


4-30-2011

Ancient Mitochondrial DNA From Pre-historic Southeastern Europe: The Presence of East Eurasian Haplogroups Provides Evidence of Interactions with South Siberians Across the Central Asian Steppe Belt

Jeremy R. Newton
Grand Valley State University

Follow this and additional works at: <http://scholarworks.gvsu.edu/theses>

 Part of the [Cell Biology Commons](#), and the [Molecular Biology Commons](#)

Recommended Citation

Newton, Jeremy R., "Ancient Mitochondrial DNA From Pre-historic Southeastern Europe: The Presence of East Eurasian Haplogroups Provides Evidence of Interactions with South Siberians Across the Central Asian Steppe Belt" (2011). *Masters Theses*. 5.
<http://scholarworks.gvsu.edu/theses/5>

This Thesis is brought to you for free and open access by the Graduate Research and Creative Practice at ScholarWorks@GVSU. It has been accepted for inclusion in Masters Theses by an authorized administrator of ScholarWorks@GVSU. For more information, please contact scholarworks@gvsu.edu.

ANCIENT MITOCHONDRIAL DNA FROM PRE-HISTORIC SOUTH-
EASTERN EUROPE: THE PRESENCE OF EAST EURASIAN HAPLOGROUPS
PROVIDES EVIDENCE OF INTERACTIONS WITH SOUTH SIBERIANS ACROSS
THE CENTRAL ASIAN STEPPE BELT

A thesis submittal in partial fulfillment of
the requirements for the degree of
Master of Science

By

Jeremy R. Newton

To

Cell and Molecular Biology Department
Grand Valley State University
Allendale, MI
April, 2011

“Not all those who wander are lost.”

J.R.R. Tolkien

ACKNOWLEDGEMENTS

I would like to extend my sincerest thanks to every person who has motivated, directed, and encouraged me throughout this thesis project. I especially thank my graduate advisor, Dr. Alex Nikitin, for his advice, direction, and vision over the past two years. I also would like to thank my Graduate Committee, Dr. Roderick Morgan and Dr. Agnieszka Szarecka, for their contributions to this manuscript and for their continual support during my time at GVSU. Additionally, I am grateful to Jessica Riley for her willingness to help with laboratory work, as it helped bring this research to a successful conclusion. I would also like to thank Dr. Ryan Thum and Dustin Weisel of Annis Water Resources Institute for their assistance with the operation of their ABI 3130xl Genetic Analyzer. Finally, I would like to thank Grand Valley State University for awarding a Presidential Research Grant to provide the funding needed to bring this project to fruition.

ABSTRACT

ANCIENT MITOCHONDRIAL DNA FROM PRE-HISTORIC SOUTH-EASTERN EUROPE: THE PRESENCE OF EAST EURASIAN HAPLOGROUPS PROVIDES EVIDENCE OF INTERACTIONS WITH SOUTH SIBERIANS ACROSS THE CENTRAL ASIAN STEPPE BELT

Studies of mitochondrial DNA (mtDNA) polymorphism have provided valuable insights for understanding patterns of human migration and interaction. The ability to recover ancient mtDNA sequence data from post-mortem bone and tissue samples allows us to view snapshots of historic gene pools firsthand, provided that great care is taken to prevent sample contamination. In this study, we analyzed the DNA sequence of the first hypervariable segment (HVSI) of the mtDNA control region, as well as a portion of the coding region, in 14 individuals from three collective burials from the Neolithic Dnieper-Donetz culture and three individuals from Bronze Age Kurgan burials, all located in modern-day Ukraine on the northern shores of the Black Sea (the North Pontic Region, or NPR). While most of our samples possessed mtDNA haplotypes that can be linked to European and Near Eastern populations, three Neolithic and all three Bronze Age individuals belonged to mtDNA haplogroup C, which is common in East Eurasian, particularly South Siberian, populations but exceedingly rare in Europe. Phylogeographic network analysis revealed that our samples are located at or near the ancestral node for haplogroup C and that derived lineages branching from the Neolithic samples were present in Bronze Age Kurgans. In light of the numerous examples of mtDNA admixture that can be found in both Europe and Siberia, it appears that the NPR and South Siberia are located at opposite ends of a genetic continuum established at some point prior to the Neolithic. This migration corridor may have been established during the Last Glacial

Maximum due to extensive glaciation in northern Eurasia and a consequent aridization of western Asia. This implies the demographic history for the European gene pool is more complex than previously considered and also has significant implications regarding the origin of Kurgan populations.

TABLE OF CONTENTS

LIST OF TABLES.....	ix
LIST OF FIGURES.....	x
INTRODUCTION.....	1
PROPERTIES OF MITOCHONDRIAL DNA.....	1
MTDNA IN POPULATION GENETICS.....	5
SURVEY OF GLOBAL MTDNA DIVERSITY.....	8
ANCIENT MTDNA RESEARCH.....	14
PHYLOGEOGRAPHIC ANALYSIS OF HUMAN MTDNA.....	18
THESIS PURPOSE STATEMENT.....	23
MATERIALS AND METHODS.....	25
SAMPLE COLLECTION, EXTRACTION, AND AMPLIFICATION.....	25
MOLECULAR CLONING.....	27
TYPING OF CODING REGION POLYMORPHISMS.....	27
SAMPLE SEQUENCING.....	28
NETWORK ANALYSIS.....	28
RESULTS.....	30
MTDNA HVSI SEQUENCES IN NEOLITHIC NPR INHABITANTS	30
MTDNA HVSI SEQUENCES IN BRONZE AGE KURGANS.....	31
NETWORK FOR HAPLOGROUP U5A.....	32
NETWORK FOR HAPLOGROUP C.....	33
DISCUSSION.....	36
WEST EURASIAN LINEAGES IN NEOLITHIC NPR.....	37
HAPLOGROUP H.....	37

HAPLOGROUP T.....	40
HAPLOGROUP U.....	41
ADMIXTURE OF EAST EURASIAN MTDNA IN WEST EURASIAN POPULATIONS.....	42
NON-GENETIC INDICATORS OF EAST EURASIAN INTERACTIONS.....	46
ADMIXTURE OF EAST EURASIAN MTDNA IN WEST EURASIAN POPULATIONS.....	47
PRESENCE OF A STEPPE BELT MIGRATION CORRIDOR.....	49
CONCLUSIONS.....	52
LITERATURE CITED.....	53

LIST OF TABLES

Table 1. MtDNA HVSI profiles of researchers performing aDNA analysis or involved in prior handling of the specimens.....	66
Table 2. Primer pairs used to amplify mtDNA HVSI in Neolithic and Bronze Age bone samples.....	66
Table 3. MtDNA HVSI sequences of Neolithic and Bronze Age individuals from the North Pontic Region.....	67
Table 4. Differences from the rCRS in mtDNA HVSI of pre-historic NPR inhabitants.....	68

LIST OF FIGURES

Figure Legends.....	69
Figure 1a. Illustration of human mtDNA (modified from Pakendorf and Stoneking, 2005).....	71
Figure 1b. Revised Cambridge Reference Sequence (Accession Number: NC_012920) for human mtDNA from positions 16001 to 16400.....	72
Figure 2. Schematic representation of mtDNA haplogroups.....	73
Figure 3. Geographic location of burial sites utilized in this study.....	74
Figure 4. Median-Joining Network for mtDNA Haplogroup U5a.....	75
Figure 5. Median-Joining Network for mtDNA Haplogroup C.....	76
Figure 6. Map of Europe showing geographic locations of haplogroup C.....	77
Figure 7. Map of Europe showing geographic locations of haplogroup T4.....	78

INTRODUCTION

Properties of Mitochondrial DNA

Mitochondrial DNA (mtDNA) polymorphism has been widely utilized to study the phylogeographic relationships within and between human populations (van Oven and Kayser, 2009). The extranuclear mtDNA genome is a 16,569-bp molecule of circular DNA containing a subset of genes required to produce proteins essential for cellular respiration (Figure 1a). The vast majority of the mtDNA genome consists of these protein coding regions. The remainder of the mtDNA molecule is composed of a 1,100-bp control region flanking either side of the origin of replication (Pakendorf and Stoneking, 2005).

The mechanism of transcription and DNA replication of the mitochondrial genome differs greatly from related nuclear processes. Although mutations can arise during any replication event, heritable mutations must occur in oocyte mitochondria. These cells are arrested in metaphase until fertilization occurs and cleavage begins. During the first rounds of cell division and through all pre-implantation stages of embryogenesis, mtDNA replication is virtually halted (St. John et al., 2010). Following implantation, mitochondrial replication is allowed to proceed as cell differentiation and proliferation progresses in the blastocyst, which provides a favorable environment for the generation of mtDNA mutations.

The mitochondrial control region contains multiple promoters and protein-binding domains required to regulate both transcription and replication. Although the majority of mitochondrial transcripts are found on the heavy strand (H-strand), so named for its elevated guanine content, there are eight tRNAs and one protein-coding gene on the light strand (L-strand; Fernández-Silva et al., 2003). Promoters for each DNA strand are

located in the control region upstream from the H-strand origin of replication (O_H). The L-strand promoter (LSP) is oriented the same direction as the O_H domain (clockwise in Fig. 1a), and the H-strand promoter (HSP) faces in the opposite direction (Clayton, 2000). The HSP is further divided into two subdomains, H_1 and H_2 (Montoya et al., 1982). Initiation is facilitated by a mitochondrial transcription factor, mtTFA, which uncoils the mtDNA helix and exposes the template (Clayton, 2000). MtTFA may also be involved with regulating mtDNA copy number (Ekstrand et al., 2004). Other possible roles for mtTFA include mtDNA damage repair and acting as a mitochondrial version of nuclear histones (Fernández-Silva et al., 2003). MtTFA transcription itself takes place in the nucleus, along with virtually all other proteins needed for mtDNA replication and transcription (Davis and Clayton, 1996) and is regulated by the nuclear transcription factor NRF-1 (Clayton, 2000). Transcription is carried out by a mitochondrial RNA polymerase that shows homology with bacterial polymerases (Clayton, 2000), consistent with the endosymbiotic origin of mitochondria in eukaryotic cells. This mitochondrial polymerase must be complexed with one of two possible transcription factors (either TFB1M or TFB2M) to carry out transcription (Fernández-Silva et al., 2003). Transcription of either strand proceeds around the entire molecule, producing a large polycistronic RNA that is processed into individual RNAs prior to translation. Since transcription and replication are fairly slow in mitochondria – covering the entire genome takes about 1 hour (Clayton, 2000) – it is unlikely that both strands are transcribed at once. Transcription termination results from an interaction between the polymerase and the termination factor mTERF. The precise mechanism is not known, although it appears

to be more sophisticated than mTERF simply acting as a physical barrier (Fernández-Silva et al., 2003).

MtDNA replication also occurs in a fairly unusual manner and is linked to mitochondrial transcription through the LSP site. The RNA primer for H-strand replication is built at the LSP by RNA polymerase after mtTFA has opened the helix (Chang and Clayton, 1985), which hints at an underlying balance of replication and transcription in mtDNA. The RNA segment is processed into a mature primer possessing a 3'-OH group by an mtRNA-processing endoribonuclease (RNase MRP; Lee and Clayton, 1998). The mitochondrial DNA polymerase (γ -DNApol) can then begin to replicate the H-strand (clockwise in Fig. 1a). However, one of two outcomes can result once replication begins (Fernández-Silva et al., 2003). Usually the DNA strand will cease replication at a termination associated segment (TAS) in the mtDNA HVSI, forming a structure called the displacement loop (D-loop) consisting of the L-strand complemented by the short, nascent H-strand and an unpaired H-strand (Clayton, 2000). The TAS may work to halt replication by acting as a protein-binding site (Roberti et al., 1998). In some cases replication will proceed past the D-loop, proceeding around the molecule and displacing the original H-strand. After replication has progressed about two-thirds around the mtDNA, the L-strand origin of replication (O_L) is reached. The O_L forms a stem-loop structure in its single stranded form. The loop region acts as a template for a mitochondrial primase to build an RNA primer to initiate L-strand replication (Fernández-Silva et al., 2003), which proceeds along the exposed parental H-strand in the opposite direction of H-strand replication (counterclockwise in Fig 1a). This method of replication allows both new DNA strands to be replicated as leading strands. MtDNA

replication is completed by separating the daughter molecules, removing, filling, and ligating primer sites, and re-introducing the supercoiled structure of the completed molecule.

Examining the details of mitochondrial replication and transcription reveals that the control region has multiple loci that are actively involved in the regulation of both processes. In spite of this knowledge, it is rare for population genetics research to acknowledge the possibility of selection on the control region, even though some portions of the control region accumulate mutations more readily than others. The main focus is on the primary DNA structure, and the possible influence of secondary structural elements is often not considered. However, a recent study by Pereira et al. (2008) identified 13 different secondary structural elements – both hairpins and tRNA-like structures – scattered throughout the control region. Most of these structures have fewer mutations in their stem regions than in their loop domains. Although most of these structures were small and did not have clear biological significance (according to the authors), the largest structure covers a 93-bp stretch (positions 16028-16120) of the mtDNA HVSI domain (Pereira et al., 2008). The density of mutations in this region is lower than in surrounding areas and appears to be under strong purifying selection (Tajima's $D = -2.182$). Another smaller (37 bp; positions 16124-16160) hairpin is present in proximity to the TAS domain, also in the HVSI. These structures may be involved in forming the D-loop by halting H-strand replication. Other possible functions include post-transcriptional processing of polycistronic RNA and the execution of transcription termination (Pereira et al., 2008). The O_H domain and LSP are both located on hairpin structures as well; these are outside the HVSI but still in the control region (Pereira et al.,

2008; Figure S2). It is noteworthy that the remainder of the HVSI has very little secondary structure, although three small hairpins are present (Pereira et al., 2008). The first is 23 bp (positions 16310-16332), the second is 14 bp (positions 16360-16373), and the third is 15 bp (positions 16377-16391). This relative lack of superstructure correlates with the general observation that the majority of mtDNA control region polymorphisms can be found in the last two-thirds of the HVSI. As further studies uncover additional structural and functional control region elements, it is incumbent upon population geneticists to evaluate the impact these structures will have on their research.

MtDNA in Population Genetics

There are several characteristics of mtDNA that make it a useful tool for population genetics research. The first and most important property of mtDNA is its strict maternal inheritance (Giles et al., 1980) which has been especially useful for forensic and population genetics research (Cann et al., 1987; Torroni et al., 1996; Richards et al., 1998; Pfeiffer et al., 1999; Torroni et al., 2006; Parson and Dür, 2007). However, one exception to the rule of maternal inheritance can be found in the literature and concerns a 28-year-old subject unable to perform even basic exercise due to defective cellular respiration in his skeletal muscle tissue (Shwartz and Vissing, 2002). While mtDNA from the subject's blood was maternally inherited, the mtDNA found within the skeletal muscle tissue was identical to the subject's father. Normally, paternal mtDNA would be eliminated in early development (Shitara et al., 2000), which would limit heteroplasmy (the presence of multiple mtDNA sequence variants in an organism). Another attribute of mtDNA is an overall lack of recombination in mtDNA molecules. One paper purported to

have directly observed recombination in human mtDNA (Hagelberg et al., 1999), but was later retracted due to problems with their DNA sequence alignments (Hagelberg et al., 2000). Even if recombination were to occur, it is unlikely to generate a detectable product, since any recombination would likely take place between two identical molecules (Rokas et al., 2003). Compared to nuclear DNA, mtDNA also has an elevated mutation rate. Early studies estimated the mtDNA mutation rate at 2.0×10^{-8} substitutions/site/yr (Brown, 1979), which exceeds the genomic mutation rate of 2.2×10^{-9} substitutions/site/yr by almost tenfold (Kumar and Subramanian, 2002). More recent estimates of 1.7×10^{-8} substitutions/site/year for the mtDNA coding region support the early results (Ingman et al., 2000). The control region mutates even more rapidly with an estimated mutation rate of 3.2×10^{-7} substitutions/site/yr in humans (Sigurðardottir et al., 2000). Additionally, there are several sites within the control region considered to be “hotspots” for mutation (Gilbert et al., 2003). For the purposes of population genetics studies, the elevated mutation rate of mtDNA is important because it leads to a fixation of single nucleotide polymorphisms (SNPs) within and between populations (Brown et al., 1979; Brown, 1980). This further leads to the emergence of diagnostic mutations in different human populations. The variation in mtDNA mutation profiles is vital for phylogeographic studies of human dispersals and interactions, since they provide the ability to determine the geographic origin of an individual’s matrilineal genome.

This group-specific genetic variation allows all human mtDNA genomes to be grouped into distinct clusters called haplogroups based on diagnostic nucleotide polymorphisms. Each haplogroup can be identified based upon the sum of control and coding region variation compared to the established Cambridge Reference Sequence

(rCRS; Figure 1b) (Anderson et al., 1981; Andrews et al., 1999). The coding region variation is diagnostic for most haplogroups (Torroni et al., 1992; Torroni et al., 1993; Santos et al., 2004). For molecular analytic purposes, the control region is split into two hypervariable regions (HVSI and HVSII). The HVSI is more widely used due to its greater variability compared to HVSII.

As a general rule, DNA sequence data from both control and coding regions is required to definitively assign a haplogroup. Although most of the mtDNA polymorphisms used for haplogroup identification are located in the coding region, many haplogroups can be readily identified using HVSI polymorphisms. The following example illustrates the haplotyping procedure based on HVSI variation. In this case, we would first obtain the HVSI DNA sequence of an individual and align it against the rCRS. For instance, one of the researchers involved in this study has differences to the rCRS at five positions (Table 1). The mutation at position 16126 indicates this individual belongs to either haplogroup J or T (Torroni et al., 1996; Richards et al., 1998), while an additional mutation at position 16294 confirms the individual belongs to haplogroup T (haplogroup J has different mutation at position 16069). While some polymorphisms are used to define entire haplogroups, additional derivation can sometimes denote sublineages within a haplogroup. For instance, the researcher in Table 1 belonging to haplogroup T has additional mutations at positions 16163, 16186, and 16189, which are diagnostic for subhaplogroup T1 (Richards et al., 1998).

