appendix Table 3: Scandinavian I-Group (I1-M253)

| | | | |
|---|---|---|---|
| Lithuania | I1a | 4.9% | Lappalainen et al. 2008 |
| Russia (Karelian Region) | I1a | 15.2% | Lappalainen et al. 2008 |

### Eastern Europe

| **Population Studied** | **Nomenclature used for Scandinavian I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Belarus | I1a-M253 | 2.7% | Rootsi et al. 2004 |
| Belarus | I1-M253 | 2.0% | Underhill et al. 2007 |
| Moldavia | I1a-M253 | 1.7% | Rootsi et al. 2004 |
| Poland | I1a-M253 | 5.8% | Rootsi et al. 2004 |
| Romania | I1a-M253 | 1.7% | Rootsi et al. 2004 |
| Russia | I1a-M253 | 3.6% | Rootsi et al. 2004 |
| Russia | I1-M253 | 6.0% | Underhill et al. 2007 |
| Russia (northern) | I1a-M253 | 6.2% | Balanovsky 2008 |
| Russia (central) | I1a-M253 | 5.3% | Balanovsky 2008 |
| Russia (southern) | I1a-M253 | 3.9% | Balanovsky 2008 |
| Russia (northwestern) | I1-M253 | 4.7% | Mirabal et al. 2009 |
| Ukraine | I1a-M253 | 4.8% | Rootsi et al. 2004 |
| Ukraine | I1-M253 | 4.9% | Underhill et al. 2007 |

# Appendix Table 3: Scandinavian I-Group (I1-M253)

## Balkans

| **Population Studied** | **Nomenclature used for Scandinavian I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Albania | I1a-M253 | 2.8% | Rootsi et al. 2004 |
| Bosnia | I1a-M253 | 2.0% | Rootsi et al. 2004 |
| Bosnia-Herzegovina | I1a-M253 | 2.3% | Marjanovic et al. 2005 |
| Bosnia-Herzegovina | I1a-M253 | 3.8% | Peričić et al 2005(b) |
| Croatia | I1a-M253 | 5.3% | Rootsi et al. 2004 |
| Croatia | I1a-M253 | 2.8% | Peričić et al 2005(b) |
| Macedonia | I1a-M253 | 8.0% | Rootsi et al. 2004 |
| Macedonia | I1a-M253 | 5.1% | Peričić et al 2005(b) |
| Serbia | I1a-M253 | 5.3% | Peričić et al 2005(b) |
| Slovenia | I1a-M253 | 10.9% | Rootsi et al. 2004 |
| Slovenia | I1-M253 | 7.4% | Underhill et al. 2007 |

## Mediterranean

| **Population Studied** | **Nomenclature used for Scandinavian I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Greece | I1a-M253 | 2.5% | Rootsi et al. 2004 |
| Greece | I1-M253 | 2.3% | Underhill et al. 2007 |
| Italy (northern) | I1a-M253 | 2.6% | Rootsi et al. 2004 |
| Italy (central) | I1a-M253 | 2.0% | Rootsi et al. 2004 |
| Italy (southern) | I1a-M253 | 0.7% | Rootsi et al. 2004 |

## Appendix Table 3: Scandinavian I-Group (I1-M253)

### Middle East

| **Population Studied** | **Nomenclature used for Scandinavian I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Turkey | I1a-M253 | 0.9% | Rootsi et al. 2004 |
| Turkey | I1-M253 | 1.1% | Underhill et al. 2007 |

## Appendix Table 4: Balkan I-Group (I2a1-M423)

### Balkans

| Population Studied | Nomenclature used for Balkan I Group | Percentage Found | Reference |
|---|---|---|---|
| Albania | I1b-P37 | 17.0% | Rootsi et al. 2004 |
| Bosnia | I1b-P37 | 40.0% | Rootsi et al. 2004 |
| Bosnia-Herzegovina | I-P37 (xM26) | 60.0% | Peričić et al 2005(b) |
| Bosnia-Herzegovina | I1b-P37 | 49.2% | Marjanovic et al. 2005 |
| Croatia | I1b-P37 | 31.2% | Rootsi et al. 2004 |
| Croatia | I-P37 | 41.7% | Peričić et al. 2005(a) |
| Croatia | I-P37 (xM26) | 32.4% | Peričić et al 2005(b) |
| Macedonia | I1b-P37 | 18.0% | Rootsi et al. 2004 |
| Macedonia | I-P37 (xM26) | 29.1% | Peričić et al 2005(b) |
| Serbia | I-P37 (xM26) | 29.2% | Peričić et al 2005(b) |
| Slovenia | I1b-P37 | 20.0% | Rootsi et al. 2004 |
| Slovenia | I2a2-M423 | 18.9% | Underhill et al. 2007 |

### Eastern Europe

| Population Studied | Nomenclature used for Balkan I Group | Percentage Found | Reference |
|---|---|---|---|
| Belarus | I1b-P37 | 15.0% | Rootsi et al. 2004 |
| Belarus | I2a2-M423 | 15.7% | Underhill et al. 2007 |
| Moldavia | I1b-P37 | 21.1% | Rootsi et al. 2004 |
| Poland | I1b-P37 | 9.9% | Rootsi et al. 2004 |

## Appendix Table 4: Balkan I-Group (I2a1-M423)

| | | | |
|---|---|---|---|
| Romania | I1b-P37 | 17.7% | Rootsi et al. 2004 |
| Russia | I1b-P37 | 7.0% | Rootsi et al. 2004 |
| Russia | I2a2-M423 | 9.0% | Underhill et al. 2007 |
| Russia (northern) | I1b-P37 | 5.7% | Balanovsky 2008 |
| Russia (central) | I1b-P37 | 10.0% | Balanovsky 2008 |
| Russia (southern) | I1b-P37 | 15.9% | Balanovsky 2008 |
| Russia (northwest) | I2a-P37.2 | 3.4% | Mirabal et al. 2009 |
| Ukraine | I1b-P37 | 16.1% | Rootsi et al. 2004 |
| Ukraine | I2a2-M423 | 20.2% | Underhill et al. 2007 |

### Baltic Region

| **Population Studied** | **Nomenclature used for Balkan I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Estonia | I1b-P37 | 2.9% | Rootsi et al. 2004 |
| Estonia | I1b | 4.2% | Lappalainen et al. 2008 |
| Latvia | I1b | 2.7% | Lappalainen et al. 2008 |
| Lithuania | I1b | 4.9% | Lappalainen et al. 2008 |
| Russia (Karelian Region) | I1b | 2.3% | Lappalainen et al. 2008 |

### Central Europe

| **Population Studied** | **Nomenclature used for Balkan I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Austria | I2a2-M423 | 4.7% | Underhill et al. 2007 |

## Appendix Table 4: Balkan I-Group (I2a1-M423)

| | | | |
|---|---|---|---|
| Czech Republic and Slovakia | I1b-P37 | 7.1% | Rootsi et al. 2004 |
| Czech Republic | I2a-P37 | 7.4% | Luca et al. 2007 |
| Czech Republic | I2a2-M423 | 17.0% | Underhill et al. 2007 |
| Hungary | I1b-P37 | 11.1% | Rootsi et al. 2004 |
| Hungary | I1b-P37 | 16.8% | Vögyi et al. 2008 |

### Northern Europe

| **Population Studied** | **Nomenclature used for Balkan I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Finland | I1b | 0.2% | Lappalainen et al. 2006 |
| Finland (western) | I1b | 0.4% | Lappalainen et al. 2008 |

### Western Europe

| **Population Studied** | **Nomenclature used for Balkan I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Ireland | I2a2-M423 | 2.0% | Underhill et al. 2007 |
| Portugal | I1b-P37 | 0.7% | Rootsi et al. 2004 |
| United Kingdom (England) | I2a2-M423 | 1.0% | Underhill et al. 2007 |

# Appendix Table 4: Balkan I-Group (I2a1-M423)

## Mediterranean

| Population Studied | Nomenclature used for Balkan I Group | Percentage Found | Reference |
|---|---|---|---|
| Greece | I1b-P37 | 8.4% | Rootsi et al. 2004 |
| Greece | I2a2-M423 | 3.0% | Underhill et al. 2007 |
| Italy (northern) | I1b-P37 | 1.0% | Rootsi et al. 2004 |
| Italy (southern) | I1b-P37 | 0.7% | Rootsi et al. 2004 |
| Sardinia | I2a2-M423 | 1.3% | Underhill et al. 2007 |

## Middle East

| Population Studied | Nomenclature used for Balkan I Group | Percentage Found | Reference |
|---|---|---|---|
| Lebanon | I1b-P37 | 1.5% | Rootsi et al. 2004 |
| Turkey | I1b-P37 | 2.3% | Rootsi et al. 2004 |
| Turkey | I2a2-M423 | 2.3% | Underhill et al. 2007 |

## Appendix Table 5: Sardinian I Group (I2a2-M26)

### Mediterranean

| Population Studied | Nomenclature used for Sardinian I Group | Percentage Found | Reference |
|---|---|---|---|
| Italy (central) | I1b2-M26 | 1.0% | Rootsi et al. 2004 |
| Italy (Calabria) | I1b2-M26 | 0.7% | Rootsi et al. 2004 |
| Italy (southern) | I1b2 | 1.5% | Capelli et al. 2006 |
| Sardinia | Eu 8 | 35.1% | Semino et al. 2000(a) |
| Sardinia | I-M26 | 33.7% | Zei et al. 2003 |
| Sardinia | I1b2-M26 | 40.9% | Rootsi et al. 2004 |
| Sardinia | I1b2 | 25.9% | Capelli et al. 2006 |
| Sardinia | I2a1-M26 | 43.0% | Underhill et al. 2007 |
| Sardinia | M26 | 37.0% | Contu et al. 2008 |
| Sicily | I1b2 | 0.5% | Capelli et al. 2006 |

### Western Europe

| Population Studied | Nomenclature used for Sardinian I Group | Percentage Found | Reference |
|---|---|---|---|
| Basque Region (Spain) | Eu 8 | 4.4% | Semino et al. 2000(a) |
| Basque Region (France) | Eu 8 | 9.1% | Semino et al. 2000(a) |
| Basques (France and Spain) | I1b2-M26 | 6.0% | Rootsi et al. 2004 |
| France (Normandy) | I1b2-M26 | 2.4% | Rootsi et al. 2004 |
| France (Bearnais in the Pyrenees | I1b2-M26 | 7.7% | Rootsi et al. 2004 |

## Appendix Table 5: Sardinian I Group (I2a2-M26)

| | | | |
|---|---|---|---|
| Region) | | | |
| France | I2a1-M26 | 1.4% | Underhill et al. 2007 |
| Ireland (Rush) | I1b2-M26 | 2.6% | Rootsi et al. 2004 |
| Portugal | I1b2-M26 | 0.3% | Rootsi et al. 2004 |
| Portugal (Madeira) | I1b2-M26 | 0.7% | Rootsi et al. 2004 |
| Spain (Cantabria) | I-M26 | 3.0% | Maca-Meyer et al. 2003 |
| Iberia (Spain and Portugal) | I1b2 | 3.4% | Flores et al. 2004 |
| Spain (Andalusia) | I1b2-M26 | 1.0% | Rootsi et al. 2004 |
| Spain (Pyrenees Region) | I2a2-M26 | 7.7% | López-Parra et al. 2009 |
| United Kingdom | I1b2 | 0.5% | Capelli et al. 2003 |
| United Kingdom (Wales) | I1b2-M26 | 0.5% | Rootsi et al. 2004 |
| United Kingdom (England) | I1b2-M26 | 0.7% | Rootsi et al. 2004 |
| United Kingdom (Scottish Isles) | I1b2-M26 | 0.4% | Rootsi et al. 2004 |

### Northern Europe

| **Population Studied** | **Nomenclature used for Sardinian I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Sweden (southern) | I1b2-M26 | 0.6% | Rootsi et al. 2004 |

## Appendix Table 5: Sardinian I Group (I2a2-M26)

### Central Europe

| **Population Studied** | **Nomenclature used for Sardinian I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Czech Republic | I2a1-M26 | 1.2% | Luca et al. 2007 |

# Appendix Table 6: Central European I-Group (I2b1-M223)

## Central Europe

| Population Studied | Nomenclature used for Central European I Group | Percentage Found | Reference |
|---|---|---|---|
| Czech Republic and Slovakia | I1c-M223 | 1.0% | Rootsi et al. 2004 |
| Czech Republic | I2b1-M223 | 2.7% | Luca et al. 2007 |
| Czech Republic | I2b1-M223 | 6.4% | Underhill et al. 2007 |
| Germany | I1c-M223 | 12.5% | Rootsi et al. 2004 |
| Germany | I2b1-M223 | 6.4% | Underhill et al. 2007 |
| Hungary | I1c-M223 | 1.2% | Rootsi et al. 2004 |
| Hungary | I1c-M223 | 4.2% | Vögyi et al. 2008 |
| Switzerland | I1c-M223 | 1.4% | Rootsi et al. 2004 |

## Western Europe

| Population Studied | Nomenclature used for Central European I Group | Percentage Found | Reference |
|---|---|---|---|
| France (southern) | I1c-M223 | 5.3% | Rootsi et al. 2004 |
| France (Normandy) | I1c-M223 | 4.8% | Rootsi et al. 2004 |
| France | I2b1-M223 | 4.3% | Underhill et al. 2007 |
| Ireland | I2b1-M223 | 1.0% | Underhill et al. 2007 |
| Netherlands | I1c-M223 | 10.0% | Rootsi et al. 2004 |
| Netherlands | I2b1-M223 | 5.4% | Underhill et al. 2007 |
| Portugal | I1c-M223 | 1.6% | Rootsi et al. 2004 |
| Portugal (Madeira) | I1c-M223 | 1.5% | Rootsi et al. 2004 |

# Appendix Table 6: Central European I-Group (I2b1-M223)

| | | | |
|---|---|---|---|
| Portugal (Azores) | I1c-M223 | 1.6% | Rootsi et al. 2004 |
| Portugal (Cape Verde) | I1c-M223 | 0.5% | Rootsi et al. 2004 |
| United Kingdom (England) | I2b1-M223 | 1.0% | Underhill et al. 2007 |

## Northern Europe

| **Population Studied** | **Nomenclature used for Central European I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Denmark | I2b1-M223 | 4.9% | Underhill et al. 2007 |
| Finland | I1c | 0.8% | Lappalainen et al. 2006 |
| Finland (eastern) | I1c-M223 | 0.7% | Lappalainen et al. 2008 |
| Finland (western) | I1c-M223 | 0.9% | Lappalainen et al. 2008 |
| Norway | I1c-M223 | 1.4% | Rootsi et al. 2004 |
| Sweden (southern) | I1c-M223 | 3.6% | Rootsi et al. 2004 |
| Sweden | I1c-M223 | 4.9% | Karlsson et al. 2006 |
| Sweden | I1c-M223 | 1.9% | Lappalainen et al. 2008 |

## Mediterranean

| **Population Studied** | **Nomenclature used for Central European I Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Greece | I1c-M223 | 1.5% | Rootsi et al. 2004 |
| Greece | I2b1-M223 | 2.3% | Underhill et al. 2007 |
| Italy (northern) | I1c-M223 | 1.0% | Rootsi et al. 2004 |