In some cases the HVSI DNA sequence of an individual is identical to the rCRS or has no differences that are diagnostic for any given lineage. While it could be argued that the lack of variation indicates membership in haplogroup H (of which the rCRS is a

member), HVSI sequences identical to the rCRS have been observed in other haplogroups (Pereira et al., 2000; Zhang et al., 2010). Multiple sites within the control region are also homoplastic, which can be problematic when attempting to accurately assign haplogroups based solely on control region variation. For example, a T-C transition at position 16298 is part of the diagnostic HVSI motif for N-derived haplogroup V and M-derived haplogroups C and Z (Fig. 2), and it has also been detected in members of haplogroup T (Finnilä et al., 2001). If control region polymorphisms are insufficient, analysis of variable positions within the coding region can be used to type ambiguous samples. Hierarchical schemes have been developed to provide assistance when assigning haplogroups based on coding region polymorphisms (Santos et al., 2004). Endonuclease digestion was originally used to define haplogroups (Brown, 1980; Torroni et al., 1994; Santos et al., 2004), but modern studies often analyze diagnostic coding region polymorphisms by DNA sequencing. It is becoming more common to collect DNA sequence data for the entire mtDNA genome (van Oven and Kayser, 2009; Derenko et al., 2010; Malyarchuk et al., 2010). Whole-genome analysis provides the best resolution to ensure accurate haplotyping, especially when studying rare haplogroups or groups with few control region polymorphisms.

Survey of Global mtDNA Diversity

One of the earliest studies of mtDNA polymorphism (Brown, 1980) noted variability in restriction endonuclease fragment sizes in individuals from various global populations, concluding there was a “strong possibility” of regional mtDNA polymorphisms. Further research confirmed distinct distributions for many mtDNA

haplogroups. Although some lineages are poorly understood because of their scarcity, much progress has been made elucidating patterns of human migration with mtDNA. A basic schematic of the mitochondrial haplogroup tree is outlined in Figure 2. More recent projects have produced trees based upon whole-mtDNA genome sequencing (van Oven and Kayser, 2009).

Determination of the origins of human mtDNA lineages is accomplished using a phylogeographic approach rooted in the principles of coalescent theory. Coalescent theory works backward to estimate the timing of evolutionary events that have produced the genetic diversity in modern populations by modeling gene genealogies that describe when alleles coalesce (Kingman, 1982; Rosenberg and Nordborg, 2002). Coalescence is the point in time when two alleles share a common ancestor; it has also been explained as the point in time when two alleles “pick the same parent” (Rosenberg and Nordborg, 2002). Ultimately, all alleles in the sample will coalesce to a most recent common ancestor (MRCA). Coalescent theory is more efficient than traditional population genetics tools for the purpose of tracing lineages, because it only describes the genealogy of the sample as opposed to the entire population (Fu and Li, 1999). This approach is useful for phylogeography, since the goal of this field is to understand how past evolutionary events have shaped current gene diversity at the population level (Freeland, 2005). Coalescent theory operates under the assumption that all alleles in a given population are identical by descent so that the operative question becomes how, not if, the sampled alleles are related to one another.

All human mtDNA sequences can be traced back to a single lineage in Africa approximately 200, 000 years ago (Cann et al., 1987), sometimes referred to as

“Mitochondrial Eve.” Members of ancestral African populations belonged to mtDNA haplogroup L (Gonder et al., 2007). Early modern humans migrated from Africa to the Middle East before spreading across the globe, ultimately producing current mtDNA lineage diversity (McNiell, 1984; Richards et al., 1998). One particular subgroup, L3, migrated “Out of Africa” into the Near East 46-62 thousand years ago (kya) (Forster et al., 2004). Macrohaplogroups M and N are descended from L3, making the L3 lineage the source of virtually all non-African mtDNA diversity (Fig. 2; Maca-Meyer et al., 2001; Forster et al., 2004).

Within macrohaplogroup N, lineages are further subdivided into two major clusters: 1) direct descendents of N (also labeled N*) or 2) descendents of the N-derived macrohaplogroup R (van Oven and Kayser, 2009), which diverged from N shortly after its arrival in the Near East (Torroni et al., 2006). Nine N-derived haplogroups (H, V, U, K, T, J, I, W, X) make up almost all mtDNAs in modern European and Near Eastern populations (Fig.2; Richards et al., 1998). Of these nine, six (H, V, U, K, T, J) are descended from haplogroup R. The other three groups (I, W, X) are descended directly from haplogroup N and are generally rare in modern populations. While some early work seemed to indicate that I, W, and X were related to one another and formed their own cluster branching from N, more recent whole genome studies suggest that I, W, and X each descend from separate sub-branches within N. Haplogroup W is very poorly studied and generally occurs in less than 5% of individuals throughout Europe and the Near East. It is difficult to say whether slightly elevated concentrations of haplogroup W in Finns (9%) and Pakistanis (5%; Metspalu et al., 2004) are related to a point of origin for this group or are simply the result of a founder effect. Haplogroup I is also rare throughout

Europe and the Near East. It appears to be one of the oldest groups to distinguish itself from N (Richards et al., 2000). Despite its general scarcity, some isolated European populations have elevated levels of haplogroup I. Croatians living on Krk Island in the Adriatic Sea and Lemkos living in the Carpathian Mountains each have haplogroup I frequencies of 11.3% (Tolk et al., 2000; Nikitin et al., 2009). The elevated frequencies of haplogroup I in these populations is again likely due to a founder effect, although in the case of the Carpathian highlanders haplogroup I appears to be diversified and is represented by both ancestral and derived lineages (Nikitin, unpublished data).

Haplogroup X has perhaps the most unusual geographic distribution of any mitochondrial lineage. It appears to be of Near Eastern origin but has been detected in both European and Native American populations (Reidla et al., 2003). The intrigue surrounding haplogroup X is based on the fact that it is absent from populations in Northeast Asia (Starikovskaya et al., 1998), which makes its route to the New World unclear. Haplogroup X was detected at very low frequencies in Siberian Altaians (Derenko et al., 2001), but this is thought to be a result of more recent gene flow from Europe.

The six European lineages descended from haplogroup R are the most common on the subcontinent today. Some deeply-rooted lineages of haplogroup R are still present at appreciable levels in South Asia (Metspalu et al., 2004) but are generally scarce elsewhere. Haplogroup H is by far the most common group in Europe, comprising 50% or more of some samples. It shares a common ancestor with Haplogroup V, which is elevated in the Basques of Northern Spain (Torroni et al., 1998). This ancestral type, referred to as haplogroup HV, originated in the Near East and entered Europe during the

middle Paleolithic (Richards et al., 2000). Haplogroups J and T also share a common ancestor; both haplogroup J and the T1 subhaplogroup are considered classic markers of the Neolithic expansion from the Near East into Europe approximately 9,000 years ago (Richards et al., 2000). Haplogroup K is another European lineage that has actually been reclassified as a sublineage of haplogroup U (van Oven and Kayser, 2009). The famous Tyrolean Iceman belongs to this lineage (Endicott et al., 2009). Haplogroup U itself is one of the oldest and most abundant mitochondrial lineages. It originated in the Near East but has various subgroups that have more specific distributions in Europe (U4 and U5), North Africa (U6), the Near East (U1 and U3), and South Asia (U2 and U7).

East Eurasian populations have a more complex mixture of both N- and M-derived haplogroups (Fig.2). M-derived lineages are generally restricted to Asian populations, and there are numerous subclades of haplogroup M still present throughout South and East Asia (Metspalu et al., 2004; Yao et al., 2002). This is consistent with the current model of Asian origins, where early modern humans spread rapidly along a southern coastal route from the Near East to Southeast Asia (Kong, et al., 2006). Haplogroups C and D are the most common lineages in Siberian populations (Derenko et al., 2010). Haplogroup D is also a common lineage throughout China, along with Haplogroup G (Yao et al., 2002; Kong et al., 2006). Haplogroup Z is another Asian-specific lineage that shares a common origin with Haplogroup C. Haplogroups Q and E are also M-derived Asian lineages, but these are less common and not well-characterized. The N-derived lineages present in Asian populations branch both directly from haplogroup N (A, O, S, Y) and haplogroup R (B, F, P). Haplogroups O, P, and S are rare lineages found in Oceania (Palanichamy et al., 2004). Haplogroups B, Y, and F are all

characteristic of East Asian populations (Yao et al., 2002), except that Haplogroup B is also characteristic of Pacific Islanders (Schurr and Wallace, 2002). Haplogroup A is found throughout East Asia at low levels, but is present at elevated frequencies in Chukchi and Siberian Eskimo populations (Starikovskaya et al., 1998) and in native North Americans, who possess a subset of four lineages (A, B, C, D) commonly observed in Northern Asia (Torroni et al., 1993; Forster et al., 1996), as well as haplogroup X, as mentioned previously.

Y-chromosome polymorphisms have also been used to study human migration, providing information about human origins from a paternal perspective (Semino et al., 2000; Underhill et al., 2000; Wells et al., 2001). Both short tandem repeats (i.e., microsatellites) and SNPs are used to assign individuals to haplogroups in the same manner used for mtDNA. Many population studies combine analysis of mtDNA and Y-chromosome diversity, since this gives evidence of both maternal and paternal lineages (Malyarchuk et al., 2004; Keyser et al., 2009). Generally speaking, Y-chromosome haplogroups show stronger patterns of geographic localization than mtDNA, possibly because patrilocality (males continue to live in the same locations as their fathers) is characteristic of many human cultures (Burton et al., 1996; Jobling and Tyler-Smith, 2003). One notable exception is a widespread Y-chromosome haplotype found throughout Central Asia that originated in Mongolia approximately 1,000 years ago and is found in descendants of Ghengis Khan (Zerjal et al., 2003).

The single-copy nature of the Y-chromosome makes obtaining data more challenging in situations where DNA degradation has occurred, such as in ancient bones and teeth. STR profiles (Y-chromosome and autosomal loci) from ancient individuals are

often incomplete because DNA sequences cannot be recovered, limiting data interpretation (Ricaud et al., 2005; Keyser et al., 2009). The scope of this study was limited to mtDNA analysis.

Ancient DNA Research

Studying mtDNA diversity in modern populations has provided a means to make inferences about ancient human populations (Willerslev and Cooper, 2004; Merriwether et al., 2005). The ability to extract genetic material from ancient bones, teeth, and other remains (Pääbo, 1988) allow for a direct assessment of the genetic makeup of founding populations and their movements. These results can then be used to make connections between ancient and modern populations for a more thorough discussion of human origins.

However, there are various additional issues that must be taken into consideration when working with ancient DNA (aDNA), all of which combine to make aDNA research technically challenging. First and foremost, the amount of surviving intact DNA is often very low. Most of the DNA found in ancient extracts is also badly fragmented, preventing large polynucleotide fragments from being readily amplified by PCR (Pääbo, 1989; Malmström et al., 2009). This complication means that multiple overlapping primer pairs are commonly utilized in order to reliably retrieve moderate to long target DNA sequences. The ability to recover aDNA is also dependent on tissue type and the method of recovery used. A recent study by Adler et al. (2011) found that pulverizing tooth samples resulted in 5-30-fold higher concentrations of recovered DNA when compared to drilling at the standard speed of 1000 RPM, likely because excess heat generated by

drilling destroys much of the surviving fragments. Low template concentrations make sensitivity to contamination the central issue when attempting to validate aDNA sequences (Cooper and Poinar, 2000). Bacterial contamination is also a major concern (Pääbo et al., 1988). Significant post-mortem damage is also commonly observed in aDNA samples. Oxidative damage results in extensive deamination of cytosine and thymine residues (Pääbo, 1989). Oxidative modifications to nucleotides can lead to replication errors and the formation of chimeric DNA sequences during PCR (Willerslev and Cooper, 2004). Ancient extracts also may contain compounds that inhibit the PCR reaction (Pruvost and Geigl, 2004), although this can be overcome by either increasing Taq polymerase concentrations (Pääbo, 1989) or by using diluted aliquots of the DNA extract (Pruvost and Geigl, 2004; Sampietro et al., 2007). Inhibition can also occur through the formation of cross-links both between DNA molecules and within the double helix (Willerslev and Cooper, 2004). Generally speaking, the level of DNA damage is not due solely to sample age, but also to the environmental conditions of the burial, since rapid post-mortem desiccation can help minimize oxidative degradation (Pääbo, 1989). Samples gathered from moist, tropical regions have been found to have higher rates of DNA damage than samples from more arid, temperate regions (Adler et al., 2011).

The first aDNA samples were recovered in 1984 by Higuchi et al. from museum specimens of quagga (*Equus quagga*), a horse relative extinct since 1883. Other early studies extracted aDNA from animal skin and mummified human remains (Pääbo, 1985; Pääbo, 1989). Ancient DNA studies were published claiming to have retrieved aDNA from extremely old sources, such as dinosaur bones (Woodward et al., 1994) and insect specimens fossilized in amber (Cano et al., 1993). However, the early optimism within

the field was badly damaged following a series of embarrassing incidents where a great deal of early research was discredited due to either contamination or irreproducibility (Willerslev and Cooper, 2004). For instance, the reported recovery of aDNA from dinosaur bones (Woodward et al., 1994) was determined to be contamination from a nuclear insert of human mitochondrial DNA (Zischler et al., 1995).

In 2000 Science published a letter by Alan Cooper and Hendrik Poinar with the contentious title, “Ancient DNA: Do It Right or Not At All.” The letter was written after the pair attended the 5th International Ancient DNA Conference and heard a presenter state that “the field was now mature and could move ahead with confidence.” Considering the amount of flawed research that had occurred in the previous decade, it was surprising that many investigators still were not providing evidence of DNA authenticity. Cooper and Poinar’s criticism was intended to put forth a clear set of criteria for DNA authenticity and call for their widespread adoption.

Cooper and Poinar suggest that aDNA research must take place in an “isolated” workspace, ideally in a separate building from other DNA work. They also advise that duplicate analyses be performed at multiple locations. All PCR reactions must be run in a controlled manner (i.e., duplication and blanks), although they shy away from endorsing the use of positive controls due to the risk of contamination. Since aDNA is commonly fragmented, PCR amplification should generally be more successful for smaller amplicons. All aDNA sequences must be reproduced in multiple extracts and purified by molecular cloning. Extracting DNA from “associated remains” (i.e., animal remains found nearby) could also improve the case for authenticity. Furthermore, aDNA sequences should “make phylogenetic sense,” although they do not clearly define this

idea. Montiel et al. (2001) submit that the requirement of phylogenetic sense could be met by agreement between RFLP and HVI haplogroup markers and by observing similar haplotype variability between the sample and surrounding populations. The extent of oxidative processes, such as deamination, also could be used to support authenticity. They also endorse the quantification of initial copy number, if possible, stating that very low copy numbers ($<1,000$) may make it “impossible to exclude the possibility of sporadic contamination (Cooper and Poinar, 2000).” Other studies have taken additional precautions not discussed by Cooper and Poinar, including publishing the mtDNA sequences of all researchers associated with the samples. Extensive precautions are also taken to ensure reagent sterility, especially those not purchased commercially, including filter-sterilization, autoclaving, and aliquoting (Montiel et al., 2001).

While there is little doubt that aDNA research required better quality control in its early years, debate continues regarding how extensive validation procedures should be. Cooper and Poinar acknowledge the process is “both expensive and time-consuming,” while other authors have described the authentication criteria as “severe” (Pruvost and Geigl, 2004). Barbujani et al. (2004) also comment on the difficulty of recovering and analyzing related animal bones. Bandelt et al. (2005) report that very few researchers are actually following all of the criteria given by Cooper and Poinar. This is likely because fulfilling every possible measure of DNA authentication is not practical for many laboratories, especially considering the limited quantity of DNA present in a successful extraction. Some common-sense precautions, like proper PCR controls, sample isolation, and molecular cloning are already widely used. Others, such as independent replication, are out of reach for all but the largest laboratories. Quality control procedures are still

active topics of conversation, but the high priority it has been given thus far has improved the overall credibility of aDNA research.

Phylogeographic Analysis of Human mtDNA

MtDNA sequence data has been analyzed using a variety of classic population genetics tools, including trees built using Neighbor-Joining and Maximum Likelihood algorithms (Cann et al., 1987, Ingman et al., 2000). However, problems arise when these methods are applied to population-level datasets, primarily because traditional phylogenetic methods were built to compare different species that had already diverged from one another (Posada and Crandall, 2001). When a tree is constructed using data from only one species or population, there may be little differentiation. This can hamper resolution by producing multiple polytomies within the tree and lead to a very large number of equally parsimonious trees (Bandelt et al., 1995). Since phylogenetic trees are bifurcating, parallel mutations and back-mutations cannot be modeled effectively. In fact, parallel mutations can lead to the presence of “ghost links” within the trees (Bandelt et al., 1995). Bifurcating trees also assume that all ancestral nodes are extinct by forcing all sampled genotypes to occupy terminal branches.