## Appendix Table 6: Central European I-Group (I2b1-M223)

| | | | |
|---|---|---|---|
| Italy (central) | I1c-M223 | 3.0% | Rootsi et al. 2004 |
| Italy (Calabria) | I1c-M223 | 1.4% | Rootsi et al. 2004 |
| Sardinia | I1c-M223 | 1.4% | Rootsi et al. 2004 |

### Eastern Europe

| Population Studied | Nomenclature used for Central European I Group | Percentage Found | Reference |
|---|---|---|---|
| Belarus | I1c-M223 | 0.7% | Rootsi et al. 2004 |
| Moldavia | I1c-M223 | 3.3% | Rootsi et al. 2004 |
| Poland | I1c-M223 | 1.0% | Rootsi et al. 2004 |
| Romania | I1c-M223 | 1.9% | Rootsi et al. 2004 |
| Russia | I1c-M223 | 1.1% | Rootsi et al. 2004 |
| Russia | I2b1-M223 | 0.5% | Underhill et al. 2007 |
| Russia (northern) | I1c-M223 | 0.4% | Balanovsky 2008 |
| Russia (central) | I1c-M223 | 1.6% | Balanovsky 2008 |
| Russia (southern) | I1c-M223 | 0.7% | Balanovsky 2008 |
| Ukraine | I1c-M223 | 0.5% | Rootsi et al. 2004 |
| Ukraine | I2b1-M223 | 0.4% | Underhill et al. 2007 |

### Baltic Region

| Population Studied | Nomenclature used for Central European I Group | Percentage Found | Reference |
|---|---|---|---|
| Albania | I1c-M223 | 3.8% | Rootsi et al. 2004 |

# Appendix Table 6: Central European I-Group (I2b1-M223)

| | | | |
|---|---|---|---|
| Estonia | I1c-M223 | 0.5% | Rootsi et al. 2004 |
| Estonia | I1c-M223 | 0.8% | Lappalainen et al. 2008 |
| Latvia | I1c-M223 | 1.2% | Rootsi et al. 2004 |
| Latvia | I1c-M223 | 0.9% | Lappalainen et al. 2008 |
| Lithuania | I1c-M223 | 1.8% | Lappalainen et al. 2008 |

## Balkans

| Population Studied | Nomenclature used for Central European I Group | Percentage Found | Reference |
|---|---|---|---|
| Bosnia-Herzegovina | I1c-M223 | 0.4% | Marjanovic et al. 2005 |
| Croatia | I1c-M223 | 0.5% | Rootsi et al. 2004 |
| Croatia | I1c-M223 | 0.9% | Peričić et al 2005(b |
| Slovenia | I1c-M223 | 1.8% | Rootsi et al. 2004 |

## Middle East

| Population Studied | Nomenclature used for Central European I Group | Percentage Found | Reference |
|---|---|---|---|
| Lebanon | I1c-M223 | 1.5% | Rootsi et al. 2004 |
| Pakistan | I1c2-M170 \M223 \M379 | 0.57% | Sengupta et al. 2006 |
| Turkey | I1c-M223 | 0.7% | Rootsi et al. 2004 |
| Turkey | I2b1-M223 | 0.6% | Underhill et al. 2007 |

# Appendix Table 7: Finno-Baltic N-Group (N1c1-M178)

## Northern Europe

| Population Studied | Nomenclature used for the Finno-Baltic N Group | Percentage Found | Reference |
|---|---|---|---|
| Denmark | Haplogroup 16 | 0.0% | Helgason et al. 2000 |
| Denmark | Haplogroup 16 | 2.0% | Rosser et al. 2000 |
| Denmark | Haplogroup 16 | 2.9% | Scozzari et al. 2001 |
| Denmark | N3 | 0.5% | Tamberts et al. 2004 |
| Finland | Haplogroup 16 | 61.0% | Rosser et al. 2000 |
| Finland | Haplogroup 16 | 64.0% | Zerjal et al. 2001 |
| Finland | N3 | 58.2% | Lappalainen et al. 2006 |
| Finland (eastern) | N3 | 70.9% | Lappalainen et al. 2008 |
| Finland (western) | N3 | 41.3% | Lappalainen et al. 2008 |
| Norway | Haplogroup 16 | 2.7% | Helgason et al. 2000 |
| Norway | Haplogroup 16 | 4.0% | Rosser et al. 2000 |
| Norway | Haplogroup 16 | 4.0% | Zerjal et al. 2001 |
| Norway | EU 14 | 6.9% | Passarino et al. 2002 |
| Norway | N3 | 3.8% | Dupuy et al. 2006 |
| Norway (northern) | N3 | 10.6% | Dupuy et al. 2006 |
| Saami | Haplogroup 16 | 42.0% | Rosser et al. 2000 |
| Saami | Eu 14 | 41.7% | Semino et al. 2000 |
| Saami | Haplogroup 16 | 42.0% | Zerjal et al. 2001 |
| Saami | N3 | 47.2% | Tamberts et al. 2004 |
| Saami | N3 | 44.7% | Karlsson et al. 2006 |
| Sweden | Haplogroup 16 | 7.3% | Helgason et al. 2000 |
| Sweden (northern) | Haplogroup 16 | 8.0% | Rosser et al. 2000 |

## Appendix Table 7: Finno-Baltic N-Group (N1c1-M178)

| | | | |
|---|---|---|---|
| Sweden (Gotland) | Haplogroup 16 | 6.0% | Rosser et al. 2000 |
| Sweden (Gotland) | Haplogroup 16 | 6.0% | Zerjal et al. 2001 |
| Sweden | Haplogroup 16 | 8.0% | Zerjal et al. 2001 |
| Sweden | N3 | 2.8% | Tamberts et al. 2004 |
| Sweden | N3 | 9.5% | Karlsson et al. 2006 |
| Sweden | N3 | 14.4% | Lappalainen et al. 2008 |

### Baltic Region

| **Population Studied** | **Nomenclature used for the Finno-Baltic N Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Estonia | Haplogroup 16 | 37.0% | Rosser et al. 2000 |
| Estonia | Haplogroup 16 | 32.5% | Zerjal et al. 2001 |
| Estonia | Haplogroup 16 | 33.9% | Laitinen et al. 2002 |
| Estonia | Haplogroup 16 | 32.4% | Scozzari et al. 2001 |
| Estonia | N3 | 30.6% | Tamberts et al. 2004 |
| Estonia | N3 | 33.9% | Lappalainen et al. 2008 |
| Latvia | Haplogroup 16 | 32.0% | Rosser et al. 2000 |
| Latvia | Haplogroup 16 | 32.0% | Zerjal et al. 2001 |
| Latvia | Haplogroup 16 | 42.1% | Laitinen et al. 2002 |
| Latvia | N3 | 41.9% | Tamberts et al. 2004 |
| Latvia | N3 | 41.6% | Lappalainen et al. 2008 |
| Lithuania | Haplogroup 16 | 47.0% | Rosser et al. 2000 |
| Lithuania | Haplogroup 16 | 47.0% | Zerjal et al. 2001 |
| Lithuania | Haplogroup 16 | 43.0% | Laitinen et al. 2002 |
| Lithuania | N3 | 36.7% | Kasperavičiūtė et al. |

## Appendix Table 7: Finno-Baltic N-Group (N1c1-M178)

| | | | |
|---|---|---|---|
| | | | 2004 |
| Lithuania | N3 | 43.9% | Lappalainen et al. 2008 |
| Russia (Karelia) | N3 | 53.0% | Lappalainen et al. 2008 |

### Eastern Europe

| **Population Studied** | **Nomenclature used for the Finno-Baltic N Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Belarus | Haplogroup 16 | 2.0% | Rosser et al. 2000 |
| Belarus | N3a | 8.8% | Kharkov et al. 2005 |
| Bulgaria | Haplogroup 16 | 0.0% | Rosser et al. 2000 |
| Poland | Haplogroup 16 | 4.0% | Rosser et al. 2000 |
| Poland | Haplogroup 16 | 2.8% | Scozzari et al. 2001 |
| Poland | N3 | 3.2% | Tamberts et al. 2004 |
| Poland | N3-M46 | 0.5% | Kayser et al. 2005 |
| Romania | Haplogroup 16 | 0.0% | Rosser et al. 2000 |
| Russia | Haplogroup 16 | 3.3% | Helgason et al. 2000 |
| Russia | Haplogroup 16 | 14.0% | Rosser et al. 2000 |
| Russia | M46 | 20.0% | Well et al. 2001 |
| Russia | Haplogroup 16 | 50.0% | Scozzari et al. 2001 |
| Russia | N3 | 13.9% | Malyarchuk et al. 2004 |
| Russia | N3 | 8.2% | Tamberts et al. 2004 |
| Russia | N3 | 14.0% | Derenko et al. 2006 |
| Russia | N3a | 11.3% | Derenko et al. 2007 |
| Russia (northern) | N3 | 35.5% | Balanovsky et al. 2008 |
| Russia (central) | N3 | 16.3% | Balanovsky et al. 2008 |

## Appendix Table 7: Finno-Baltic N-Group (N1c1-M178)

| | | | |
|---|---|---|---|
| Russia (southern) | N3 | 9.5% | Balanovsky et al. 2008 |
| Russia (Europe) | N3 (TAT) | 14.7% | Fechner et al. 2008 |
| Russia (Europe) | N3a | 14.0% | Malyarchuk and Derenko 2008 |
| Russia (northwestern) | N1c1-M178 | 29.7% | Mirabal et al. 2009 |
| Ukraine | Haplogroup 16 | 11.0% | Rosser et al. 2000 |
| Ukraine | Eu 14 | 6.0% | Semino et al. 2000 |
| Ukraine (eastern) | N3 | 9.6% | Kharkov et al. 2004 |

### Balkans

| **Population Studied** | **Nomenclature used for the Finno-Baltic N Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Slovenia | Haplogroup 16 | 0.0% | Rosser et al. 2000 |
| Yugoslavia (former) | Haplogroup 16 | 0.0% | Rosser et al. 2000 |

### Central Europe

| **Population Studied** | **Nomenclature used for the Finno-Baltic N Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Czech Republic | Haplogroup 16 | 0.0% | Rosser et al. 2000 |
| Czech and Slovak Republics | EU 14 | 2.2% | Semino et al. 2000 |
| Czech Republic | N3-TAT | 1.6% | Luca et al. 2007 |
| Germany | Haplogroup 16 | 0.0% | Helgason et al. 2000 |
| Germany (Bavaria) | Haplogroup 16 | 0.0% | Rosser et al. 2000 |

# Appendix Table 7: Finno-Baltic N-Group (N1c1-M178)

| | | | |
|---|---|---|---|
| Germany | Haplogroup 16 | 3.0% | Rosser et al. 2000 |
| Germany | N3-M46 | 1.6% | Kayser et al. 2005 |
| Hungary | Haplogroup 16 | 0.0% | Rosser et al. 2000 |
| Slovak Republic | Haplogroup 16 | 3.0% | Rosser et al. 2000 |

**Middle East**

| **Population Studied** | **Nomenclature used for the Finno-Baltic N Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Turkey | Haplogroup 16 | 1.0% | Rosser et al. 2000 |
| Turkey | Eu 14 | 3.3% | Semino et al. 2000 |
| Turkey | N3a-M178 | 1.0% | Cinnioğlu et al. 2004 |

**Caucasus**

| **Population Studied** | **Nomenclature used for the Finno-Baltic N Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Armenia | Haplogroup 16 | 3.0% | Rosser et al. 2000 |
| Georgia | Haplogroup 16 | 0.0% | Rosser et al. 2000 |

**Mediterranean**

| **Population Studied** | **Nomenclature used for the Finno-Baltic N Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Cyprus | Haplogroup 16 | 0.0% | Rosser et al. 2000 |
| Greece | Haplogroup 16 | 2.4% | Helgason et al. 2000 |
| Greece | Haplogroup 16 | 0.0% | Rosser et al. 2000 |

## Appendix Table 7: Finno-Baltic N-Group (N1c1-M178)

| | | | |
|---|---|---|---|
| Italy | Haplogroup 16 | 0.0% | Helgason et al. 2000 |
| Italy | Haplogroup 16 | 0.0% | Rosser et al. 2000 |
| Italy | Haplogroup 16 | 0.0% | Scozzari et al. 2001 |
| Sardinia | Haplogroup 16 | 0.0% | Rosser et al. 2000 |

### Western Europe

| **Population Studied** | **Nomenclature used for the Finno-Baltic N Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Belgium | Haplogroup 16 | 0.0% | Rosser et al. 2000 |
| France | Haplogroup 16 | 0.0% | Rosser et al. 2000 |
| France | Haplogroup 16 | 0.0% | Scozzari et al. 2001 |
| Iceland | Haplogroup 16 | 0.6% | Helgason et al. 2000 |
| Iceland | Haplogroup 16 | 0.0% | Rosser et al. 2000 |
| Ireland | Haplogroup 16 | 0.5% | Helgason et al. 2000 |
| Ireland | Haplogroup 16 | 0.5% | Rosser et al. 2000 |
| Netherlands | Haplogroup 16 | 0.0% | Rosser et al. 2000 |
| Spain and Portugal (Iberia) | Haplogroup 16 | 0.0% | Rosser et al. 2000 |
| Spain | Haplogroup 16 | 0.0% | Scozzari et al. 2001 |
| Spain and Portugal (Iberia) | N3a-TAT | 0.1% | Alonso et al. 2005 |
| United Kingdom (Scotland) | Haplogroup 16 | 0.0% | Helgason et al. 2000 |
| United Kingdom (British) | Haplogroup 16 | 0.0% | Helgason et al. 2000 |
| United Kingdom | Haplogroup 16 | 0.0% | Rosser et al. 2000 |

## Appendix Table 7: Finno-Baltic N-Group (N1c1-M178)

### Siberia

| Population Studied | Nomenclature used for the Finno-Baltic N Group | Percentage Found | Reference |
|---|---|---|---|
| Siberia | N-M178 | 22.7% | Karafet et al. 2002 |
| Siberia | N3 | 29.2% | Tamberts et al. 2004 |
| Siberia (southern) | N3 | 13.3% | Derenko et al. 2006 |

# Appendix Table 8: European E-Group (E1b1b1a2-V13)

## Mediterranen

| Population Studied | Nomenclature used for European<br>E-Group | Percentage Found | Reference |
|---|---|---|---|
| Crete | E3b1a2-V13 | 6.7% | King et al. 2008 |
| Greece (continental) | E-V13 | 17.7% | Cruciani et al. 2007 |
| Greece | E3b1a2-V13 | 28.1% | King et al. 2008 |
| Greece | E1b1b1a2-V13 | 16.3% | Battaglia et al. 2009 |
| Italy (northern) | E-V13 | 5.3% | Cruciani et al. 2007 |
| Italy (central) | E-V13 | 5.3% | Cruciani et al. 2007 |
| Italy (southern) | E-V13 | 8.5% | Cruciani et al. 2007 |
| Italy (Sicily) | E-V13 | 7.2% | Cruciani et al. 2007 |
| Sicily | E3b1a2-V13 | 5.9% | Di Gaetano et al. 2009 |
| Sardinia | E-V13 | 1.1% | Cruciani et al. 2007 |