Programs capable of producing haplotype networks were developed in order to address these issues and provide a more effective means of analysis for population-level data. Haplotype networks are by far the most common method for examining human interrelationships using mtDNA sequence variation (Richards et al., 1998; Malyarchuk et al., 2002; Derenko et al., 2003; Li et al., 2010). Networks are useful for examining diversity at the species- and population-level, where the level of DNA sequence

divergence is often lower compared to between-species divergence (Bandelt et al., 1999). Furthermore, haplotype networks can be used to visualize reticulations due to parallel mutations by forming loops within the network. Ancestor-descendant relationships can also be clearly described using haplotype networks, where interior nodes represent ancestral lineages, while derived nodes represent descendant lineages (see Figure 2 for an example network; Posada and Crandall, 2001). Interior nodes are also predicted to be the most abundant and have the broadest distribution (Freeland, 2005). Networks also have no issues displaying polytomies; in fact, the most ancestral nodes in a network may have many links branching from them (Fig. 2.).

These properties of haplotype network have their foundation in the principles of coalescent theory. As mentioned previously, coalescent theory is a stochastic process that provides a retrospective view of the time depths at which alleles coalesce to a common ancestor. (Kingman, 1982). In a haplotype network, coalescence can be visualized by tracing inward from the most peripheral nodes to the node at the core of the network representing the MRCA. The point at which two peripheral branches unite represents a coalescent event. The rate at which coalescence occurs is related to effective population size (N_e) and allelic diversity (Fu and Li, 1999). For haploid genes (such as mtDNA), the probability of coalescence (the probability that two alleles chosen at random coalesce in the previous generation) at each generational step is $1/N_{ef}$, where N_{ef} is the effective number of females, and the probability they do not coalesce is $1-1/N_{ef}$ (Freeland, 2005). The probability of remaining distinct after t generations is $(1-1/N_{ef})^t$ (Nordborg, 2000). Under this model, N_{ef} generations are needed, on average, for all lineages to coalesce to a MRCA (Nordborg, 2001; Freeland, 2005). Coalescent theory is a powerful tool for

performing population genetics or phylogeographic research, but the theory rests on several key assumptions (Fu and Li, 1999; Freeland, 2005). The locus of interest must not undergo recombination or be acted on by selection (which also implies random mating). The effective population size must remain constant, which can be a demanding assumption to meet at times (Forster, 1996). Also, the accuracy of coalescent models is higher if N_e is much greater than the sample size (Fu and Li, 1999). This does not imply that a large sample size is needed to perform coalescent simulations, since the probability that a sample of n individuals will contain the MCRA of the entire population is $(n - 1)(n + 1)$ (Nordborg, 2001). For instance, a sample size of only $n = 10$ will have an 81.8% chance of accurately estimating the MCRA.

In this study, we have utilized the network program most widely used in human population genetics, aptly named NETWORK (v4.6) (Bandelt et al., 1995; Bandelt et al., 1999). This program contains a Median-Joining algorithm, which is a modified minimum-spanning algorithm (Kruskal, 1956), which generates a single minimum-spanning tree. In a minimum-spanning algorithm, a distance matrix is used to construct a graph containing all possible links between clusters of unique DNA sequence types. Links are included in the tree in order of their distance with shorter links having higher priority. Links with identical distances are ordered arbitrarily during this prioritization. The shortest link is selected for inclusion in the minimum-spanning tree, followed by the next shortest, and so on. During this process, any links not yet selected that connect two nodes already involved in the minimum-spanning tree are removed from consideration. The MJ algorithm in NETWORK differs from a classic minimum-spanning algorithm in two fundamental ways, leading to the production of a network that includes multiple

minimum-spanning trees. First, the MJ algorithm does not “break ties” by arbitrarily prioritizing links of identical distance. Second, theoretical ancestral nodes referred to as “median vectors” can be inserted into the network in order to achieve minimum distances between nodes.

Once the network has been generated, it may contain many minimum-spanning trees within itself. While the network contains many *possible* explanations for mapping the observed genetic diversity, some trees within the network are more *plausible* than others. This is related to the idea of parsimony, which suggests that the most likely explanation is the one that involves the fewest steps to arrive at a conclusion. It would be incorrect to assume that all paths through the network are equally parsimonious. As a solution, NETWORK contains a Maximum Parsimony (MP) calculation (Polzin and Daneshmand, 2003), which generates a list of shortest trees within the network. The list will contain the minimum number of trees required to account for every link and node present in the full network. In short, applying the MP calculation allows for a determination of the most likely path(s) of evolution by resolving the reticulations present in the full network.

Proper evaluation of DNA sequence data is also important for constructing phylogeographic networks. Not all variation is necessarily informative; some sites may have elevated mutation rates, making them prone to parallel mutation (Malyarchuk et al., 2002), while others may be prone to back-mutation or (in the case of aDNA) post-mortem damage (Gilbert et al., 2003). There is a particularly problematic region in the mtDNA HVSI that contains a poly-cytosine stretch at positions 16180-16193 that is prone to length variation (Fig 1b; Bendall and Sykes, 1995). There are also four positions

(16093, 16189, 16311, and 16362) in the HVSI that are routinely excluded from network analyses. There is an additional position of concern (16296) in members of haplogroup T that is routinely discarded. It is thought that the diagnostic T mutation at position 16294 facilitates mutation at 16296, but the mechanism is unknown (Malyarchuk and Derenko, 1999). When all of the mutations listed above were located within the secondary structures identified by Pereira et al. (2008), three out of five were either in loop regions of identified structures (16093) or in regions where no structures were detected (16189, including the entire poly-C stretch, and 16296). Somewhat surprisingly, the two remaining unstable positions (16311 and 16362) are located in stem regions of small hairpins.

Including uninformative sites in network calculations can result in the improper subdivision of certain lineages. This will lead to the formation of excessive reticulations and median vectors within the network that hinder resolution of relevant links, even if the MP calculation is applied. Removing these uninformative positions from consideration will improve network resolution and clarify evolutionarily relevant lineages within the network (Pike, 2006). Some large networks also can contain very high numbers of reticulations due to large numbers of different DNA sequence types. Since many of the reticulations in large datasets are caused by unique, unusual DNA polymorphisms, NETWORK contains an option where all unique lineages can be excluded from the network. Since major lineages should be present more than once in a large dataset, excluding unique lineages should not compromise overall network structure. This also provides a measure of quality control, since this will limit the influence of spurious data points that may exist in the dataset.

Just as each haplogroup has a distinct global distribution, various sublineages also are also often confined to specific regions. There are times when it may be more useful to look at the distributions of these sublineages as opposed to the entire haplogroup when investigating human migrations (Loogväli et al., 2004). In this thesis, haplogroup networks were used to identify mtDNA sublineages within overall haplogroup diversity and evaluate their geographic distributions. Networks were also used to visualize the ancestor-descendant relationships existing between lineages in our dataset.

Thesis Purpose Statement

While there are only scattered studies that have recovered DNA from Mesolithic individuals (Bramanti et al., 2009; Krause et al., 2010), Neolithic remains from multiple sites across Eurasia (Ricaud et al., 2005; Mooder et al., 2006; Sampietro et al., 2007; Bramanti et al., 2009; Haak et al., 2010; Nikitin et al., 2010) have yielded a great deal of mtDNA sequence data. The Neolithic Revolution in Europe began approximately 9,000 years ago (Thorpe, 1996; Bellwood, 2005) with the emergence of farming and animal domestication and the gradual disappearance of hunter-gatherers. Studies utilizing both modern and ancient mtDNA (Richards et al., 2000; Haak et al., 2005; Sampietro et al., 2007) support a demic diffusion model, where a long-term, low-intensity immigration from the Near East brought farming culture into the Europe without replacing the indigenous population.

In this study, we analyzed the DNA sequence of the HVSI in the mtDNA control region and a portion of the coding region in 14 sets of Neolithic human remains from the North Pontic Region (NPR) in Eastern Europe (modern-day Ukraine). These individuals

are members of the Dnieper-Donetz culture complex (DD) present in the area north of the Black Sea during the Middle Neolithic (Telegin, 1987). We also analyzed the mtDNA HVSI region in three individuals belonging to the Kurgan culture complex present in southern Ukraine during the Bronze Age. Following mtDNA extraction and authentication using molecular cloning, Ukrainian mtDNA sequences were compared to a global pool of modern and ancient samples using phylogeographic network analysis in order to evaluate the relationships between samples in our dataset and understand how our samples are linked to modern and ancient individuals in the geographic area and beyond.

MATERIALS AND METHODS

Sample Collection, Extraction and Amplification

Neolithic bone samples were collected during archeological excavations in south-central Ukraine at three Neolithic cemetery sites along the Dnieper River, north of the Black Sea (Figure 3). The Nikolskoye and Yasinkovatka cemeteries are located in the Podnieprovie steppe region and are relatively close to each other, albeit on opposite sides of the Dnieper River (Fig. 3; Telegin et al., 2002). The Dereivka site is located to the northwest, upstream of Nikolskoye and Yasinkovatka (Fig. 3). All three Neolithic sites are characteristic of the “Mariupol-type” pit graves in this region attributed to the DD culture complex (Telegin, 1987). Radiocarbon dating was performed in prior studies (Telegin and Potekhina, 1987; Hedges et al., 1995; Lillie and Richards, 2000; Telegin et al., 2002; Lillie et al., 2009; Nikitin, unpublished data) (Table 3). While most sample dates appear to be fairly accurate, one individual from Nikolskoye (Ni58; Table 3) has been assigned a much younger age (Lillie and Richards, 2000) than any other individual in our dataset. It is likely that this date is erroneous, especially considering Ni58 was found near the bottom of the burial pit at Nikolskoye, buried underneath other individuals with much older dates (Nikitin, pers. comm.).

We also analyzed three Bronze Age individuals from southern Ukraine (Odessa province) belonging to the Kurgan culture complex (Ivanova et al., 2005) (Table 3). Specimen L8 was assigned to the Yamna (Pit Grave) archeological culture, D1.8 to the Catacomb culture, and L15 to the Mnogovalikovaya (KMK), or Babino, culture. These specimens are not meant to describe the overall gene pool of Ukrainian Kurgans as was recently done for Siberian Kurgans (Keyser et al., 2009). These samples were included in

order to evaluate whether any population genetic continuity existed between Neolithic and Bronze Age NPR inhabitants.

Throughout all steps of aDNA analysis, multiple precautions were taken to prevent sample contamination. All aDNA extractions were carried out under a laminar hood in sterile conditions. All researchers directly involved with lab work, as well as the person who performed the anthropological evaluation of the specimens prior to DNA extraction, had their HVSI regions analyzed as a means of detecting possible contamination (Table 1). All bone samples were thoroughly cleaned and irradiated under UV light overnight to remove surface contamination from post-mortem sources (i.e., bacteria, rodents, etc.). After cleaning, ~1g bone samples were removed using a sterilized Dremel tool. Ancient DNA was extracted into 50µL of sterile water using a modified Qiagen protocol for extracting DNA from bone. The first hypervariable region (HVSI) of the mtDNA control region was amplified using four overlapping primer pairs (Table 2) with negative controls accompanying each set of reactions. Positive controls were prepared in isolation from negative controls and ancient DNA extracts and were only used to confirm effective PCR reaction chemistry. Template volumes of 0.5 and 1.0 µl were used to carry out Fast-Cycling PCR (Qiagen) following PCR conditions essentially as outlined in the Fast-Cycling Kit protocol, except that the PCR cycles were kept at 40 rounds. The reduced template volume of 0.5 µl was used to overcome potential PCR inhibition by salts and other compounds in the bone extract. Successful amplifications were cleaned up using a QIAquick PCR Purification Kit and eluted into 50 µl sterile H₂O. In cases of weak PCR amplification, DNA was concentrated using the Qiagen MinElute kit.

Molecular Cloning

Bone extractions containing aDNA were purified by molecular cloning. PCR products were ligated into Qiagen pDrive vectors using the Qiagen PCR Cloning Kit. Transformed cells were plated onto sterile LB-Amp plates and allowed to incubate overnight at 37°C. Successful clones were identified the next day using blue-white selection and subcultured overnight at 37°C. Subcultured clones were used to inoculate 250 µl DI H₂O using a sterile loop. Hot-Start colony PCR was performed using 1 µl resuspended cells as the template. Following an initial cycle of 10 minutes at 94°C in order to lyse the cells and release mtDNA, we performed 30 PCR cycles under the following conditions: 94°C for 30 seconds, 42°C for 45 seconds, and 72°C for 90 seconds, ending with a final elongation step of 5 minutes at 72°C. SP6 and T7 universal primers were used to amplify the plasmid insertion site. Following amplification, PCR products were run on 2.5% agarose using 100 bp and Low Molecular Weight DNA Ladders (New England Biolabs) to confirm the presence of the aDNA insert.

Typing of Coding Region Polymorphisms

In some cases, the HVSI did not contain sufficient polymorphisms to confidently assign an individual to a given lineage. Restriction fragment length polymorphisms (RFLPs) were originally used to identify mtDNA polymorphism and determine haplogroup status and distribution prior to the widespread use of mtDNA sequence data (Brown, 1980; Torroni et al., 1994). With this in mind, we amplified sites within the mtDNA coding region to assay any homoplastic individuals for diagnostic RFLP polymorphisms, using the hierarchical typing scheme and primer pairs developed by Santos, et al. (2004). RFLP digestion was carried out in a total reaction volume of 10 µl

(8 µl PCR product, 1 µl enzyme, 1 µl buffer) at 37°C for 2-4 hours. All enzymes and compatible buffers were purchased from New England Biolabs. Positive and negative controls were also used to confirm reaction purity and enzyme activity. Following digestion, samples were run on 3.5% agarose to determine RFLP status of each individual.

MtDNA Sequencing

DNA sequence analysis for all samples was performed at the U-M Sequencing Core at the University of Michigan and the Annis Water Resources Institute at Grand Valley State University. Sequencing reactions were carried out on 96-well plates using a BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Samples were cleaned up prior to sequencing using a standard Sephadex protocol to remove proteins, primers, and unincorporated dNTPs. Samples were run using an ABI 3130xl Genetic Analyzer with a 50-cm capillary array. Any ambiguous nucleotide positions remaining in the DNA chromatograms after basecalling were manually assigned using ABI Sequence Scanner (v1.0).

Network analysis

Obtained HVSI DNA sequences were compared to modern and ancient populations by developing phylogeographic networks. We utilized the Median-Joining algorithm (Bandelt et al., 1999) within the program NETWORK (www.fluxus-engineering.com). Resulting networks were simplified to trees using the Maximum Parsimony option (Polzin and Daneshmand, 2003). Unstable nucleotide positions within the control region and length variation within the poly-cytosine region were not

considered for analysis (Fig. 1b; Malyarchuk et al., 2002; Gilbert et al., 2003). During initial phases of network construction, excessive numbers of reticulations and median vectors were observed that could not be resolved effectively using MP. Since most of the reticulations in these initial network constructions were the result of unique mtDNA sequence types, we initially decided to utilize the option within NETWORK that removes unique DNA sequence variants prior to calculation. This did resolve the vast majority of reticulations and produced much “cleaner” networks. However, this option had two major drawbacks: 1) many individuals with unique DNA polymorphisms that were not involved in reticulations would be removed and 2) multiple Neolithic and Bronze Age study subjects would be eliminated, since most have unique mtDNA sequence types. Furthermore, none of our ancient samples had HVSI mutations involved with network reticulations. To resolve this issue, we chose a moderate approach where samples were removed from the dataset only if they were unique *and* the cause of a reticulation. This simplified the network to the extent that the MP calculation was able to resolve any remaining reticulations between well-represented nodes without major decreases in network resolution.

RESULTS

MtDNA HVSI Sequences in Neolithic NPR Inhabitants

We obtained mtDNA HVSI sequences for 14 individuals from three collective burials belonging to the DD culture from the Middle Neolithic (Table 3). All differences from the revised Cambridge Reference Sequence (rCRS; Anderson et al., 1981; Andrews et al., 1999) for the HVSI region (nucleotide positions 16024-16383) are listed in Table 4. All differences listed are nucleotide transitions (Table 4). Of the three cemeteries, Yasinkovatka is the best-represented (9 individuals), followed by Nikolskoye (3 individuals) and Dereivka (2 individuals).

We were able to assign 12 of 14 Neolithic individuals to mtDNA haplogroups based on HVSI variations and RFLP data (when necessary). Nine of these individuals belonged to haplogroups characteristic of European and Near Eastern populations (Table 3). DD33 (HVSI motif 16311), DD38 (16080), Ya57 (16241), and Ya64 (16064-16240) were members of haplogroup H. Ya32, Ya54, and Ni79 belonged to haplogroup T and shared the HVSI motif 16294-16296. Ya19 (16343) belonged to haplogroup U3, and Ni94 (16256-16270-16356) belonged to haplogroup U5a. Interestingly, Ni58 and Ya45 carried the HVSI motif characteristic of the root sequence that defines haplogroup C, 16223-16298-16327, although Ni58 possessed an additional mutation at position 16061. A third individual, Ya34 (16223-16298-16327-16357), belonged to the C4a sublineage. Haplogroup C is one of the most common lineages in East Eurasian, particularly Siberian, populations but is very rare in modern Europe.

DD38, Ya32, Ya54, and Ni79 were assigned to haplogroups even though we were unable to obtain DNA sequence data for their entire HVSI. We could not amplify the 145 bp fragment of HVSI (Table 2) in Ya32, Ya54, or Ni79 (shared HVSI motif 16294-

16296), but sufficient mtDNA sequence data was collected from other fragments to successfully type them to haplogroup T. We were only able to recover data from one HVSI fragment in sample DD38 which contained only one mutation at position 16080 (142 bp; Table 2). We were able to assign this sample to haplogroup H by confirming the absence of the AluI restriction enzyme site at position 7025 (Table 3).

Two individuals (Ya17 and Ya36) could not be definitively assigned to a haplogroup due ambiguous HVSI motifs (16241 and 16320, respectively), and we could not successfully amplify DNA to check the RFLP status of these samples. It is quite possible that Ya17 is a member of haplogroup H, considering that another sample from the same cemetery (Ya57; Table 3) has the same HVSI motif as Ya17 (16241) and was successfully typed to haplogroup H. Ya36 possessed only one difference from the rCRS at position 16320, but it is not a diagnostic position for any particular haplogroup (van Oven and Kayser, 2009). Ya36 may also be a member of haplogroup H due to its overall lack of differentiation, but this is only a tentative assignment in the absence of RFLP data.

MtDNA HVSI Sequences in Bronze Age Kurgans

In addition to the 14 Neolithic specimens mentioned above, we also recovered mtDNA HVSI sequences from three individuals from Bronze Age Kurgan burial mounds in southern Ukraine in order to examine whether any genetic continuity existed between Neolithic and Bronze Age NPR populations. All differences to the rCRS are nucleotide transitions, except for an A-T transversion at position 16305 in sample L8 (Table 4). Samples D1.8, L8, and L15 all possess highly derived HVSI motifs (D1.8: 16223-16278-

16298-16327-16357; L8: 16218-16223-16288-16298-16305AT-16327-16357; L15: 16218-16223-16298-16327-16357; Table 4). All three samples were identified as members of haplogroup C4a, similar to Ya34 (16223-16298-16327-16357). D1.8 belonged to haplogroup C4a3, as defined by 16278. L8 and L15 belonged to a branch of C4a not previously observed in any of the modern and ancient humans, which we have labeled C4a6 (Table 4).

Network for Haplogroup U5a

Sample Ni94 (16256-16270-16356) is a member of haplogroup U5a, which is one of the oldest European mitochondrial lineages (Richards et al., 2000; Malyarchuk et al., 2010). Although members of haplogroup U5 are present at low frequencies in modern Europeans, it is present at higher frequencies in aDNA studies from Central Europe (Haak et al., 2005; Bramanti et al., 2009). There are disagreements in the literature concerning which mutations should be used to describe haplogroup U5 and its subgroups (Richards et al., 2000; cf. Bramanti et al., 2009, van Oven and Kayser, 2009), which can complicate lineage assignments and phylogeographic analysis. In order to determine the placement of the U5a individual in our dataset (Ni94; Table 3) within U5a as a whole, we constructed a phylogeographic network of U5a diversity from across Eurasia (Figure 4). We used mtDNA sequences from modern and ancient European, Near Eastern, and Asian populations to construct the tree (Figure 4) and followed the classic mutation scheme for delineating U5 lineages, where a transition at position 16270 defines U5 in general and transitions at 16192 and 16189 define U5a and U5b, respectively (as described in Richards et al., 2000). The derived lineage U5a1 has an additional mutation at position

16256. An additional derivation of U5a1, subgroup U5a1a, has a back mutation at position 16192. As expected, Ni94 appears in the U5a tree one step away from the U5a1a cluster (Fig. 4).

Network for Haplogroup C

Ni58, Ya45, and Ya34 are members of haplogroup C, which is one of the most common mtDNA lineages in East Eurasia and likely originated in South Siberia, although it is virtually absent in other locations (Derenko et al., 2007; Derenko et al., 2010), except North America, where one sublineage of C is present in Native American populations (Torroni et al., 1993). Ya45 had the consensus HVSI motif for haplogroup C (16223-16298-16327), while Ya34 had an additional mutation at position 16357, which identified this sample as a member of haplogroup C4a. Ni58 has a mutation at position 16061 in addition to the C root motif.

All three Bronze Age samples (D1.8, L8, and L15) also belong to haplogroup C, bringing the total number of haplogroup C individuals in the present study to six. These samples share the HVSI motif 16223-16298-16327-16357, placing them in haplogroup C4a with Ya34, as mentioned previously. D1.8 has an additional mutation at position 16278 which is diagnostic for the C4a3 sublineage (Table 4). L8 and L15 share a mutation at position 16218, but L8 has two more mutations at positions 16288 and 16305 (Table 4).

In order to understand how the haplogroup C individuals observed in this study relate to overall group diversity, we constructed a phylogeographic network for C (Figure 5). This network draws from the same pool of sources used to construct the network of

U5a, but also includes haplogroup C individuals from an additional subset of articles from Europe and Asia in order to more fully characterize haplogroup C variation (Fig. 5). Papers included in the haplogroup C network but not the U5a network either 1) do not include U5a mtDNA sequences or 2) cover geographic regions already well-represented by other articles in the U5a tree. Two aDNA articles from northwestern China (Zhang et al., 2010) and south Siberia (Mooder et al., 2006) were also not included because they only analyzed a subset of the HVSI. They are still useful for lineage comparisons, but would have reduced network resolution.

The completed network confirms Ya45 possesses the 16223-16298-16327 root motif for haplogroup C and is located at the core of the network (Fig. 5). Ni58 appears one mutational step from the ancestral node, but appears to be a unique lineage with no descendant nodes. Ya34 is also one step away from the C root motif and is a member of the C4a subgroup. This sample represents an ancestral node for a large portion of the network defined by a mutation at position 16357 (Fig. 5). Van Oven and Kayser (2009) have named this lineage C4a2'3'4', since the Ya34 motif is ancestral for the C4a2, C4a3, and C4a4 sublineages.

Interestingly, the HVSI motifs present in the three Kurgan individuals appear to represent unique branches within the haplogroup C network. D1.8, L8, and L15 all branch directly from the ancestral node defined by Ya34, although D1.8 occupies a separate terminal branch from L8 and L15 (Fig. 5). This “L branch” is defined by the mutation at position 16218. L15 is separated from Ya34 by this mutation alone, whereas L8 occupies a terminal node due to its additional HVSI mutations mentioned previously

(Table 4; Fig. 5). We have labeled this branch “C4a6,” since it has not been previously observed in other mtDNA studies of modern and ancient humans.

DISCUSSION

It is generally accepted that the modern European gene pool is composed of lineages that arrived in the region from the Near East during the Paleolithic and repopulated the continent following the LGM, along with another, smaller influx at the onset of the Neolithic, which spread the use of agriculture via demic diffusion without replacing the indigenous population (Sokal et al., 1991; Cavalli-Sforza and Minich, 1997; Richards et al., 2000; Pinhasi and von Cramon-Taubadel, 2009). Overall, our results suggest that the gene pool of the Neolithic DD culture from the NPR contained genetic contributions from both West and East Eurasian populations. We detected multiple specimens belonging to Haplogroup C, which is a lineage characteristic of East Eurasian populations that is thought to have originated in South Siberia in the area surrounding Lake Baikal (Starikovskaya et al., 2005; Derenko et al., 2010). Archeological and anthropological evidence documenting cultural and physical characteristics of members of the DD culture during the Neolithic also can be brought to bear to reinforce the genetic data. Additional evidence from Bronze Age sites and modern populations show the retention of these Asian lineages following the Neolithic in the NPR. Taken together, these results have significant implications with respect to the current model of human migrations into Europe at the onset of the Neolithic which assume a predominantly Near Eastern source population.

West Eurasian Lineages in Neolithic NPR

Haplogroup H

Of the four haplogroups found in our sample of the Neolithic NPR population, three (H, U, and T) are generally considered West Eurasian lineages and are commonly found in European populations. Haplogroup H is the most abundant lineage in our sample (4 - possibly up to 6 - of 14 individuals; Table 3). Haplogroup H is the most common lineage throughout Europe, occurring in approximately 46% of the population, although it likely originated in the Near East (Richards et al., 2000). The rCRS is also located within this group (Andrews et al., 1999). Extensive study of haplogroup H has revealed a high level of diversity within this group (Loogväli et al., 2004).

Genbank was analyzed using BLAST searches (<http://blast.ncbi.nlm.nih.gov>; Altschul et al., 1990) for similar HVSI mtDNA sequences belonging to haplogroup H and look for any geographic tendencies that may be present in these lineages. BLAST algorithms operate by breaking the query DNA sequence into smaller fragments prior to searching Genbank. Once these fragments are found, the algorithm builds alignments while extending the length of each fragment, scoring alignments higher or lower based upon DNA sequence similarity, provided the score remains above a predetermined threshold. This algorithm operates more rapidly than if the initial search is conducted using the entire query DNA sequence. The Megablast search algorithm was used, since it is optimized to search for target sequences identical or very similar to the query DNA sequence.

Some samples could be connected with other individuals in Europe and the Near East, while others had ambiguous relationships. Sample DD33 has one difference from

the rCRS at position 16311 (Table 3). However, mutations at this position are considered unstable and uninformative for the purposes of phylogeographic analysis (Malyarchuk et al., 2002; Gilbert et al., 2003). Phylogenetic trees produced from whole-genome mtDNA sequences belonging to haplogroup H support this contention since 16311 appears in multiple separate sublineages throughout haplogroup H (van Oven and Kayser, 2009). Disregarding position 16311 equates the HVSI of DD33 with the rCRS motif, which is widespread across Europe and the Near East. Despite this ambiguity, it is worth mentioning that the 16311 mutation is present in two individuals from a Neolithic cemetery near Lake Baikal (Mooder et al., 2006), although the authors did not assign these samples to a haplogroup, in part because the DNA sequence for the entire HVSI was not available, leaving open the possibility that additional mutations were present. Likewise, no exact matches of the Ya64 HVSI (16064-16240) were found in Genbank. However, four individuals were detected possessing a mutation at position 16240 alone. Three of these individuals are from Western Europe: one each from the Netherlands (Accession No. HM100712), Norway (Helgason et al., 2001; AY025964), and Portugal (Pereira et al., 2007; EF177436). The fourth 16240 match is a Druze from Northern Israel (Shlush et al., 2008; EU600333), which is intriguing since Druze populations have an unusual mtDNA pool that may reflect prior genetic diversity in the region due to their religious and social isolation (Shlush et al., 2008). No samples were found that contained only the 16064 mutation. DD38 has an HVSI motif with one difference from the rCRS at position 16080 which appears infrequently in haplogroup H and appears to cluster geographically. Other samples matching the DD38 motif appear in populations surrounding the North Sea, including Denmark (AY847787), the Shetland and Orkney

Islands (Goodacre et al., 2005; AY951224, AY951217, AY951093, AY951080, AY950910, AY950876, AY950875, AY950831.1), and Iceland (Helgason et al., 2003; AY314252; AY314456; AY314516). How these individuals relate to DD38, if at all, is unclear. Unfortunately, we were only able to amplify the first segment of the HVSI in DD38 (presumably due to DNA degradation), so it is unknown whether additional polymorphisms are present that would influence our search results.

Samples Ya17 and Ya57 both have identical HVSI motifs (a single mutation at 16241). This lineage was only detected in a single Bosnian individual (AY005635), suggesting it may be scarce in modern populations. Finally, Ya36 has a single mutation at position 16320. All matches in Genbank are all members of haplogroup H, supporting the tentative assignment of Ya36 to this group. The Ya36 motif is found in five Druze individuals one from Shlush et al., 2008 (EU600338), and another four mtDNA sequences directly submitted to Genbank (EU566049, EU566059, EU566080, EU566091). Ya36 matches were also found in one Armenian (HM775971) and one Greek (Irwin et al., 2008; DQ418146). To summarize, although geographic affinities could not be determined or had unclear implications for some haplogroup H individuals, others were only found in individuals localized in proximity to the greater Black Sea region in locations such as the Balkans, Caucasus, and Near East. This suggests that Near Eastern populations were influential in shaping haplogroup H diversity in the NPR during the Neolithic.

Haplogroup T

Three of 14 Neolithic NPR specimens were members of haplogroup T and all shared the same lineage (16294-16296). Haplogroup T is a fairly recent arrival into Europe from the Near East with some sublineages entering as late as the Neolithic (Richards et al., 1998; Richards et al., 2000). Normally haplogroup T is defined by mutations at positions 16126 and 16294, but we were unable to amplify the HVSI segment containing position 16126 (Table 2). Overlapping sequencing of the primer sites from this segment did not reveal any mutations that would hinder primer annealing during PCR in any of the three samples. The formation of primer dimers or hairpins does not seem likely to be the cause, as this lack of amplification was not a systemic problem in all samples.

The 145 bp fragment in question begins in the middle of the large secondary structure detected by Pereira et al. (2008) and contains another 37 bp hairpin that immediately follows the 93 bp structure. The TAS element that is influential for H-strand arrest and D-loop formation is also located in this segment (Roberti et al., 1998). The presence of these structures did not prevent amplification in other samples or hinder amplification of the 142 bp fragment (Table 2) which also includes a large portion of the 93 bp structure. However, if a mutation were to occur that stabilized these structures, it might prevent elongation during PCR by physically blocking *Taq* polymerase.

Additionally, there is widespread doubt regarding the usefulness of mutations at position 16296 within haplogroup T (Richards et al., 2000; Pike, 2006; Malyarchuk and Derenko, 1999). These studies have recommended removing 16296 from consideration in phylogeographic analyses, speculating that it may be destabilized by the C-T transition at

16294 that partially defines haplogroup T, although the mechanism is unknown (Richards et al., 2000). This information suggests that our samples are effectively at the root ancestral node of haplogroup T, or at least close to it.

Haplogroup U

Samples Ya19 and Ni94 belong to haplogroup U, albeit to different sublineages. Ya19 belongs to subhaplogroup U3, which occurs most often in the Near East (Abu-Amero et al., 2008) and Caucasus Mountains (Behar et al., 2008). U3 is present in Europe only in low frequencies but is more visible in Balkan populations (Bosch et al., 2005). U3 has been found in very high frequencies in Polish Roma populations, likely due to a strong founder effect (Malyarchuk et al., 2006a). This lineage is also abundant (39%) in the Southern Levant (González et al., 2008). Ni94 is a member of haplogroup U5a, and our network for U5a (Fig. 4) indicates Ni94 belongs to the highly derived lineage U5a1a, which has the consensus motif 16256-16270. The additional mutation in Ni94 at position 16356 is unusual, since this is a diagnostic mutation for a subhaplogroup U4. However, other samples have been identified in the literature that possess mutations diagnostic to both U4 and U5, and in all cases the samples have been assigned to U5 (Álvarez-Iglesias et al., 2009; Pereira et al., 2000).

U5a is one of the oldest haplogroups and is present throughout Europe, the Near East (Abu-Amero et al., 2008; Metspalu et al., 2004) and Siberia (Derenko et al., 2007). The phylogeographic genetic network for U5a presented here (Fig. 4) confirms the accepted European origin for U5a (Richards et al., 2000; Malyarchuk et al., 2010). The majority of samples are European, although substantial minorities of Near Eastern and

Asian individuals are present. Further support for a European origin of U5a is the fact that all Near Eastern and Asian samples are derived from HVSI motifs present in Europe. There is only one European sample (from the Açores Islands, Brehm et al., 2003) that has derived from a node that is exclusively Asian or Near Eastern in our dataset, possessing the HVSI motif 16086-16256-16270-16342 (Fig. 4) and branching from a node represented by a single individual from Central Iran (16256-16270-16342) (Metspalu et al., 2004). This would seem to suggest that, following the initial expansion of U5a from Europe, the Açorean haplotype originated in the Near East before eventually back-migrating to Europe. It is difficult to pinpoint how recently the Açorean mtDNA lineage returned to Europe from the Near East, but it would have arrived in the Açores Islands no earlier than the settlement of the archipelago by the Portuguese in the late 15th Century (Brehm, 2003).

Admixture of East Eurasian mtDNA in West Eurasian Populations

In this study, we have detected East Eurasian mtDNA lineages in Neolithic and Bronze Age NPR inhabitants. Members of haplogroup C have been recovered in multiple aDNA studies (Ricaud et al., 2005; Mooder et al., 2006; Keyser et al., 2009; Li et al., 2010) and, to our knowledge, both Ya34 and Ya45 are the oldest haplogroup C individuals ever observed outside of Siberia (Table 3). Ya34 and Ni58 appear to have mtDNA sequence types not previously observed in extensive studies of modern Siberian populations (Starikovskaya et al., 1998; Derenko et al., 2003; Starikovskaya et al., 2005; Derenko et al., 2007). Ya34 is especially noteworthy, as it appears to possess an ancestral HVSI motif within the portion of the C4a subhaplogroup defined by 16357 (Fig. 5).

Three specimens from Bronze Age Kurgan burials were also typed to haplogroup C4a. Our phylogeographic network of haplogroup C (Fig. 5) illustrates that these Kurgan samples are directly derived from the root motif of C possessed by the Ya34 individual, indicating that the diversification of the C haplogroup could have happened in the Black Sea region *in situ*. This hypothesis is supported by the fact that the C-bearing L8 and L15 individuals share a unique mutation at position 16218, making this a diagnostic polymorphism for Bronze Age C-bearing NPR populations. The presence of haplogroup C in Bronze Age Kurgans and the Neolithic DD culture is evidence of population genetic continuity in the NPR through the maternal line.

Asian lineages have also been detected in other aDNA studies of European populations. In their 2007 article discussing the evolution of lactase persistence in Europe, Burger et al. reported a partial HVSI profile from an individual belonging to the Körös culture inhabiting the eastern Hungarian plain during the Neolithic. The individual in question appears to be a member of haplogroup N9a, a deep-rooted subclade of N that is characteristic of Southeast Asian populations (Yao et al., 2002; Kong et al., 2006).

Numerous studies of modern European mtDNA diversity have also detected East Eurasian lineages in European populations, especially in and around the Carpathian Basin. A recent study of mtDNA haplogroup diversity in Carpathian highlanders (Nikitin et al., 2009) detected several individuals in Lemko populations that are members of haplogroup C, all containing the HVSI motif 16051-16223-16298-16327 (Nikitin, unpublished data), which was also detected in a Kalmyk individual (Derenko et al., 2007). Kalmyks currently live along the northwestern coast of the Caspian Sea in southern Russia, following a migration from western Mongolia approximately 300 years

ago, but remain genetically similar to other South Siberian populations (Nasidze et al., 2005; Derenko et al., 2007). One Lemko individual also belongs to haplogroup Y, which is especially intriguing because haplogroup Y is another Asian mtDNA lineage present in Siberia (Derenko et al., 2007; Yao et al., 2002) that descended from haplogroup N9 (van Oven and Kayser, 2009). Another recent study documenting mtDNA diversity in modern Hungarian populations (Nádasi et al., 2007) found several individuals that belong to haplogroup M. Unfortunately, this study only utilized RFLP data and did not specify the lineage of these samples further.