## Balkans

| Population Studied | Nomenclature used for European<br>E-Group | Percentage Found | Reference |
|---|---|---|---|
| Albania | E-V13 | 32.3% | Cruciani et al. 2007 |
| Albania | E3b1-M78 | 23.3% | Bosch et al. 2006 |
| Albania | E1b1b1a2-V13 | 23.6% | Battaglia et al. 2009 |
| Bosnia-Herzegovina | E3b1-M78 | 13.7% | Marjanovic et al. 2005 |

## Appendix Table 8: European E-Group (E1b1b1a2-V13)

| | | | |
|---|---|---|---|
| Bosnia-Herzegovina | E3b1-M78 | 9.1% | Peričić et al. 2006(b) |
| Croatia | E-78 | 7.0% | Semino et al. 2004 |
| Croatia | E-M78 | 4.9% | Peričić et al. 2006(a) |
| Croatia (mainland) | E3b1-M78 | 5.6% | Peričić et al. 2006(b) |
| Macedonia | E3b1-M78 | 24.1% | Peričić et al. 2006(b) |
| Macedonia | E-V13 | 17.2% | Cruciani et al. 2007 |
| Macedonia | E3b1-M78 | 21.2% | Bosch et al. 2006 |
| Serbia | E3b1-M78 | 20.4% | Peričić et al. 2006(b) |
| Slovenia | E-V13 | 2.9% | Cruciani et al. 2007 |
| Slovenia | E1b1b1a2-V13 | 2.7% | Battaglia et al. 2009 |

### Western Europe

| **Population Studied** | **Nomenclature used for European E-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Basque Region (Spain) | E-V13 | 0.0% | Cruciani et al. 2007 |
| Basque Region (France) | E-V13 | 0.0% | Cruciani et al. 2007<br>Cruciani et al. 2007 |
| France | E-V13 | 4.0% | Cruciani et al. 2007 |
| Netherlands | HG E | 0.0% | Semino et al. 2004 |
| Portugal (northern) | E-V13 | 4.0% | Cruciani et al. 2007 |
| Portugal (southern) | E-V13 | 4.1% | Cruciani et al. 2007 |
| Spain (southern) | E-V13 | 0.0% | Cruciani et al. 2007 |
| United Kingdom | E3b | 2.2% | Capelli et al. 2003 |

# Appendix Table 8: European E-Group (E1b1b1a2-V13)

| | | | |
|---|---|---|---|
| United Kingdom (English) | E-V13 | 0.0% | Cruciani et al. 2007 |

## Central Europe

| Population Studied | Nomenclature used for European E-Group | Percentage Found | Reference |
|---|---|---|---|
| Czech Republic | E-V13 | 4.9% | Cruciani et al. 2007 |
| Czech Republic | E3b1-M78 | 5.1% | Luca et al. 2007 |
| Germany | E3b-M35 | 6.2% | Kayser et al. 2005 |
| Germany | E-V13 | 3.9% | Cruciani et al. 2007 |
| Hungary | E-V13 | 9.4% | Cruciani et al. 2007 |
| Hungary | E3b1-M78 | 2.5% | Völgyi et al. 2008 |
| Hungary | E1b1b1a2-V13 | 7.5% | Battaglia et al. 2009 |
| Slovakia | E-V13 | 8.3% | Cruciani et al. 2007 |

## Northern Europe

| Population Studied | Nomenclature used for European E-Group | Percentage Found | Reference |
|---|---|---|---|
| Denmark/Schleswig-Holstein | E3b | 3.0% | Capelli et al. 2003 |
| Denmark | E-V13 | 2.9% | Cruciani et al. 2007 |
| Sweden | E3b1-M78 | 1.0% | Karlsson et al. 2006 |

## Appendix Table 8: European E-Group (E1b1b1a2-V13)

### Baltic Region

| **Population Studied** | **Nomenclature used for European E-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Estonia | E-V13 | 4.1% | Cruciani et al. 2007 |

### Eastern Europe

| **Population Studied** | **Nomenclature used for European E-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Bulgaria | E-V13 | 16.2% | Cruciani et al. 2007 |
| Moldavia | E-V13 | 7.8% | Cruciani et al. 2007 |
| Poland | E3b-M35 | 4.5% | Kayser et al. 2005 |
| Poland | E-V13 | 2.5% | Cruciani et al. 2007 |
| Poland | E1b1b1a2-V13 | 4.0% | Battaglia et al. 2009 |
| Romania | E3b1-M78 | 11.9% | Bosch et al. 2006 |
| Russia (northern) | E-V13 | 3.7% | Cruciani et al. 2007 |
| Russia (southern) | E-V13 | 2.2% | Cruciani et al. 2007 |
| Russia (northern) | E3b1-M78 | 0.2% | Balanovsky et al. 2008 |
| Russia (central) | E3b1-M78 | 4.6% | Balanovsky et al. 2008 |
| Russia (southern) | E3b1-M78 | 1.8% | Balanovsky et al. 2008 |
| Russia (Northwestern) | E1b1b1a2-V13 | 1.7% | Mirabal et al. 2009 |

# Appendix Table 8: European E-Group (E1b1b1a2-V13)

| | | | |
|---|---|---|---|
| Rumania | E-V13 | 7.2% | Cruciani et al. 2007 |
| Ukraine | E-V13 | 9.1% | Cruciani et al. 2007 |
| Ukraine | E1b1b1a2-V13 | 7.6% | Battaglia et al. 2009 |

**Middle East**

| **Population Studied** | **Nomenclature used for European E-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Qatar | E3b1a2-V13 | 1.4% | Cadenas et al. 2008 |
| Turkey | E-V13 | 2.7% | Cruciani et al. 2007 |
| United Arab Emirites | E3b1a2-V13 | 0.6% | Cadenas et al. 2008 |

**Caucasus**

| **Population Studied** | **Nomenclature used for European E-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Georgia | E1b1b1a2-V13 | 1.5% | Battaglia et al. 2009 |

## Appendix Table 9: Near Eastern J-Group (J2-M172)

### Eastern Europe

| Population Studied | Nomenclature used for Near Eastern J-Group | Percentage Found | Reference |
|---|---|---|---|
| Belarus | Haplogroup 9 | 2.0% | Rosser et al. 2000 |
| Belarus | J2 | 4.4% | Kharkov et al. 2005 |
| Bulgaria | Haplogroup 9 | 12.0% | Rosser et al. 2000 |
| Bulgaria | HG 9 | 23.1% | Malaspina et al. 2001 |
| Bulgaria | M-304 | 18.1% | Karachanak et al. 2009 |
| Moldavia | HG9 | 2.8% | Malaspina et al. 2001 |
| Poland | Haplogroup 9 | 4.0% | Rosser et al. 2000 |
| Poland | Eu 9 | 0.0% | Semino et al. 2000(a) |
| Poland | M172 | 1.0% | Semino et al. 2004 |
| Poland | J2-M172 | 2.5% | Kayser et al. 2005 |
| Poland | M172 | 1.0% | Battaglia et al. 2009 |
| Romania | Haplogroup 9 | 24.0% | Rosser et al. 2000 |
| Romania | HG 9 | 13.3% | Malaspina et al. 2001 |
| Romania | Haplogroup 9 | 12.1% | Stefan et al. 2001 |
| Romania | J2-M172 | 11.9% | Bosch et al. 2006 |
| Russia | Haplogroup 9 | 4.0% | Rosser et al. 2000 |
| Russia | HG 9 | 5.6% | Malaspina et al. 2001 |
| Russia (northern) | M172 | 4.0% | Wells et al. 2001 |
| Russia (northern) | J2 | 1.6% | Balanovsky et al. 2008 |
| Russia (central) | J2 | 2.4% | Balanovsky et al. 2008 |
| Russia (southern) | J2 | 3.0% | Balanovsky et al. 2008 |
| Russia | J2-M172 | 2.9% | Fechner et al. 2008 |