The C5 subgroup (HVSI motif 16223-16288-16298-16327) has a distinct presence in Europe. In fact, it contains a haplogroup C lineage unique to Europe, which possesses a derived mtDNA sequence type with mutations at positions 16223, 16234, 16288, 16298, and 16327. It is geographically restricted to northern Poland (Malyarchuk et al., 2002; Grzybowski et al., 2007) and northeastern Germany (Poetsch et al., 2003; Poetsch et al., 2004). This derived subcluster extends the presence of haplogroup C in Europe from the Carpathian Basin north to the Baltic coast. One individual belonging to the same European-specific lineage (except with two additional mutations) was reported in a study of Romanian Aromuns (Bosch et al., 2005) suggesting this subcluster has a persistent presence within Europe. Other examples of haplogroup C5 in Europe include another individual from Poland lacking the 16234 mutation (Malyarchuk et al., 2002) and one individual from Northern Greece with the HVSI motif 16223-16261-16288-16298 (Irwin et al., 2008). An additional member of C is located in Greece (Bosch et al., 2005) but belongs to an entirely different lineage.

The C1 subgroup is present in three different studies from Germany, Iceland, and the Canary Islands, two of which (Maya-Mayer et al., 2001; Ebenesersdóttir et al., 2011) likely reflect recent gene flow from the Americas (Torroni et al., 1993). One German from Lower Saxony belonging to C1 was detected by Pfeiffer et al. (2001), but no other C1 individuals were observed in Europe. A major study of Siberian haplotype diversity (Derenko et al., 2007) only detected C1 in 4 out of 1,432 individuals tested, which seems to also limit the possibility of a Siberian origin for the German C1 sample, leaving recent gene flow (possibly from Iceland) as the most likely explanation for the presence of this lineage in northwest Europe.

In some cases the geographic origins of certain individuals are unclear. An analysis of haplogroup diversity in European Russians detected two haplogroup C individuals, but the study (Orekhov et al., 1999) does not specify which location in the study (Kostroma, Ryazan, or Kursk) the individuals in question originate from. The only other study to report haplogroup C in European Russia (aside from Kalmyks) was a study of Druze and Adygei mtDNA diversity (Macaulay et al., 1999). Three Adygei in this study were typed to haplogroup C, which is intriguing due to their proximity to our study site. However, it is difficult to determine whether haplogroup C in Adygei is representative of their genetic history or because of their close proximity to Kalmykia. Another difficult haplogroup C sample that cannot be precisely assigned to a geographic location was also detected in mainland Turkey (Calafell et al., 1996).

There seem to be multiple locations throughout Europe that harbor Asian lineages, as summarized in Figure 6. The first is the greater Black Sea region, which includes our Neolithic study sites in Ukraine, the Balkans, and the adjacent Carpathian

Basin. The Baltic coast of Europe and Poland also contains a unique C lineage, which may have expanded north from the Black Sea. Also, Haplogroup C1 has been detected in Europe, but these individuals are most likely examples of gene flow from North America.

Non-Genetic Indicators of East Eurasian Interactions

The various characteristics of prehistoric burials are often diagnostic for defining to which culture the burial belongs. When DD burial sites are compared with similar sites representing other Neolithic cultures, they are found to differ from other sites elsewhere in Europe. Importantly, those differing features often align with characteristics of Neolithic Siberian burials. For instance, individuals belonging to the Neolithic LBK culture of Central Europe (one of the largest cultural groups in Neolithic Europe) are often found buried in the fetal position (Haak et al., 2008; Bramanti et al., 2009). Alternatively, Neolithic Siberian burials most commonly feature individuals buried in an extended supine position with arms to the side (Mooder et al., 2006), indicating the body was tightly wrapped. Heavy use of red ochre is also recorded at Siberian sites (Mooder et al., 2005). DD individuals were found buried in the extended supine position, covered in red ochre, and had likely been tightly wrapped (Telegin, 1987), suggesting a cultural affinity with Neolithic Siberians, at least with respect to burial traditions.

The discovery of Siberian mtDNA lineages in Neolithic Ukraine gives insight regarding an interesting study of Neolithic Ukrainian tooth and jaw measurements by Jacobs (1994). After extensive analysis of Mesolithic and Neolithic specimens recovered from multiple burial sites throughout Ukraine, Jacobs concluded that both tooth and jaw dimensions had increased in Ukrainian populations from the Mesolithic to the early

Neolithic before decreasing somewhat by the late Neolithic. This effect is also more pronounced in females than in males (Jacobs, 1994). These findings run counter to observations made in several other Neolithic populations, where it is hypothesized that the observed reduction in dento-gnathic parameters is due to a shift in subsistence style from a hunter-gatherer mode to the agricultural approach that defines the Neolithic (Calcagno and Gibson, 1988). Jacobs concluded that “low-intensity gene flow” (i.e., demic diffusion) from outside the region was responsible for the observed increase in tooth and jaw size after ruling out other possible causes, such as sampling effects, increases in overall body size, and local selection effects. In his concluding remarks, Jacobs hypothesizes that the source of the gene flow was in the Near East, specifically at a Neolithic site in Western Iran.

Based on our recovery of East Eurasian mtDNA lineages in members of the DD culture, we propose that the gene flow observed by Jacobs may have been derived from multiple sources. Considering that several of the Neolithic Ukrainians in this study show affinities with Near Eastern populations based on their mtDNA sequence variation, gene flow from Western Iran is still quite possible. However, the presence of haplogroup C in Neolithic Ukraine indicates that there is likely a Siberian component to the observed gene flow as well.

Admixture of West Eurasian mtDNA in East Eurasian Populations

Just as Asian mtDNA haplogroups are found scattered throughout European populations, there is corresponding evidence of European haplogroups in Asia. Cemeteries dating to the Early and Middle Neolithic in the Lake Baikal region contain

three individuals belonging to haplogroup U5a (Mooder et al., 2006). The presence of U5 Neolithic Siberia may be related to its early estimated time of divergence (25-30 kya; Malyarchuk et al., 2010). There is also evidence of an influx of European haplotypes during the Bronze and Iron Ages that followed the Neolithic. Bronze and Iron Age Kurgan populations in the vicinity of Lake Baikal are composed of primarily West Eurasian lineages, such as haplogroups U, T, H, and I (Lalueza-Fox et al., 2004; Keyser et al., 2009). In particular, the west-to-east migration of Kurgans can be evidenced by the presence of haplogroup T4 in Bronze Age Altaian Kurgans (Keyser et al., 2009) and modern Khakassians in South Siberia (Derenko et al., 2003). T4 lineages are extremely rare in modern populations, but are most often found in Eastern Europe and the Balkans (Figure 7), including one T4 individual recovered from an Eneolithic burial site in Western Ukraine (Nikitin et al., 2010) belonging to the Trypillian culture complex. Studies in Xinjiang province in northwestern China also confirm the presence of haplogroups H and K during the Bronze Age (Li et al., 2010) with additional observations of haplogroups U, T, H, and K in the region during the Iron Age (Zhang et al., 2010). Analysis of remains from the Iron Age Egyin Gol necropolis, located at the southern edge of the Lake Baikal region, also identified four individuals belonging to West Eurasian haplotypes, two from haplogroup J and two from haplogroup U5a (Keyser-Tracqui et al., 2003). An individual belonging to haplogroup U5 was also recovered from a gravesite in Eastern China dating to the late Iron Age (Xie et al., 2007). To summarize, West Eurasian haplogroups are more pronounced in South Siberia by the Iron Age, but there is evidence of population admixture as early as the Early Neolithic.

Modern South Siberian populations also display a considerable amount of admixture with European and Near Eastern populations. In their study of South Siberian mtDNA diversity, Derenko et al. (2007) report a wide variety of West Eurasian mtDNA lineages in native populations, but do not detect basal lineages of macrohaplogroups N, M, or R. This suggests that the initial peopling of the region was accomplished by dispersals from East Asia after initial migrations from the Near East by a southern coastal route (Kong et al., 2006). European haplogroups would have entered Siberia by migrations across the Central Asian steppe at a later date (Derenko et al., 2007). Studies of mtDNA and Y-chromosome diversity in Central Asian populations show widespread admixture of West and East Eurasian lineages (Semino et al., 2000; Quintana-Murci et al., 2004). This is in agreement with the finding that admixture in south Siberia was greater in southern and western populations and diminished to the north and east (Derenko et al., 2007). In contrast, a large-scale mtDNA analysis in Han Chinese (Yao et al., 2002) only detected two individuals with West Eurasian haplotypes, one from haplogroup T and another from haplogroup HV (the parent lineage of haplogroups H and V). Therefore, it appears that West Eurasian interactions with Chinese populations have been restricted with regard to mtDNA, as opposed to South Siberian populations, which are clearly admixed, but still retain a predominantly Asian haplogroup profile.

Presence of a Steppe-Belt Migration Corridor

There is clearly a continuum of genetic variation present between human populations in Eastern Europe and south Siberia. On one end of the spectrum is an Eastern European and Near Eastern mtDNA pool consisting primarily of typical West Eurasian haplogroups (such as U, K, T, J, H, etc.) with a minority of East Eurasian

lineages. At the other end, South Siberian populations display a haplogroup profile composed mostly of Asian-specific lineages, but European and Near Eastern lineages are also present. The Central Asian steppe linking the two regions is highly admixed and contains a blend of Near Eastern and Asian lineages.

It could be argued that East-West admixture observed in modern populations could be the result of recent migration, and for some individuals (such as the C1 lineage observed in northwest Germany) that is likely the case. However, recent migration is less likely for other East Eurasian lineages in Europe, especially when the presence of these lineages shows a clear regional specificity observed in multiple independent research projects. Admittedly, there appears to be an increased amount of admixture in Siberian populations following the Bronze Age, but the presence of haplogroup U5a at Neolithic cemeteries in Siberia, coupled with our observation of haplogroup C individuals in Neolithic Ukraine, supports the idea that human migration from East to West, and *vice versa*, had already been taking place prior to the onset of the Neolithic. Recent studies of the origins of haplogroup C suggest that this lineage diverged and began expanding in East Asia approximately 27 kya (Derenko et al., 2010). This means the earliest lineages of haplogroup C would have been present in East Asia prior to the Last Glacial Maximum, which occurred 22-19 kya (Yokoyama et al., 2000).

During the Last Glacial Maximum, a massive Arctic ice sheet covered large portions of northern Eurasia (Grosswald and Hughes, 2002; Forster, 2004). Furthermore, rivers such as the Ob and Lena, which would otherwise drain into the Arctic Ocean, were instead dammed by ice, forming a massive inland lake in northern Siberia (Grosswald and Hughes, 2002). Populations that had already spread throughout northern Eurasia

would have been forced to retreat to the south. The effect of this population compression is apparent in Southwest Europe, where a glacial refuge was established in what is now northwest Spain, as evidenced by the re-expansion of mtDNA and Y-chromosome haplogroups from this region (Torroni et al., 1998; Semino et al., 2000; Loogväli et al., 2004). In the case of Siberia, populations moving down from the north to escape the ice would have eventually been blocked by the mountain ranges and deserts that separate Northern Asia from India and China. In areas surrounding the Black Sea, migration would certainly have been restricted to the north and west, but not to the south and east. In this context, it would seem appropriate to classify South Siberia and the Black Sea region as boundaries of human population flow during the LGM, essentially outlining the limits of the LGM population continuum within which population migration continued along a Central Asian steppe “corridor” delineated by northern ice sheets and southern geological impediments.

CONCLUSIONS

For this thesis, ancient DNA was obtained from Neolithic and Bronze Age bone samples from the NPR. We performed DNA sequence analysis on these specimens using the HVSI region of the mtDNA genome. Our mtDNA HVSI data revealed that the gene pool of Neolithic NPR inhabitants contained mtDNA haplogroups characteristic of both Near Eastern and South Siberian populations. Bronze Age specimens belonging to the Kurgan culture were also identified as members of haplogroup C with HVSI motifs that could have been derived directly from haplotypes present in the Neolithic NPR. These results suggest substantial connections were established between populations in the Black Sea region and South Siberia at some point prior to the Neolithic. These interactions significantly influenced the NPR mtDNA pool, as evidenced by haplogroup C lineages that persisted until at least the Bronze Age. This is further supported by multiple examples of West-East admixture from both modern and ancient populations in Europe and East Asia. We propose that a migration corridor established by glacial and geographic barriers during the Last Glacial Maximum linked the Black Sea region to Siberian populations and allowed for bidirectional gene flow. This research indicates that the European gene pool at the onset of the Neolithic was influenced by an Asian source in addition to the accepted demic influx from the Near East.

LITERATURE CITED

- Abu-Amero KK, Larruga JM, Cabrera VM, Gonzalez AM. 2008. Mitochondrial DNA structure in the Arabian Peninsula. *BMC Ev Biol* 8:45.
- Adler CJ, Haak W, Donlon D, Cooper A, The Genographic Consortium. 2011. Survival and recovery of DNA from ancient teeth and bones. *J Archaeol Sci* 38:956-964.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403-410.
- Álvarez-Iglesias V, Mosquera-Miguel A, Cerezo M, Quintáns B, Zarrabeitia MT, Cuscó I, Lareu MV, García O, Pérez-Jurado L, Carracedo A, Salas A. 2009. New population and phylogenetic features of the internal variation within mitochondrial haplogroup R0. *PLoS One* 4:e5112.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457-465.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147.
- Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of populations using median networks. *Genetics* 141:743-753.
- Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37-48.
- Bandelt HJ. 2005. Mosaics of ancient mitochondrial DNA: positive indicators of nonauthenticity. *Eur J Hum Genet* 13:1106-1112.
- Barbujani G, Vernesi C, Caramelli D, Castri L, Lalueza-Fox C, Bertorelle G. 2004. Etruscan artifacts: much ado about nothing. *Am J Hum Genet* 75:923-927.
- Behar DM, Metspalu E, Kivisild T, Rosset S, Tzur S, Hadid Y, Yudkovsky G, Rosengarten D, Pereira L, Amorim A, Kutuev I, Gurwitz D, Bonne-Tamir B, Vilems R, Skorecki K. 2008. Counting the founders: the matrilineal genetic ancestry of the Jewish diaspora. *PLoS One* 3:e2062.
- Bellwood P. 2005. *First farmers: the origins of agricultural societies*. Malden, MA: Blackwell.
- Bendall KE, Sykes BC. 1995. Length heteroplasmy in the first hypervariable segment of the human mitochondrial control region. *Am J Hum Genet* 57:248-256.
- Bosch E, Calafell F, González-Neira A, Flaiz C, Mateu E, Scheil HG, Huckenbeck W, Efremovska L, Mikerezi I, Xirotis N, Grasa C, Schmidt H, Comas D. 2005.

- Paternal and maternal lineages in the Balkans show a homogenous landscape over linguistic barriers, except for the isolated Aromuns. *Ann Hum Genet* 70:459-487.
- Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, Antanaitis-Jacobs I, Haidle MN, Jankauskas R, Kind CJ, Lueth F, Terberger T, Hiller J, Matsumura S, Forster P, Burger J. 2009. Genetic discontinuity between local hunter-gatherers and Central Europe's first farmers. *Science* 326:137-140.
- Brehm A, Pereira L, Kivisild T, Amorim A. 2003. Mitochondrial portraits of the Madeira and Açores archipelagos witness different genetic pools of its settlers. *Hum Genet* 114:77-86.
- Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. *P Natl Acad Sci USA* 76:1967-1971.
- Brown WM. 1980. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *P Natl Acad Sci USA* 77:3605-3609.
- Burton ML, Moore CC, Whiting JWM, Romney AK, Aberle DF, Barcelo JA, Dow MM, Guyer JI, Kronenfeld DB, Levy JE, Linnekin J. 1996. Regions based on social structure. *Curr Anthropol* 37:87-123.
- Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjeiva L. 1996. From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann Hum Genet* 60:35-49.
- Cann RA, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31-36.
- Calcagno JM, Gibson KR. 1988. Human dental reduction: natural selection or the probable mutation effect. *Am J Phys Anthropol* 77:505-517.
- Cavalli-Sforza LL, Minch E. 1997. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 61:247-251.
- Chang DD, Clayton DA. 1985. Priming of human mitochondrial DNA replication occurs at the light-strand promoter. *P Natl Acad Sci USA* 82:351-355.
- Clayton DA. 2000. Transcription and replication of mitochondrial DNA. *Human Reprod* 15:11-17.
- Cooper A, Poinar HN. 2000. Ancient DNA: Do it right or not at all. *Science* 289:1139.
- Davis AF, Clayton DA. 1996. In situ localization of mitochondrial DNA replication in intact mammalian cells. *J Cell Biol* 135:883-893.
- Derbeneva OA, Starikovskaya EB, Wallace DC, Sukernik RI. 2002. Traces of early Eurasians in the Mansi of Northwest Siberia revealed by mitochondrial DNA analysis. 2002. *Am J Hum Genet* 70:1009-1014.

- Derenko MV, Grzybowski T, Malyarchuk BA, Czarny J, Miścicka-Śliwka D, Zhakarov IA. 2001. The presence of mitochondrial haplogroup X in Altaians from South Siberia. *Am J Hum Genet* 69:237–241.
- Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Miścicka-Śliwka D, Woźniak M, Zakharov IA. 2003. Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet* 67:391-411.
- Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I, Perkova M, Dorzhu C, Luzina F, Lee HK, Vanecek T, Villems R, Zakharov I. 2007. Phylogeographic analysis of mitochondrial DNA in Northern Asian populations. *Am J Hum Genet* 81:1025-1041.
- Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Rogalla U, Perkova M, Dambueva I, Zakharov I. 2010. Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in Northern Asia. *PLoS One* 5:e15214.
- Ebenesersdóttir SS, Sigurðsson A, Sánchez-Quinto F, Lalueza-Fox C, Stefánsson, Helgason A. 2011. A new subclade of mtDNA haplogroup C1 found in Icelanders: Evidence of pre-Columbian contact? *Am J Phys Anthropol* 144:92-99.
- Ekstrand MI, Falkenberg M, Rantanen A, Park CB, Gaspari M, Hultenby K, Rustin P, Gustafsson CM, Larsson NG. 2004. Mitochondrial transcription factor A regulates mtDNA copy number in mammals. *Hum Mol Genet* 13:935-944.
- Endicott P, Sanchez JJ, Pilcher I, Brotherton P, Brooks J, Egarter-Vigl E, Cooper A, Pramstaller P. 2009. Genotyping human ancient mtDNA control and coding region polymorphisms with a multiplexed single-base-extension assay: the singular maternal history of the Tyrolean Iceman. *BMC Genet* 10:29.
- Fernández-Silva P, Enriquez JA, Montoya J. 2003. Replication and transcription of mammalian mitochondrial DNA. *Exp Physiol* 88:41-56.
- Finnilä S, Lehtonen MS, Majamaa K. 2001. Phylogenetic network for European mtDNA. *Am J Hum Genet* 68:1475-1484.
- Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59:935-945.
- Forster P. 2004. Ice Ages and the mitochondrial DNA of human dispersals: a review. *Phil Trans R Soc Lond B* 359:255-264.
- Freeland JR. 2005. Phylogeography. In: Freeland, JR. *Molecular Ecology*. Chichester, England: John Wiley and Sons, p 155-200.
- Fu YX, Li WH. 1999. Coalescing into the 21st Century: an overview and prospects of coalescent theory. *Theor Popul Biol* 56:1-10.

- Gilbert MTP, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N, Cooper A. 2003. Distribution patterns of post-mortem damage in human mitochondrial DNA. *Am J Hum Genet* 72:32-47.
- Giles RE, Blanc H, Cann HM, Wallace DC. 1980. Maternal inheritance of human mitochondrial DNA. *P Natl Acad Sci USA* 77:6715-6719.
- Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA. 2007. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol Biol Evol* 24:757-768.
- González AM, Karadsheh N, Maca-Meyer N, Flores C, Cabrera VM, Larruga JM. 2008. Mitochondrial DNA variation in Jordanians and their genetic relationship to other Middle East populations. *Ann Hum Biol* 35:212-231.
- Goodacre S, Helgason A, Nicholson J, Southam L, Ferguson L, Hickey E, Vega E, Stefánsson K, Ward R, Sykes B. 2005. Genetic evidence for a family-based Scandinavian settlement of Shetland and Orkney during the Viking periods. *Heredity* 95:129-135.
- Grosswald MK, Hughes TJ. 2002. The Russian component of an Arctic Ice Sheet during the Last Glacial Maximum. *Quaternary Sci Rev* 21:121-146.
- Grzybowski T, Malyarchuk BA, Derenko MV, Perkova MA, Bedarek J, Woźniak M. 2007. Complex interactions of the Eastern and Western Slavic populations with other European groups as revealed by mitochondrial DNA analysis. *For Sci Int-Gen* 1:141-147.
- Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, Tänzer M, Villems R, Renfrew C, Gronenborn D, Alt KW, Burger J. 2005. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310: 1016-1018.
- Haak W, Brandt G, de Jong HN, Meyer C, Ganslmeier R, Heyd V, Hawkesworth C, Pike AWG, Meller H, Alt KW. 2008. Ancient DNA, strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. *P Natl Acad Sci* 105: 18226-18231.
- Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CSI, Brandt G, Schwarz C, Nicklisch N, Dresely V, Fritsch B, Balanovska E, Villems R, Meller H, Alt KW, Cooper A, The Genographic Consortium. 2010. Ancient DNA from European early Neolithic farmers reveals their Near Eastern affinities. *PLoS Biol* 8:e1000536.
- Hagelberg E, Goldman N, Lió P, Whelan S, Schiefenhövel W, Clegg JB. 1999. Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proc R Soc Lond B* 266:485-492.

- Hagelberg E, Goldman N, Lió P, Whelan S, Schiefenhövel W, Clegg JB, Bowden DK. 2000. Evidence for mitochondrial DNA recombination in a human population of island Micronesia: correction. *Proc R Soc Lond B* 267:1595-1596.
- Helgason A, Hickey E, Goodacre S, Bosnes V, Stefánsson K, Ward R, Sykes B. 2001. MtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* 68: 723-737.
- Helgason A, Nicholson G, Stefánsson K, Donnelly P. 2003. A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Ann Hum Genet* 67:281-297.
- Hedges REM, Lee-Thorp JA, Tuross NC. 1995. Is tooth enamel carbonate a suitable material for radiocarbon dating? *Radiocarbon* 37:285-290.
- Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312:282-284.
- Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708-713.
- Irwin J, Saunier J, Strouss K, Paintner C, Diegoli T, Sturk K, Kovatsi L, Brandstätter A, Cariolou MA, Parson W, Parsons TJ. 2008. *Int J Leg Med* 122:87-89.
- Ivanova SV, Petrenko VG, Betchinnikova NE. 2005. Kurgans of ancient herdsmen from the South Bug and Dnister interfluvium. *Odessa: KP OGT*, 207 p.
- Jacobs K. 1994. Human dento-gnathic metric variation in Mesolithic/Neolithic Ukraine: possible evidence of demic diffusion in the Dnieper Rapids region. *Am J Phys Anthropol* 95:1-26.
- Jobling MA, Tyler-Smith C. 2003. The human Y-chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598-612.
- Keyser C, Bouakaze C, Crubézy E, Nikolaev VG, Montagnon D, Reis T, Ludes B. 2009. Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum Genet* 126:395-410.
- Keyser-Tracqui C, Crubézy E, Ludes B. 2003. Nuclear and mitochondrial DNA analysis of a 2,000-year-old necropolis in the Egyin Gol Valley of Mongolia. *Am J Hum Genet* 73:247-260.
- Kingman JFC. 1982. On the genealogy of large populations. *J Appl Probab* 19:27-43.
- Kong QP, Bandelt HJ, Sun C, Yao YG, Salas A, Achilli A, Wang CY, Zhong L, Zhu CL, Wu SF, Torroni A, Zhang YP. 2006. Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations *Hum Mol Genet* 15:2076-2086.

- Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A, Pääbo S. 2010. A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol* 20:231-236.
- Kruskal JB. 1956. On the shortest spanning subtree of the graph and the travelling salesman problem. *Proc Amer Math Soc* 7:48–57.
- Kumar S, Subramanian S. 2002. Mutation rates in mammalian genomes. *P Natl Acad Sci* 99:803-808.
- Lalueza-Fox C, Sampietro ML, Gilbert MTP, Castri L, Facchini F, Pettener D, Bertranpetit J. 2004. Unraveling migrations in the steppe: mitochondrial DNA sequences from ancient Central Asians. *Proc R Soc B* 271:941-947.
- Lee DY, Clayton DA. 1998. Initiation of mitochondrial DNA replication by transcription and R-loop processing. *J Biol Chem* 273:30614-30621.
- Li C, Li H, Cui Y, Xie C, Cai D, Li W, Mair VH, Xu Z, Zhang Q, Abuduresule I, Jin L, Zhu H, Zhou H. 2010. Evidence that a west-east admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC Biol* 8:15.
- Lillie M, Budd C, Potekhina I, Hedges R. 2009. The radiocarbon reservoir effect: new evidence from the cemeteries of the middle and lower Dnieper basin, Ukraine. *J Archaeol Sci* 36:256-264.
- Lillie M, Richard M. 2000. Stable isotope analysis and dental evidence of diet at the Mesolithic-Neolithic transition in Ukraine. *J Archaeol Sci* 27:965-972.
- Loogväli EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk HV, Parik J, Pennarun E, Laos S, Lunkina A, Golubenko M, Barac L, Perićić M, Balanovsky OP, Gusar V, Khusnutdinova EK, Stepanov V, Puzyrev V, Rudan P, Balanovska EV, Grechanina E, Richard C, Moisan JP, Chaventré A, Anagnou NP, Pappa KI, Michalodimitrakis EN, Claustres M, Gölge M, Mikerezi I, Usanga E, Villems R. 2004. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol* 21: 2012-2021.
- Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM. 2001. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2:13.
- Macauley V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonnét-Tamir B, Sykes B, Torroni A. 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 64:232-249.
- Malmström H, Gilbert MTP, Thomas MG, Brandström M, Storå J, Molnar P, Andersen PK, Bendixen C, Holmlund G, Götherström A, Willerslev E. 2009. Ancient DNA

- reveals lack of continuity between Neolithic hunter-gatherers and contemporary Scandinavians. *Curr Biol* 19:1758-1762.
- Malyarchuk BA, Derenko MV. 1999. Molecular instability of the mitochondrial haplogroup T sequences at nucleotide positions 16292 and 16296. *Ann Hum Genet* 63:489-497.
- Malyarchuk BA, Derenko MV. 2001. Mitochondrial DNA variability in Russians and Ukrainians: implication to the origin of Eastern Slavs. *Ann Hum Genet* 65:63-78.
- Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Woźniak M, Miścicka-Śliwka D. 2002. Mitochondrial DNA variability in Poles and Russians. *Ann Hum Genet* 66:261-283.
- Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Drobníč K, Miścicka-Śliwka D. 2003. Mitochondrial DNA variability in Bosnians and Slovenians. *Ann Hum Genet* 67:412-425.
- Malyarchuk B, Derenko M, Grzybowski T, Lunkina A, Czarny J, Rychkov S, Morozova I, Denisova G, Miścicka-Śliwka D. 2004. Differentiation of mitochondrial DNA and Y chromosomes in Russian populations. *Hum Biol* 76:877-900.
- Malyarchuk BA, Grzybowski T, Derenko MV, Czarny, and Miścicka-Śliwka D. 2006a. Mitochondrial DNA diversity in the Polish Roma. *Ann Hum Genet* 70:195-206.
- Malyarchuk BA, Vanecek T, Perkova MA, Derenko MV, Sip M. 2006b. Mitochondrial DNA variability in Czech population, with application to the ethnic history of Slavs. *Hum Biol* 78:681-696.
- Malyarchuk BA, Perkova MA, Derenko MV, Vanecek T, Lazur J, Gomolcak P. 2008. Mitochondrial DNA variability in Slovaks, with application to the Roma origin. *Ann Hum Genet* 72:228-240.
- Malyarchuk BA, Derenko MV, Grzybowski T, Perkova M, Rogalla U, Vanecek T, Tsybovsky I. 2010. The peopling of Europe from the mitochondrial haplogroup U5 perspective. *PLoS One* 5:e10285.
- McNeill WH. 1984. Human migration in historical perspective. *Popul Dev Rev* 10:1-18.
- Merriwether DA, Hodgson JA, Friedlaender, Allaby R, Cerchio S, Koki G, Friedlaender JS. 2005. Ancient mitochondrial M haplogroups identified in the Southwest Pacific. *P Natl Acad Sci USA* 102:13034-13039.
- Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MTP, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R. 2004. Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5:26.

- Mogentale-Profizi N, Chollet L, Stévanovitch A, Dubut V, Poggi C, Pradié MP, Spadoni JL, Gilles A, Béraud-Colomb E. 2001. Mitochondrial DNA sequence diversity in two groups of Italian Veneto speakers from Veneto. *Ann Hum Genet* 65:153-166.
- Montiel R, Malgosa A, Francalacci P. 2001. Authenticating ancient human mitochondrial DNA. *Hum Biol* 73:689-713.
- Montoya J, Christianson T, Levens D, Rabinowitz M, Attardi G. 1982. Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. *Proc Natl Acad Sci USA* 79:7195-7199.
- Mooder KP, Thomson TA, Weber AW, Bazaliiskii VI, Bamforth FJ. 2005. Uncovering the landscape of prehistoric cis-Baikal. In: Weber A, Katzenberg MA, Schurr TG, editors. *Prehistoric hunter-gatherers of the Baikal region, Siberia: bioarcheological studies of past life ways*. Philadelphia: University of Pennsylvania Museum of Archaeology and Anthropology Press, p 107-119.
- Mooder KP, Schurr TG, Bamforth FJ, Bazaliiski VI, Savel'ev NA. 2006. Population affinities of Neolithic Siberians: a snapshot from prehistoric Lake Baikal. *Am J Phys Anthropol* 129:349-361.
- Nádasi E, Gyűrűs P, Czakó M, Bene J, Kosztolányi Sz., Fazekas Sz., Dömösi P, Melegh B. 2007. Comparison of mtDNA haplogroups in Hungarians with four other European populations: a small incidence of descents with Asian origin. *Acta Biol Hung* 58:245-256.
- Nikitin AG, Kochkin IT, June CM, Willis CM, McBain I, Videiko MY. 2009. Mitochondrial DNA sequence variation in the Boyko, Hutsul, and Lemko populations of the Carpathian Highlands. *Hum Biol* 81:43-58.
- Nikitin AG, Sokhatsky MP, Kovalyukh MM, Videiko MY. 2010. Comprehensive site chronology and ancient mitochondrial DNA analysis from Verteba Cave – a Trypillian culture site of Eneolithic Ukraine. *IANSA* 1:9-18.
- Nordborg M. 2001. Coalescent theory. In: DJ Balding, MJ Bishop, C Cannings, editors. *Handbook of Statistical Genetics*. Chichester, UK: Wiley, p 843-877.
- Orekhov V, Poltoraus A, Zhivotovsky LA, Spitsyn V, Ivanov P, Yankovsky N. 1999. Mitochondrial DNA sequence diversity in Russians. *FEBS Lett* 445:197-201.
- Pääbo S. 1985. Molecular cloning of ancient Egyptian mummy DNA. *Nature* 314:644-645.
- Pääbo S, Gifford JA, Wilson AC. 1988. Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids Res* 16:9775-9787.
- Pääbo S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *P Natl Acad Sci USA* 86:1939-1943.

- Pakendorf B, Stoneking M. 2005. Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* 6:165–183.
- Parson W, Dür A. 2007. EMPOP – a forensic mtDNA database. *Forensic Sci Int-Gen* 1:88-92.
- Pereira L, Prata MJ, Amorim A. 2005. Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann Hum Genet* 64:491-506.
- Pereira L, Gonçalves J, Franco-Duarte R, Silva J, Rocha T, Arnold C, Richards M, Macaulay V. 2007. No evidence for an mtDNA role in sperm motility: data from complete sequencing of asthenozoospermic males. *Mol Biol Evol* 24:868-874.
- Pereira F, Soares P, Carneiro J, Pereira L, Richards MB, Samuels DC, Amorim A. 2008. Evidence for variable selective pressures at a large secondary structure of the human mitochondrial DNA control region. *Mol Biol Evol* 25:2759-2770.
- Pfeiffer H, Brinkmann B, Hühne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P. 1999. Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *Int J Legal Med* 112 :291–298.
- Pfeiffer H, Forster P, Ortmann C, Brinkmann B. 2001. The results of an mtDNA study of 1200 inhabitants of a German village in comparison to other Caucasian databases and its relevance for forensic casework. *Int J Leg Med* 114:169-172.
- Pike DA. 2006. Phylogenetic networks for the human mtDNA haplogroup T. *JoGG* 2:1-11.
- Pinhasi R, von Cramon-Taubadel, N. 2009. Craniometric data supports demic diffusion model for the spread of agriculture into Europe. *PLoS ONE* 4:e6747.
- Poetsch M, Wittig H, Krause D, Lignitz E. 2003. Mitochondrial diversity of a northeast German population sample. *For Sci Int* 137:125-132.
- Poetsch M, Wittig H, Krause D, Lignitz E. 2004. Corrigendum to “Mitochondrial diversity of a northeast German population sample.” 145:73-77.
- Polzin T, Daneschmand SV. 2003. On Steiner trees and minimum spanning trees in hypergraphs. *Oper Res Lett* 31:12-20.
- Posada D, Crandall KA. 2001. Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol* 16:37-45.
- Pruvost M, Geigl EM. 2004. Real-time quantitative PCR to assess the authenticity of ancient DNA amplification. *J Archaeol Sci* 31:1191-1197.
- Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santchiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Mehdi SQ, Torroni A, McElreavey K. 2004. Where East meets

West: the complex mtDNA landscape of the Southwest and Central Asian corridor. *Am J Hum Genet* 74:827-845.

Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk HV, Parik J, Loogväli EL, Derenko M, Malyarchuk B, Bermisheva M, Zhadanov S, Pennarun E, Gubina M, Golubenko M, Damba L, Fedorova S, Gusar V, Grechanina E, Mikerezi I, Moisan JP, Chaventré A, Khusnutdinova E, Osipova L, Stepanov V, Voevoda M, Achilli A, Rengo C, Rickards O, Franco de Stefano G, Papiha S, Beckman L, Janicijevic B, Rudan P, Anagnou N, Michalodimitrakakis E, Koziel S, Usanga E, Geberhiwot T, Herrnstadt C, Howell N, Torroni A, Villems R. 2003. Origin and diffusion of mtDNA haplogroup X. *Am J Hum Genet* 73:1178-1190.

Ricaut FX, Fedoseeva A, Keyser-Tracqui C, Crubézy E, Ludes B. 2005. Ancient DNA analysis of human Neolithic remains found in northeastern Siberia. *Am J Phys Anthropol* 126:458-462.