## Appendix Table 9: Near Eastern J-Group (J2-M172)

| | | | |
|---|---|---|---|
| Russia | J | 1.4% | Malyarchuk et al. 2008 |
| Russia (northeastern) | M172 | 1.3% | Mirabal et al. 2009 |
| Ukraine | Haplogroup 9 | 0.0% | Rosser et al. 2000 |
| Ukraine | Eu 9 | 6.0% | Semino et al. 2000(a) |
| Ukraine | M172 | 7.3% | Semino et al. 2004 |
| Ukraine | M172 | 6.6% | Battaglia et al. 2009 |

**Balkans**

| **Population Studied** | **Nomenclature used for Near Eastern J-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Albania | Eu 9 | 23.5% | Semino et al. 2000(a) |
| Albania | HG 9 | 24.4% | Malaspina et al. 2001 |
| Albania | M172 | 19.6% | Semino et al. 2004 |
| Albania | J2-M172 | 16.7% | Bosch et al. 2006 |
| Albania | M172 | 19.9% | Battaglia et al. 2009 |
| Bosnia-Herzegovina | J2-M172 | 6.3% | Marjanovic et al. 2005 |
| Bosnia-Herzegovina | M172 | 7.5% | Battaglia et al. 2009 |
| Croatia | Eu 9 | 5.2% | Semino et al. 2000(a) |
| Croatia | M172 | 6.2% | Semino et al. 2004 |
| Croatia | M172 | 5.9% | Battaglia et al. 2009 |
| Macedonia | Eu 9 | 15.0% | Semino et al. 2000(a) |
| Macedonia | J2-M172 | 11.5% | Bosch et al. 2006 |
| Macedonia | M172 | 14.9% | Battaglia et al. 2009 |

| | | | |
|---|---|---|---|
| Slovenia | Haplogroup 9 | 6.0% | Rosser et al. 2000 |
| Slovenia | M172 | 2.6% | Battaglia et al. 2009 |
| Former Yugoslavia | Haplogroup 9 | 8.0% | Rosser et al. 2000 |

**Mediterranean**

| **Population Studied** | **Nomenclature used for Near Eastern J-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Crete | HG 9 | 40.0% | Malaspina et al. 2001 |
| Crete | M172 | 37.5% | Martinez et al. 2007 |
| Crete | M172 | 30.6% | King et al. 2008 |
| Cyprus | Haplogroup 9 | 33.0% | Rosser et al. 2000 |
| Cyprus | HG 9 | 28.3% | Malaspina et al. 2001 |
| Cyprus | J2 | 6.2% | Capelli et al. 2006 |
| Greece | Haplogroup 9 | 28.0% | Rosser et al. 2000 |
| Greece | Eu 9 | 21.0% | Semino et al. 2000(a) |
| Greece | HG 9 | 20.3% | Malaspina et al. 2001 |
| Greece | J2 | 25.1% | Di Giacomo et al. 2003 |
| Greece | M172 | 20.6% | Semino et al. 2004 |
| Greece | J2-M172 | 22.0% | Bosch et al. 2006 |
| Greece | M172 | 14.6% | King et al. 2008 |
| Greece | M172 | 20.7% | Battaglia et al. 2009 |
| Italy | Haplogroup 9 | 20.0% | Rosser et al. 2000 |
| Italy (north-central) | Eu 9 | 14.0% | Semino et al. 2000(a) |
| Italy (Calabria) | Eu 9 | 21.6% | Semino et al. 2000(a) |
| Italy | HG9 | 26.6% | Malaspina et al. 2001 |

## Appendix Table 9: Near Eastern J-Group (J2-M172)

| | | | |
|---|---|---|---|
| Italy | J2 | 20.6% | Di Giacomo et al. 2003 |
| Italy (north-central) | M172 | 26.9% | Semino et al. 2004 |
| Italy (southern) | J2 | 16.1% | Capelli et al. 2006 |
| Italy | J2 | 20.0% | Capelli et al. 2007 |
| Italy (northeastern) | M172 | 15.5% | Battaglia et al. 2009 |
| Malta | J2 | 21.2% | Capelli et al. 2006 |
| Sardinia | Haplogroup 9 | 0.0% | Rosser et al. 2000 |
| Sardinia | Eu 9 | 5.2% | Semino et al. 2000(a) |
| Sardinia | HG 9 | 16.1% | Malaspina et al. 2001 |
| Sardinia | M172 | 9.7% | Semino et al. 2004 |
| Sardinia | J2 | 9.9% | Capelli et al. 2006 |
| Sardinia | M172 | 9.8% | Contu et al. 2008 |
| Italy (Sicily) | HG 9 | 8.7% | Malaspina et al. 2001 |
| Sicily | M172 | 16.7% | Semino et al. 2004 |
| Sicily | J2 | 22.6% | Capelli et al. 2006 |
| Sicily | M172/M241<br>M67/DYS445-6<br>M92/M12 | 25.8% | Di Gaetano et al. 2009 |

### Central Europe

| **Population Studied** | **Nomenclature used for Near Eastern J-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Czechoslovakia | Eu 9 | 8.9% | Semino et al. 2000(a) |
| Czech Republic | Haplogroup 9 | 11.0% | Rosser et al. 2000 |
| Czech Republic | HG9 | 6.5% | Malaspina et al. 2001 |

## Appendix Table 9: Near Eastern J-Group (J2-M172)

| | | | |
|---|---|---|---|
| Czech Republic | J2-M172/J2f-M67 | 3.5% | F. Luca et al. |
| Czech Republic | M172 | 5.3% | Battaglia et al. 2009 |
| Germany (Bavaria) | Haplogroup 9 | 5.0% | Rosser et al. 2000 |
| Germany | Haplogroup 9 | 3.0% | Rosser et al. 2000 |
| Germany | Eu 9 | 0.0% | Semino et al. 2000(a) |
| Germany | J2-M172 | 4.0% | Kayser et al. 2005 |
| Hungary | Haplogroup 9 | 3.0% | Rosser et al. 2000 |
| Hungary | Eu 9 | 2.2% | Semino et al. 2000(a) |
| Hungary | M172 | 2.0% | Semino et al. 2004 |
| Hungary | J2-M172/J2f-M67 | 6.7% | Völgyi et al. 2008 |
| Hungary | M172 | 1.9% | Battaglia et al. 2009 |
| Slovakia | Haplogroup 9 | 3.0% | Rosser et al. 2000 |
| Slovakia | HG 9 | 13.0% | Malaspina et al. 2001 |

### Western Europe

| **Population Studied** | **Nomenclature used for Near Eastern J-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Basques | Haplogroup 9 | 0.0% | Rosser et al. 2000 |
| Basques (Spain) | EU 9 | 0.0% | Semino et al. 2000(a) |
| Basques (France) | EU 9 | 4.5% | Semino et al. 2000(a) |
| Basques (Spain) | HG 9 | 1.2% | Malaspina et al. 2001 |
| Belgium | Haplogroup 9 | 5.0% | Rosser et al. 2000 |
| France | Haplogroup 9 | 5.0% | Rosser et al. 2000 |