Richards MB, Macaulay VA, Bandelt HJ, Sykes BC. 1998. Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 62:241-260.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Calì F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Bedelli M, Di Rienzo A, Novelletto A, Oppenheim A, Nørby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozzari R, Torroni A, Bandelt HJ. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67:1251-1276.

Roberti M, Musicco C, Polosa PL, Milella F, Gadaleta MN, Cantatore P. 1998. Multiple protein-binding sites in the TAS-region of human and rat mitochondrial DNA. *Biochem Biophys Res Commun* 243:36-40.

Rokas A, Ladoukakis E, Zouros E. 2003. Animal mitochondrial DNA recombination revisited. *TRENDS Ecol Evol* 18:411-417.

Rosenberg NA, Nordborg M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat Rev Genet* 3:380-390.

Sampietro ML, Caramelli D, Lao O, Calafell F, Comas D, Lari M, Agusti B, Bertranpetit J, Lalueza-Fox C. 2005. The genetics of the pre-Roman Iberian Peninsula: a mtDNA study of ancient Iberians. *Ann Hum Genet* 69:535-548.

Sampietro ML, Lao O, Caramelli D, Lari M, Pou R, Marti M, Bertranpetit J, Lalueza-Fox C. 2007. Palaeogenetic evidence supports a dual mode of Neolithic spreading into Europe. *Proc R Soc B* 274:2161-2167.

Santos C, Montiel R, Anglés N, Lima M, Francalacci P, Malgosa A, Abade A, Aluja MP. 2004. Determination of human Caucasian mitochondrial DNA haplogroups by means of a hierarchical approach. *Hum Biol* 76:431-453.

- Schurr TG, Wallace DC. 2002. Mitochondrial DNA diversity in Southeast Asian populations. *Hum Biol* 74:431-452.
- Schwartz M, Vissing J. 2002. Paternal inheritance of mitochondrial DNA. *New Engl J Med* 347:576-580.
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA. 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290:1155-1159.
- Shitara H, Kaneda H, Sato A, Inoue K, Ogura A, Yonekawa, H, Hayashi JI. 2000. Selective and continuous elimination of mitochondria microinjected into mouse eggs from spermatids, but not from liver cells, occurs throughout embryogenesis. *Genetics* 156:1277-1284.
- Shlush LI, Behar DM, Yudkovsky G, Templeton A, Hadid Y, Basis F, Hammer M, Itzkovitz S, Skorecki K. 2008. The Druze: a population genetic refugium of the Near East. *PLoS One* 3:e2105.
- Sigurðardóttir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P. 2000. The mutation rate in the human mtDNA control region. *Am J Hum Genet* 66:1599-1609.
- Sokal RR, Oden NL, Wilson C. 1991. Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* 351:143-145.
- Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC. 1998. MtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. *Am J Hum Genet* 63:1473-1491.
- Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E, Torroni A, Brown MD, Lott MT, Hosseini SH, Huoponen K, Wallace DC. 2005. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet* 69:67-89.
- St. John JC, Facucho-Oliveira J, Jiang Y, Kelly R, Salah R. 2010. Mitochondrial DNA transmission, replication and inheritance: a journey from the gamete through the embryo and into the offspring and embryonic stem cells. *Hum Reprod Update* 16:488-509.
- Telegin DJ. 1987. Neolithic cultures of the Ukraine and adjacent areas and their chronology. *J World Prehist* 1:307-331.
- Telegin DY, Potekhina ID. 1987. Neolithic cemeteries and populations in the Dnieper Basin. Oxford: British Archeological Reports. International series S383.

- Telegin DYa, Potekhina ID, Lillie M, Kovaliukh MM. 2002. The chronology of the Mariupol-type cemeteries of Ukraine revisited. *Antiquity* 76:356-363.
- Thorpe, IJ. 1996. The origins of agriculture in Europe. New York: Routledge.
- Tolk HV, Perićić M, Barać L, Klarić IM, Janićijević B, Rudan I, Parik J, Villems R, Rudan P. 2000. MtDNA haplogroups in the populations of Croatian Adriatic islands. *Coll Antropol* 24:267-279.
- Torroni A, Schurr TG, Yang CC, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130:153-162.
- Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* 53:563-590.
- Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, Wallace DC. 1994. MtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet* 55:760-776.
- Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC. 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835-1850.
- Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus ML, Bonn -Tamir B, Scozzari R. 1998. MtDNA analysis reveals a major Late Paleolithic population expansion from Southwestern to Northeastern Europe. *Am J Hum Genet* 62:1137-1152.
- Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ. 2006. Harvesting the fruit of the human mtDNA tree. *TRENDS Genet* 22:339-345.
- Underhill PA, Shen P, Lin AA, Passarino G, Yang WH, Kauffman E, Bonn -Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ. 2000. Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358-361.
- Van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30:E386-E394.
- Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev

- A, Bodmer WF. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 98:10244-10249.
- Willerslev E, Cooper A. 2005. Ancient DNA. *Proc R Soc B* 272:3-16.
- Woodward SR, Weyand NJ, Bunnell M. 1994. DNA sequence from Cretaceous period bone fragments. *Science* 266:1229-1232.
- Xie CZ, Li CX, Cui YQ, Zhang QC, Fu YQ, Zhu H, Zhou H. 2007. Evidence of ancient DNA reveals the first European lineage in Iron Age Central China. *Proc R Soc B* 274:1597-1601.
- Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70:635–651.
- Yokoyama Y, Lambeck K, De Deckker P, Johnston P, Fifield LK. 2000. Timing of the Last Glacial Maximum from observed sea-level minima. *Nature* 406: 713-716.
- Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P, Yuldasheva N, Ruzibakiev R, Xu J, Shu Q, Du R, Yang H, Hurles ME, Robinson E, Gerelsaikhan T, Dashnyam B, Mehdi SQ, Tyler-Smith C. 2003. The genetic legacy of the Mongols. *Am J Hum Genet* 72:717-721.
- Zhang F, Xu Z, Tan J, Sun Y, Xu B, Li S, Zhao X, Zhou H, Gong G, Zhang J, Jin L. 2010. Prehistorical East-West admixture of maternal lineages in a 2,500-year-old population in Xinjiang. *Am J Phys Anthropol* 142:314-320.
- Zischler H, Höss M, Handt O, von Haeseler A. 1995. Detecting dinosaur DNA. *Science* 268:1192-1193.

Table 1. MtDNA HVSI profiles of researchers performing aDNA analysis or involved in prior handling of the specimens.

Researcher	Differences from rCRS (Andrews et al., 1999)
#1	16126-16163-16186-16189-16294
#2	16304
#3	16298

Table 2. Primer pairs used to amplify mtDNA HVSI in Neolithic and Bronze Age bone samples.

Primer	Sequence	Section Amplified	Fragment Size	Tm (°C)
HVS1FF HVS13R	5'- ACTCCACCATTAGCACCCAA-3' 5'- GGTGGCTGGCAGTAATGTACGAA-3'	15994-16092	142bp	53.0 57.1
HVS13F HVS12R	5'- TGACTCACCCATCAACAACCGC-3' 5'- CTTGCTTGTAAGCATGGGGA-3'	16086-16188	145bp	57.5 60.0
L16163 H16278	5'- ACTTGACCACCTGTAGTACATAA-3' 5'- GTTAAGGGTGGGTAGGTTTGTGG-3'	16164-16277	161bp	52.8 57.6
HVS14F HVS14R	5'- GCAACTCCAAAGCCACCCCTCA-3' 5'- GATGGTGGTCAAGGGACCCCTA-3'	16266-16385	164bp	59.3 57.1

Table 3. MtDNA HVSI sequences of Neolithic and Bronze Age individuals from the North Pontic Region.

Sample	Age (BP)	Time Period	HVSI Sequence (+16000) ^{c, d}	RFLPs (where available)	Haplogroup
Ni58	2,305±45 ^b	Neolithic	061, 223, 298, 327		C
Ya34	6,195±80	Neolithic	223, 298, 327, 357		C4a2'3'4'
Ya45	6,360±60	Neolithic	223, 298, 327		C
D1.8	3,940±70	Bronze Age	223, 278, 298, 327, 357		C4a3
L8	3,990±70	Bronze Age	218, 223, 288, 298, 305A-T, 327, 357		C4a6
L15	3,740±70	Bronze Age	218, 223, 298, 327, 357		C4a6
DD33	6,175±60	Neolithic	311	-7025AluI	H
DD38	N/A	Neolithic	<i>080</i>	-7025AluI	H
Ya17	6,360±75	Neolithic	241		H? ^e
Ya36	6,260±180	Neolithic	320		H? ^e
Ya57	N/A	Neolithic	241	-7025AluI	H
Ya64	6,330±90	Neolithic	064, 240	-7025AluI	H
Ya32	4,500±120	Neolithic	294, 296		T
Ya54	6,593±35	Neolithic	294, 296		T
Ni79	5,200±30	Neolithic	294, 296		T
Ya19	6,370±60	Neolithic	343		U3
Ni94	6,225±75	Neolithic	256, 270, 356		U5a

^aYa – Yasinkovatka; DD – Dereivka; Ni – Nikolskoye;

^bThe radiocarbon date for this sample is likely inaccurate (see text for explanation)

^cAll mutations listed are transitions compared to the rCRS, unless noted explicitly.

^dItalics denote incomplete HVSI sequences.

^eRFLP status could not be determined for these samples.

Table 4. Differences from the rCRS in mtDNA HVSI of pre-historic NPR inhabitants.

Sample Name	Hg	16061	16064	16080	16218	16223	16240	16241	16249	16256	16270	16278	16288	16294	16296	16298	16305	16311	16320	16327	16343	16356	16357
rCRS		T	T	A	C	C	A	A	T	C	C	C	T	C	C	T	A	T	C	C	A	T	T
Ni58	C	C	-	-	-	T	-	-	-	-	-	-	-	-	-	C	-	-	-	T	-	-	-
Ya34	C4a2'3'4'	-	-	-	-	T	-	-	-	-	-	-	-	-	-	C	-	-	-	T	-	-	C
Ya45	C	-	-	-	-	T	-	-	-	-	-	-	-	-	-	C	-	-	-	T	-	-	-
D1.8	C4a3	-	-	-	-	T	-	-	-	-	-	T	-	-	-	C	-	-	-	T	-	-	C
L8	C4a6	-	-	-	T	T	-	-	-	-	-	-	C	-	-	C	T	-	-	T	-	-	C
L15	C4a6	-	-	-	T	T	-	-	-	-	-	-	-	-	-	C	-	-	-	T	-	-	C
DD33	H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-
DD38	H	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ya17	H	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ya36	H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-
Ya57	H	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ya64	H	-	C	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ya32	T	-	-	-	-	-	-	-	-	-	-	-	-	T	T	-	-	-	-	-	-	-	-
Ya54	T	-	-	-	-	-	-	-	-	-	-	-	-	T	T	-	-	-	-	-	-	-	-
Ni79	T	-	-	-	-	-	-	-	-	-	-	-	-	T	T	-	-	-	-	-	-	-	-
Ya19	U3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-
Ni94	U5a	-	-	-	-	-	-	-	-	T	T	-	-	-	-	-	-	-	-	-	-	C	-

Figure Legends

Fig. 1.

1a) Illustration of human mtDNA (modified from Pakendorf and Stoneking, 2005). The coding region contains genes for cellular respiration proteins, including NADH dehydrogenase (7 subunits), cytochrome b, cytochrome oxidase (3 subunits), and ATPase (2 subunits), as well as genes for 12S and 16S rRNA. Twenty-two tRNA genes are also present (serine and leucine are represented twice). The first hypervariable region (HVSI) of the non-coding control region is indicated by a gray wedge.

1b) Revised Cambridge Reference Sequence (Accession Number: NC_012920) for human mtDNA from positions 16001 to 16400. The region highlighted in blue denotes the HVSI (positions 16024-16383). The yellow segment denotes the poly-c region (positions 16180-16193) prone to length variation. Nucleotide positions labeled in red (positions 16093, 16189, 16311, 16362) are considered unstable positions not useful for phylogeographic analysis.

Fig. 2. Schematic representation of mtDNA haplogroups. Node colors describe the geographic affinities of each haplogroup on a continental scale (Black = Africa; Dark Grey = Asia ; Light Grey = Europe and Near East). Haplogroups have historically been defined based upon mtDNA sequence polymorphisms and/or restriction enzyme cutting sites. The diagnostic sites for each haplogroup are listed on the link leading to that group.

Fig. 3. Geographic location of burial sites utilized in this study. Neolithic sites are marked by arrows (Ya – Yasinkovatka; DD – Dereivka; Ni – Nikolskoye). Numbers in parentheses indicate number of individuals sequenced from each site. The approximate location of Bronze Age Kurgan sites in southwest Ukraine (Odessa province) is also marked.

Fig. 4. Median-Joining Network for mtDNA Haplogroup U5a. Branch labels are nucleotide differences compared to rCRS(+16000). Node sizes are proportional to the frequency of each lineage. The Neolithic Ukrainian sample within the network (Ni94) is marked by an arrow. Dark grey nodes signify European samples, light grey nodes are Near Eastern, and black nodes are Asian. Nucleotide positions that have undergone back-mutations have been underlined. Sources for European populations: Álvarez-Iglesias et al., 2009; Bosch et al., 2005; Bramanti et al., 2009; Brehm et al., 2003; Grzybowski et al., 2007; Haak et al., 2008; Haak, pers. comm.; Irwin et al., 2008; Maca-Meyer et al., 2001; Malyarchuk and Derenko, 2001; Malyarchuk et al., 2002; Malyarchuk et al., 2003; Malyarchuk et al., 2004; Malyarchuk et al., 2006b; Malyarchuk et al., 2008; Mogentale-Profizi et al., 2001; Pereira et al., 2000; Poetsch et al., 2004; Sampietro et al., 2005; Torroni et al., 1996. Sources for Near Eastern populations: Abu-Amero et al., 2008; Behar et al., 2008; Macaulay et al., 1999; Shlush et al., 2008. Sources for Asian

populations: Derbeneva et al., 2002; Derenko et al., 2003; Derenko et al., 2007; Keyser et al., 2009; Lalueza-Fox et al., 2004; Metspalu et al., 2004; Quintana-Murci et al., 2004.

Fig. 5. Median-Joining Network for mtDNA Haplogroup C. Branch labels are nucleotide differences compared to rCRS(+16000). Node sizes are proportional to the frequency of each lineage. Black nodes represent European samples. Neolithic Ukrainian samples (samples Ya34, Ya45, Ni58) have been labeled **in bold** and their positions within the network have been marked with arrows. Three additional samples from Bronze Age Ukrainian populations (D1.8, L8, L15) have been labeled *in italics* and marked with arrows. Note that all Bronze Age samples descend directly from the node defined by Neolithic sample Ya34. Samples L8 and L15 occupy a previously undefined branch of haplogroup C, which we have labeled C4a6. Nucleotide positions that have undergone back-mutations have been underlined. Sources in addition to Fig. 4 for European populations include: Calafell et al., 1996; Nikitin, unpublished data; Orekhov et al., 1999; Pfeiffer et al., 2001; Ebenesersdottir et al., 2011. Additional Asian sources include: Ingman et al., 2000; Keyser-Tracqui et al., 2003; Li et al., 2010; Ricaut et al., 2005; Starikovskaya et al., 1998; Starikovskaya et al., 2005.

Fig. 6. Map of Europe showing geographic locations of haplogroup C. Red nodes are Neolithic samples, orange nodes are Bronze Age samples, and blue nodes are samples from mtDNA studies of modern populations. Large nodes represent populations with 3+ individuals, medium nodes represent 2 individuals, and small nodes represent unique occurrences. Source papers for labeled individuals are cited in the text.

Fig. 7. Map of Europe showing geographic locations of haplogroup T4. The orange node is an Eneolithic individual from Verteba Cave in Western Ukraine. Blue nodes are individuals from modern populations. Medium nodes represent 2 individuals, and small nodes represent unique occurrences. European T4 individuals are represented in studies by Belyaeva et al. (2003), Bosch et al. (2005), Irwin et al. (2008), Macaulay et al. (1999), Malyarchuk et al. (2002), Malyarchuk et al. (2003), Passarino et al. (2002), Pereira et al. (2000), and Torroni et al. (1996). Two Palestinian members of T4 are from Behar et al. (2008). Eight additional T4 individuals were detected by searching the public database provided by FamilyTreeDNA (www.mitosearch.org).

Fig. 1a.

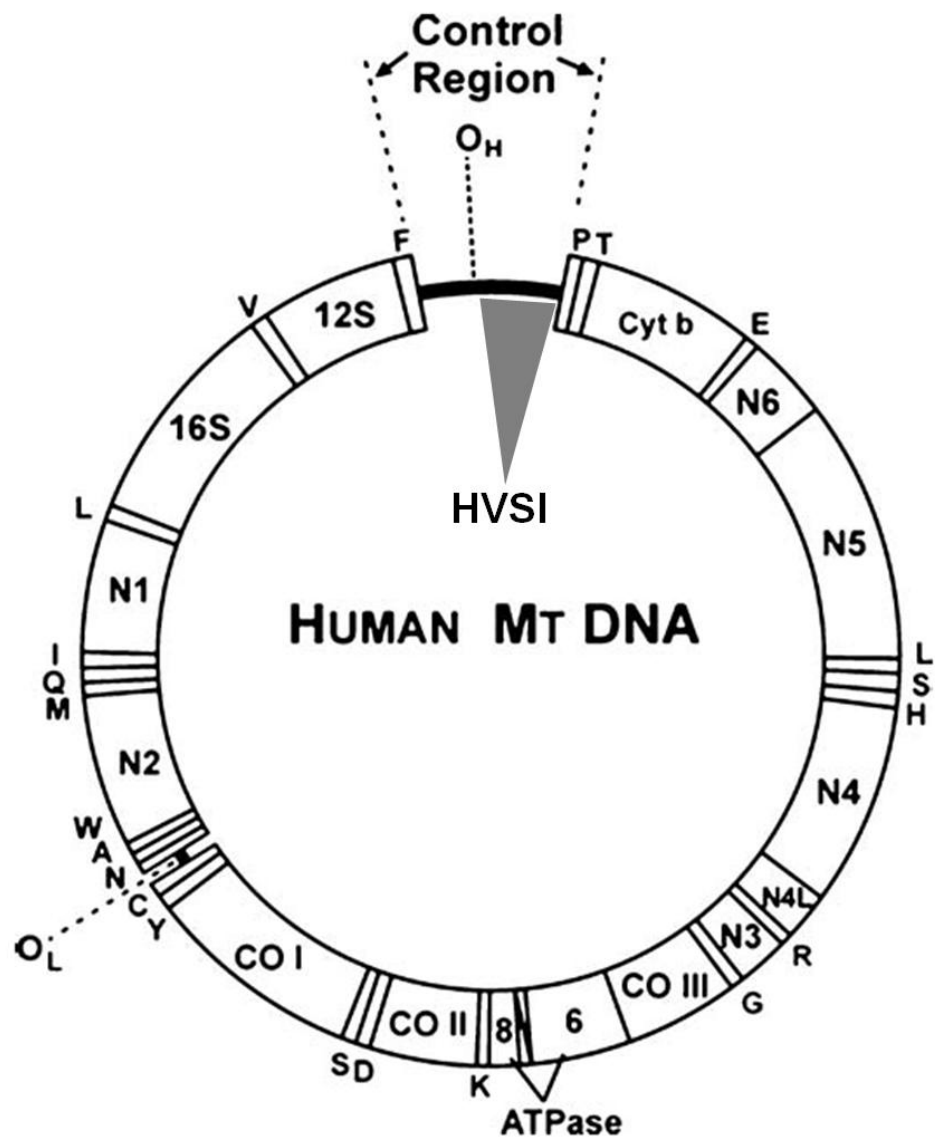


Fig. 1b.

16001 attctaattt aaactattct ctgttctttc atggggaagc agatttgggt accaccaag
16061 tattgactca cccatcaaca accgctatgt atctcgtaca ttactgccag ccaccatgaa
16121 tattgtacgg taccataaat acttgaccac ctgtagtaca taaaaccca atccacatca
16181 aaacccctc cccatgctta caagcaagta cagcaatcaa cctcaacta tcacacatca
16241 actgcaactc caagccacc cctcaccac taggatacca acaaactac ccacccttaa
16301 cagtacatag tacataaagc catttaccgt acatagcaca ttacagtcaa atcccttctc
16361 gtcccatgg atgaccccc tcagataggg gtcccttgac

Fig. 2.

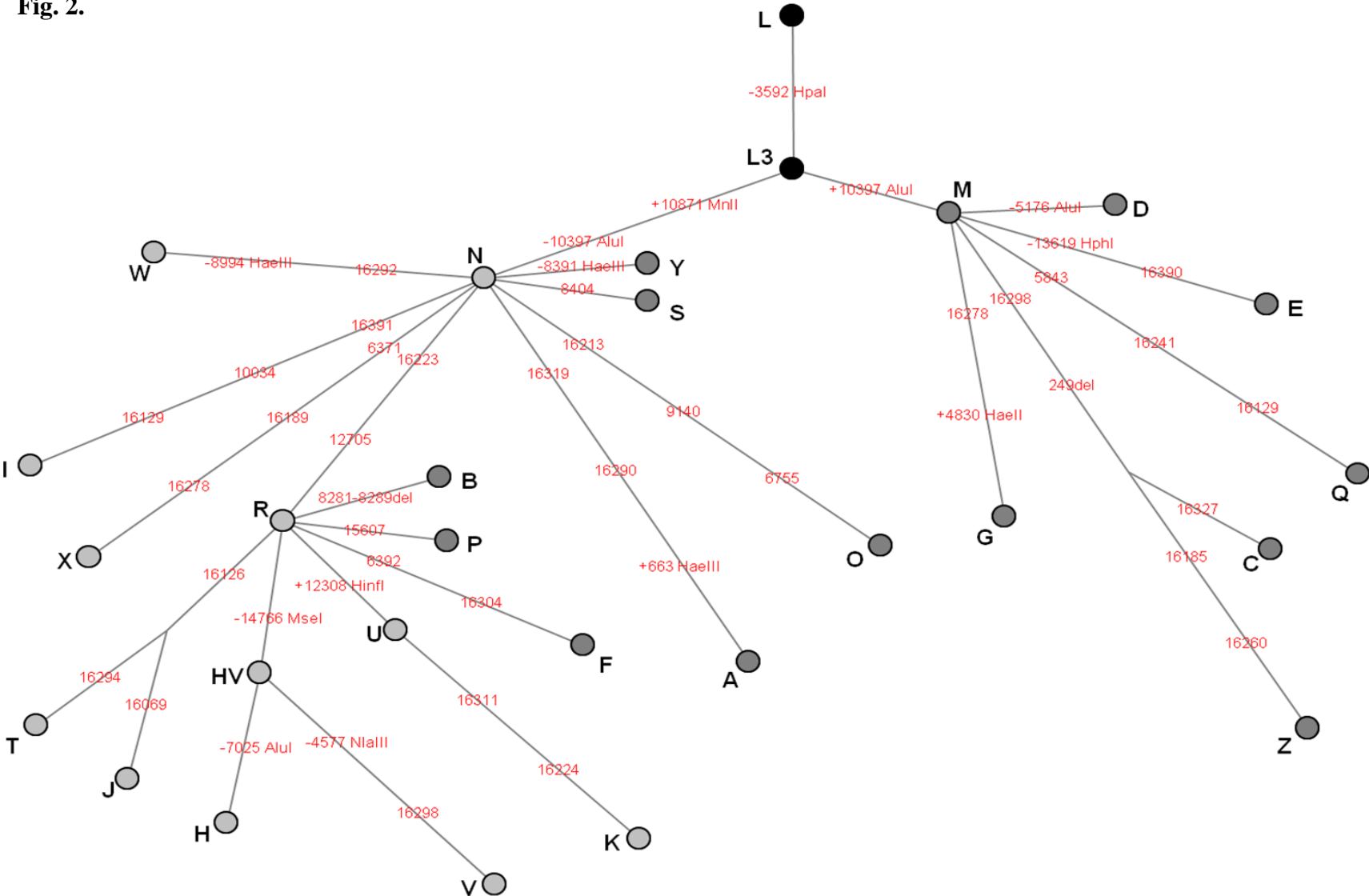


Fig 3.



Fig. 4.

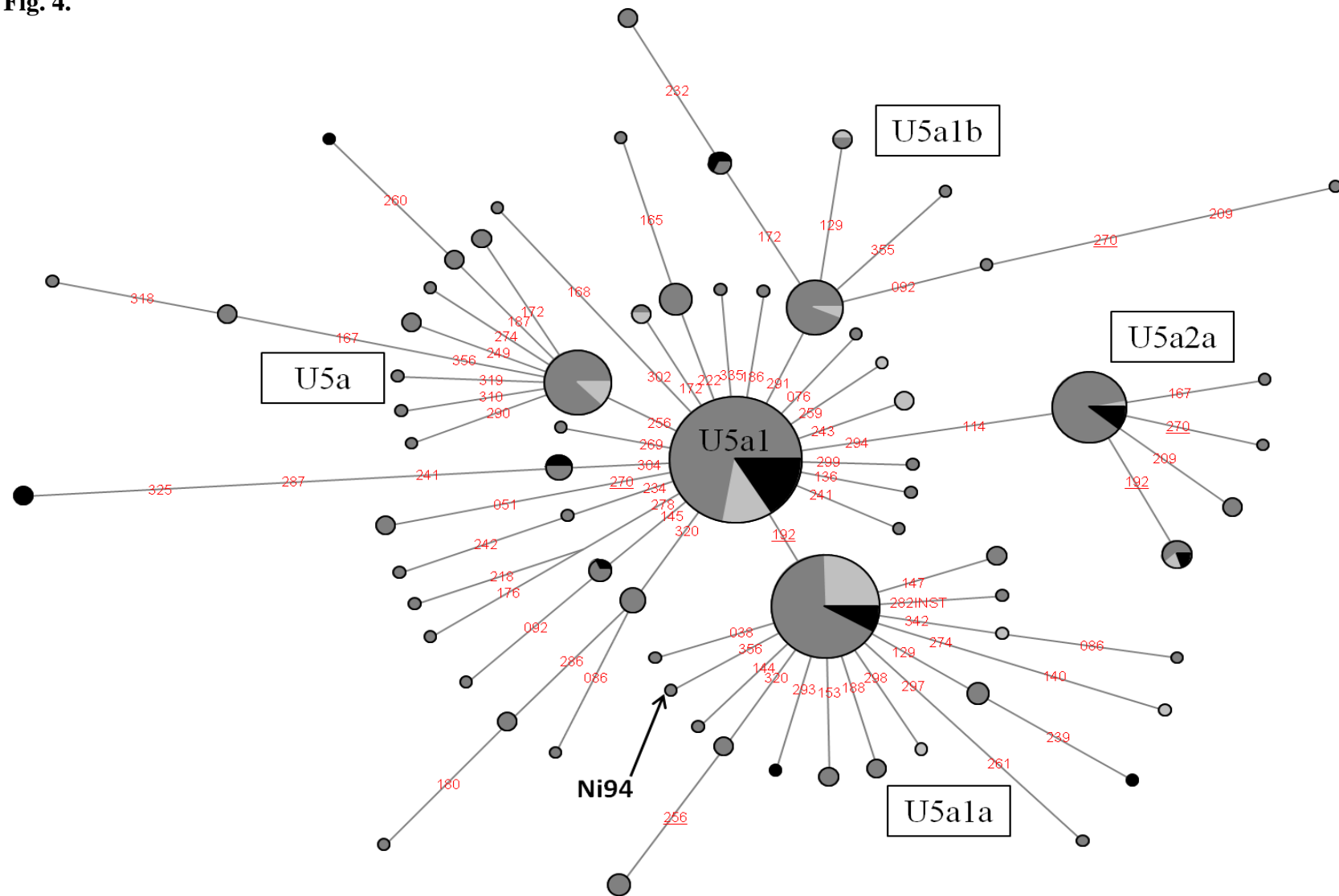


Fig. 5.

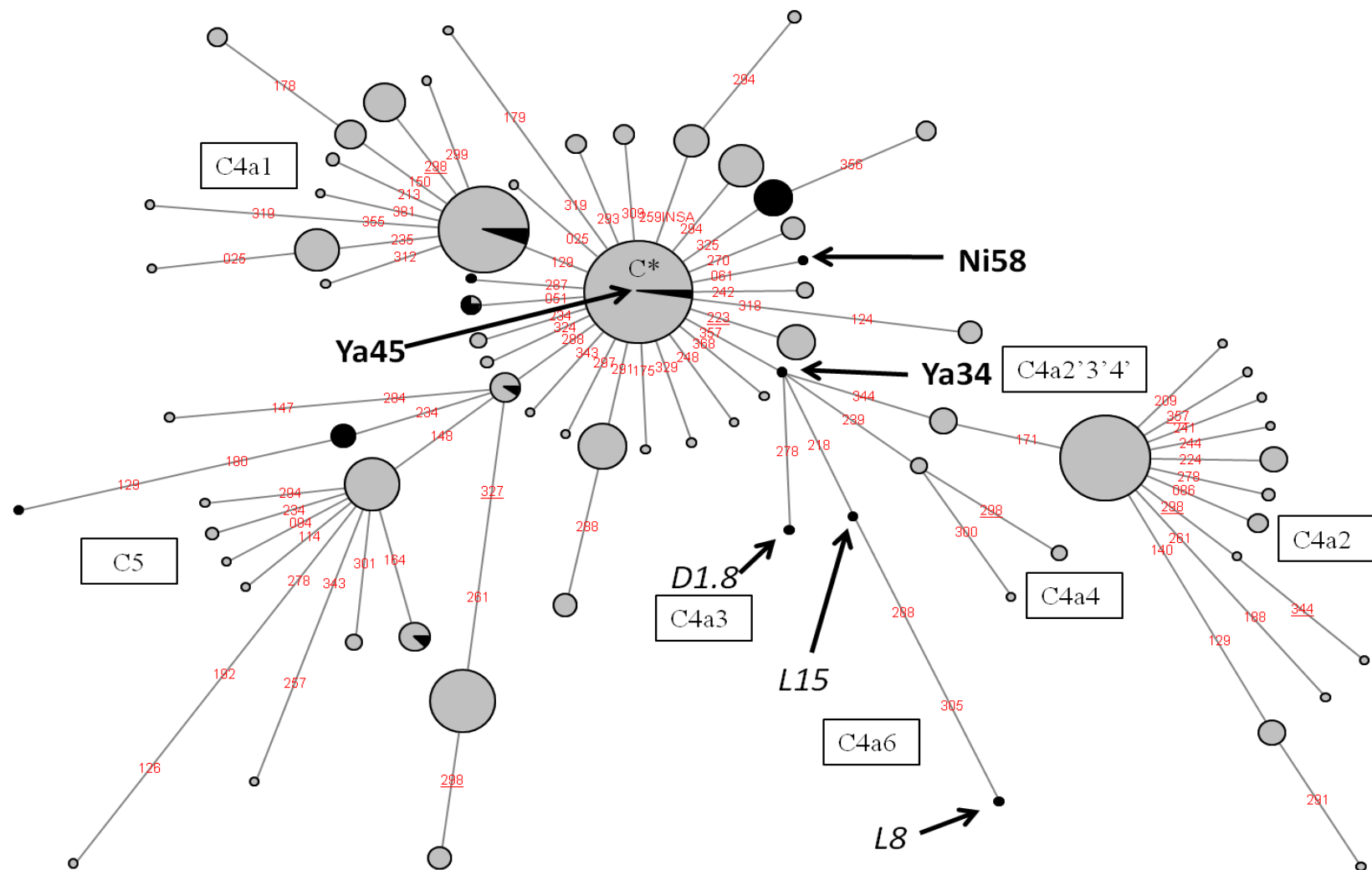


Fig. 6.

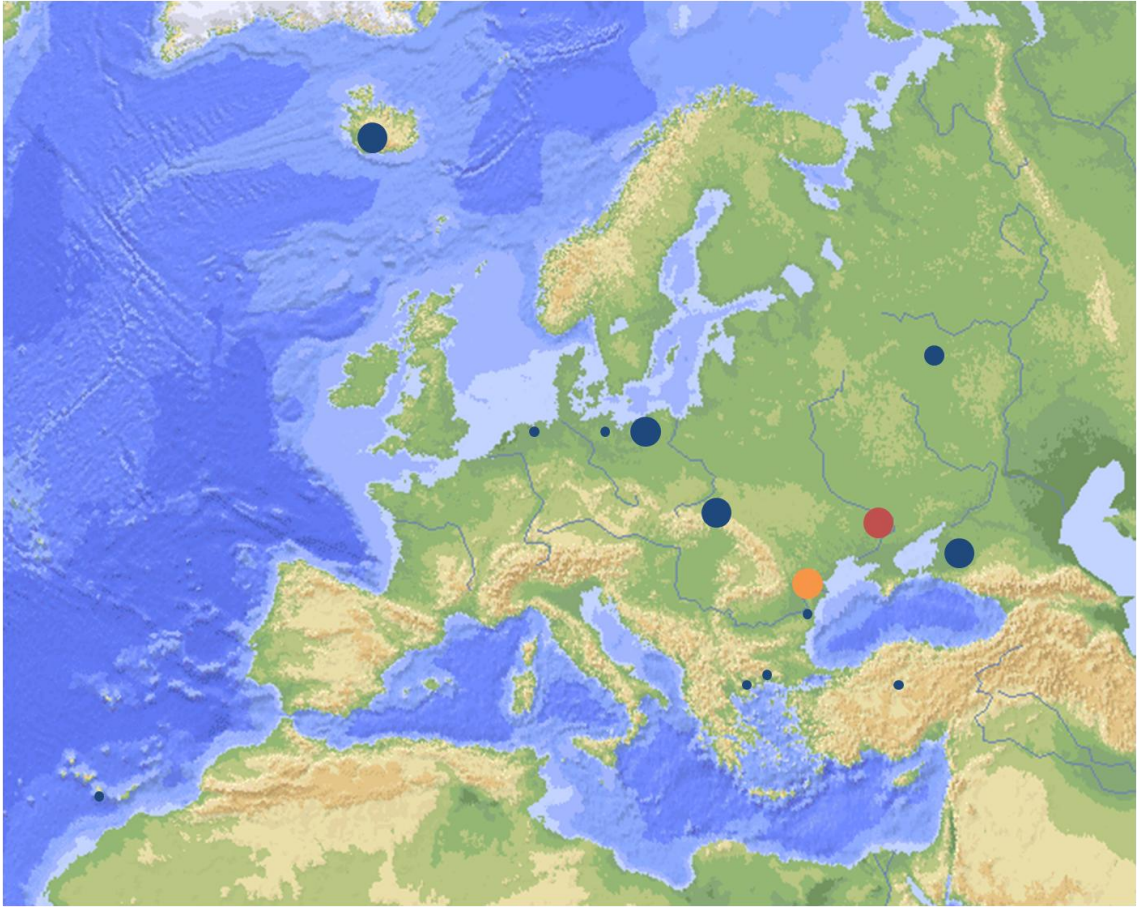


Fig. 7.