## Appendix Table 9: Near Eastern J-Group (J2-M172)

| | | | |
|---|---|---|---|
| France | Eu 9 | 13.0% | Semino et al. 2000(a) |
| Iberia | J2/J2f | 7.3% | Flores et al. 2004 |
| Iberia | J2 | 7.7% | Adams et al. 2008 |
| Iceland | Haplogroup 9 | 0.0% | Rosser et al. 2000 |
| Ireland | Haplogroup 9 | 1.0% | Rosser et al. 2000 |
| Netherlands | Haplogroup 9 | 7.0% | Rosser et al. 2000 |
| Netherlands | Eu 9 | 0.0% | Semino et al. 2000(a) |
| Netherlands | M172 | 0.0% | Semino et al. 2004 |
| Portugal (northern) | Haplogroup 9 | 6.0% | Rosser et al. 2000 |
| Portugal (southern) | Haplogroup 9 | 9.0% | Rosser et al. 2000 |
| Portugal | HG 9 | 4.5% | Malaspina et al. 2001 |
| Portugal | J2 | 7.0% | Beleza et al. 2006 |
| Spain | Haplogroup 9 | 3.0% | Rosser et al. 2000 |
| Spain (Andalusia) | Eu 9 | 6.9% | Semino et al. 2000(a) |
| Spain (Catalan) | Eu 9 | 4.2% | Semino et al. 2000(a) |
| United Kingdom | Haplogroup 9 | Less than 1.0% | Rosser et al. 2000 |
| United Kingdom | HG9 | 5.0% | Malaspina et al. 2001 |
| United Kingdom | J2 | 1.9% | Capelli et al. 2003 |

### Northern Europe

| **Population Studied** | **Nomenclature used for Near Eastern<br>J-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Denmark | Haplogroup 9 | 7.0% | Rosser et al. 2000 |
| Finland | Haplogroup 9 | 0.0% | Rosser et al. 2000 |
| Finland | Haplogroup 9 | 0.0% | Zerjal et al. 2001 |

# Appendix Table 9: Near Eastern J-Group (J2-M172)

| | | | |
|---|---|---|---|
| Finland | J2 | 0.0% | Lappalainen et al. 2006 |
| Finland | J | 0.0% | Lappalainen et al. 2008 |
| Norway | Haplogroup 9 | 2.0% | Rosser et al. 2000 |
| Norway | Haplogroup 9 | 2.0% | Zerjal et al. 2001 |
| Saami | Haplogroup 9 | 0.0% | Rosser et al. 2000 |
| Saami | Eu 9 | 0.0% | Semino et al. 2000(a) |
| Saami | Haplogroup 9 | 0.0% | Zerjal et al. 2001 |
| Sweden (northern) | Haplogroup 9 | 2.0% | Rosser et al. 2000 |
| Sweden (Gotland) | Haplogroup 9 | 0.0% | Rosser et al. 2000 |
| Sweden (Gotland) | Haplogroup 9 | 0.0% | Zerjal et al. 2001 |
| Sweden | Haplogroup 9 | 0.0% | Zerjal et al. 2001 |
| Sweden | J | 3.6% | Karlsson et al. 2005 |
| Sweden | J | 0.0% | Lappalainen et al. 2008 |

## Baltic Region

| **Population Studied** | **Nomenclature used for Near Eastern J-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Estonia | Haplogroup 9 | 1.0% | Rosser et al. 2000 |
| Estonia | Haplogroup 9 | 0.0% | Zerjal et al. 2001 |
| Estonia | J | 1.7% | Lappalainen et al. 2008 |
| Latvia | Haplogroup 9 | 0.0% | Rosser et al. 2000 |
| Latvia | Haplogroup 9 | 0.0% | Zerjal et al. 2001 |
| Latvia | J | 0.0% | Lappalainen et al. 2008 |
| Lithuania | Haplogroup 9 | 0.0% | Rosser et al. 2000 |
| Lithuania | Haplogroup 9 | 0.0% | Zerjal et al. 2001 |

# Appendix Table 9: Near Eastern J-Group (J2-M172)

| | | | |
|---|---|---|---|
| Lithuania | J | 1.8% | Lappalainen et al. 2008 |

**Middle East**

| **Population Studied** | **Nomenclature used for Near Eastern J-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Lebanon | Eu 9 | 29.0% | Semino et al. 2000(a) |
| Lebanon | M172 | 30.0% | Wells et al. 2001 |
| Lebanon | M172 | 25.9% | Zalloua et al. 2008a |
| Iran | M172 | 23.1% | Wells et al. 2001 |
| Iran (northern) | M172 | 24.2% | Regueiro et al. 2006 |
| Iran (southern) | M172 | 23.1% | Regueiro et al. 2006 |
| Jordan | J-M172 | 15.8% | Flores et al. 2005 |
| Syria | Eu 9 | 15.0% | Semino et al. 2000(a) |
| Turkey | Haplogroup 9 | 33.0% | Rosser et al. 2000 |
| Turkey | Eu 9 | 40.0% | Semino et al. 2000(a) |
| Turkey | HG 9 | 34.9% | Malaspina et al. 2001 |
| Turkey | M172 | 24.3% | Cinnioğlu et al. 2004 |

**Caucasus**

| **Population Studied** | **Nomenclature used for Near Eastern J-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Armenia | Haplogroup 9 | 29.0% | Rosser et al. 2000 |
| Armenia | J2 | 24.0% | Nasidze et al. 2003 |
| Georgia | Haplogroup 9 | 23.0% | Rosser et al. 2000 |

## Appendix Table 9: Near Eastern J-Group (J2-M172)

| | | | |
|---|---|---|---|
| Georgia | Eu 9 | 33.0% | Semino et al. 2000(a) |
| Georgia | M172 | 32.8% | Wells et al. 2001 |
| Georgia | M172 | 26.7% | Semino et al. 2004 |
| Georgia | M172 | 31.8% | Battaglia et al. 2009 |
| North Caucasus | J2 | 12.5% | Nasidze et al. 2004a |
| South Caucasus | J2 | 25.6% | Nasidze et al. 2004a |

### South-Central Asia

| **Population Studied** | **Nomenclature used for Near Eastern J-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| India | M172 | 10.8% | Wells et al. 2001 |
| India | M172 | 11.4% | Kivisild et al. 2003 |
| India (northeastern) | J-M172 | 0.6% | Cordaux et al. 2004 |
| Pakistan/India | M172 | 15.9% | Semino et al. 2004 |
| India | J-M172 | 9.1% | Sengupta et al. 2006 |
| Pakistan | J-M172 | 11.9% | Sengupta et al. 2006 |

## Appendix Table 10: J2a-M410 and J2b-M12/M102

| | **J2a as a percentage of gene pool** | **J2b as a percentage of gene pool** | **J2a as a percentage of J2** | **J2b as a percentage of J2** |
|---|---|---|---|---|
| Balkans[4] | 3.69% | 5.45% | 40.35% | 59.65% |
| Greece[5] | 8.77% | 5.85% | 60.00% | 40.00% |
| Crete[6] | 27.50% | 3.10% | 89.83% | 10.17% |
| Italy[7] (mainland) | 22.92% | 3.33% | 87.30% | 12.70% |
| Italy[8] (northeastern) | 13.50% | 0.00% | 100.00% | 0.00% |
| Sicily[9] | 15.25% | 0.85% | 94.44% | 5.56% |
| Czech Republic[10] | 0.00% | 5.30% | 0.00% | 100.00% |

---

[4] Battaglia et al. 2009
[5] King et al. 2009
[6] King et al. 2008
[7] Semino et al. 2004
[8] Battaglia et al. 2009
[9] Di Gaetano et al. 2009
[10] Battaglia et al. 2009

# Appendix Table 10: J2a-M410 and J2b-M12/M102

| | | | | |
|---|---|---|---|---|
| Poland[11] | 0.00% | 1.00% | 0.00% | 100.00% |
| Russia[12] | 0.42% | 0.85% | 33.33% | 66.66% |
| Ukraine[13] | 3.30% | 3.30% | 50.00% | 50.00% |
| Turkey[14] | 22.56% | 1.72% | 92.91% | 7.09% |
| Georgia[15] | 31.80% | 0.00% | 100.00% | 0.00% |

---

[11] Battaglia et al. 2009
[12] Mirabal et al. 2009
[13] Battaglia et al. 2009
[14] Cinnioğlu et al 2004
[15] Battaglia et al. 2009

## Appendix Table 11: Caucasus G-Group (G-M201)

### Mediterranean

| **Population Studied** | **Nomenclature used for Caucasus G-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Crete | G-M201 | 7.0% | Di Giacomo et al. 2003 |
| Crete | G-M201 | 7.1% | Martinez et al. 2007 |
| Crete | G-M201 | 10.9% | King et al. 2008 |
| Corsica | G-M201 | 11.8% | Francalacci et al. 2003 |
| Greece | Eu 11 | 2.6% | Semino et al. 2000(a) |
| Greece | G-M201 | 5.2% | Di Giacomo et al. 2003 |
| Greece | G-M201 | 4.9% | Bosch et al. 2006 |
| Greece | G-M201 | 4.7% | King et al. 2008 |
| Greece | G-M201 | 3.3% | Battaglia et al. 2009 |
| Italy (Central-Northern) | Eu 11 | 10.0% | Semino et al. 2000(a) |
| Italy (Calabria) | Eu 11 | 8.0% | Semino et al. 2000(a) |
| Italy | G-M201 | 6.3% | Di Giacomo et al. 2003 |
| Italy | G-M201 | 11.0% | Capelli et al. 2006 |
| Italy (northeastern) | G-M201 | 11.9% | Battaglia et al. 2009 |
| Sardinia | Eu 11 | 14.2% | Semino et al. 2000(a) |
| Sardinia | G-M201 | 13.9% | Zei et al. 2003 |
| Sardinia | G-M201 | 14.1% | Francalacci et al. 2003 |
| Sardinia | G-M201 | 12.6% | Contu et al. 2008 |
| Sicily | G-M201 | 11.8% | Francalacci et al. 2003 |
| Sicily | G-M201 | 5.9% | Di Gaetano et al. 2009 |

# Appendix Table 11: Caucasus G-Group (G-M201)

## Balkans

| Population Studied | Nomenclature used for Caucasus G-Group | Percentage Found | Reference |
|---|---|---|---|
| Albania | Eu 11 | 2.0% | Semino et al. 2000(a) |
| Albania | G-M201 | 3.3% | Bosch et al. 2006 |
| Albania | G-M201 | 1.8% | Battaglia et al 2009 |
| Bosnia-Herzegovina | G-M201 | 2.0% | Marjanovic et al. 2005 |
| Bosnia-Herzegovina | G-M201 | 1.4% | Peričić et al. 2005(b) |
| Bosnia-Herzegovina | G-M201 | 3.1% | Battaglia et al. 2009 |
| Croatia | Eu 11 | 1.7% | Semino et al. 2000(a) |
| Croatia (mainland) | G-M201 | 0.9% | Barac et al. 2003 |
| Croatia (mainland) | G-M201 | 0.9% | Peričić et al. 2005(a) |
| Croatia | G-M201 | 6.8% | Battaglia et al. 2009 |
| Macedonia | G-M201 | 5.1% | Peričić et al. 2005(b) |
| Macedonia | G-M201 | 3.8% | Bosch et al. 2006 |
| Macedonia | G-M201 | 1.7% | Battaglia et al. 2009 |
| Slovenia | G-M201 | 2.6% | Battaglia et al. 2009 |

# Appendix Table 11: Caucasus G-Group (G-M201)

## Western Europe

| **Population Studied** | **Nomenclature used for Caucasus**<br>**G-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Portugal | G-M201 | 5.5% | Beleza et al. 2006 |
| Spain (Cantabria) | G-M201 | 7.6% | Maca-Meyer et al, 2003 |
| Spain (Catalan) | Eu 11 | 8.3% | Semino et al. 2000(a) |
| Spain (Pyrenees Region) | G-M201 | 1.8% | López-Parra et al. 2009 |
| Iberia | G-M201 | 4.0% | Flores et al. 2004 |
| Iberia | G-M201 | 4.3% | Alonso et al. 2005 |
| Iberia | G-M201 | 5.0% | Adams et al. 2008 |

## Central Europe

| **Population Studied** | **Nomenclature used for Caucasus**<br>**G-Group** | **Percentage Found** | **Reference** |
|---|---|---|---|
| Czech Republic | G-M201 | 9.3% | Luca et al. 2007 |
| Czech Republic | M201 | 4.0% | Battaglia et al. 2009 |
| Czechoslovakia | Eu 11 | 4.4% | Semino et al. 2000(a) |
| Hungary | Eu 11 | 2.2% | Semino et al. 2000(a) |
| Hungary | G-M201 | 4.2% | Völgyi et al. 2008 |
| Hungary | G-M201 | 1.9% | Battaglia et al. 2009 |

## Appendix Table 11: Caucasus G-Group (G-M201)

### Northern Europe

| Population Studied | Nomenclature used for Caucasus G-Group | Percentage Found | Reference |
|---|---|---|---|
| Sweden | G-M201 | 1.6% | Karlsson et al. 2006 |

### Eastern Europe

| Population Studied | Nomenclature used for Caucasus G-Group | Percentage Found | Reference |
|---|---|---|---|
| Belarus | G-M201 | 1.5% | Kharkov et al. 2005 |
| Bulgaria | G-M201 | 1.6% | Karachanak et al. 2009 |
| Romania | G-M201 | 10.5% | Bosch et al. 2006 |
| Moldavia | G-M201 | 4.3% | Nasidze et al. 2006 |
| Russia (European) | G-M201 | 1.8% | Fechner et al. 2008 |
| Russia (northern) | G-M201 | 1.2% | Balanovsky et al. 2008 |
| Russia (central) | G-M201 | 0.0% | Balanovsky et al. 2008 |
| Russia (southern) | G-M201 | 1.0% | Balanovsky et al. 2008 |
| Russia | G-M201 | 1.2% | Malyarchuk and Derenko 2008 |
| Ukraine | Eu 11 | 4.0% | Semino et al. 2000(a) |
| Ukraine | G-M201 | 3.3% | Battaglia et al. 2009 |

## Appendix Table 11: Caucasus G-Group (G-M201)

**Middle East**

| Population Studied | Nomenclature used for Caucasus G-Group | Percentage Found | Reference |
|---|---|---|---|
| Iran | G-M201 | 5.4% | Nasidze 2004(a) |
| Iran (northern) | G-M201 | 15.2% | Regueiro et al. 2006 |
| Iran (southern) | G-M201 | 12.8% | Regueiro et al. 2006 |
| Jordan | G-M201 | 5.9% | Flores et al. 2005 |
| Lebanon | G-M201 | 6.6% | Zalloua et al. 2008 |
| Turkey | Eu 11 | 6.6% | Semino et al. 2000(a) |
| Turkey | G-M201 | 10.9% | Cinnioğlu et al. 2004 |
| Turkey | G-M201 | 0.0% | Nasidze 2004(a) |

**Caucasus**

| Population Studied | Nomenclature used for Caucasus G-Group | Percentage Found | Reference |
|---|---|---|---|
| Armenia | G-M201 | 11.0% | Nasidze et al. 2003 |
| Azerbaijan | G-M201 | 18.0% | Nasidze et al. 2003 |
| Georgia | Eu 11 | 30.1% | Semino et al. 2000(a) |
| Georgia | G-M201 | 31.0% | Nasidze et al. 2003 |
| Georgia | M201 | 31.8% | Battaglia et al. 2009 |
| Caucasus Region (17 different populations) | G-M201 | 21.1% | Nasidze 2004(a) |

# Appendix Table 11: Caucasus G-Group (G-M201)

## South-Central Asia

| Population Studied | Nomenclature used for Caucasus G-Group | Percentage Found | Reference |
|---|---|---|---|
| Pakistan | G-M201 | 6.3% | Sengupta et al. 2006 |