

Mitochondrial DNA
in Ancient Human Populations
of Europe

Clio Der Sarkissian

Australian Centre for Ancient DNA
Ecology and Evolutionary Biology
School of Earth and Environmental Sciences
The University of Adelaide
South Australia

A thesis submitted for the degree of Doctor of Philosophy at The
University of Adelaide

July 2011

TABLE OF CONTENTS

Abstract	10
Thesis declaration.....	11
Acknowledgments	12
General Introduction	14
RECONSTRUCTING PAST HUMAN POPULATION HISTORY USING MODERN MITOCHONDRIAL DNA	15
Mitochondrial DNA: presentation	15
Studying mitochondrial variation	16
Genetic variation.....	16
Phylogenetics and phylogeography	16
Dating using molecular data, and its limits	17
Population genetics.....	19
The coalescent theory and coalescent simulations	21
Reconstructing past genetic history of Europeans using mitochondrial data.....	23
African origins of humans	23
Out-of-Africa	25
Genetic origins of Europeans	26
Limits of modern DNA studies	29
RECONSTRUCTING PAST HUMAN POPULATION HISTORY USING ANCIENT DNA.....	29
First ancient DNA studies	30
The nature of ancient DNA	30
Post-mortem degradation of DNA.....	31
Contamination	33
Authenticating ancient DNA data	34
Survival of DNA.....	38
Applications of ancient DNA	39
Phylogeny and phylogeography of extant and extinct species	40

Domestication processes.....	41
Evolution of biological functions	41
Temporal population dynamics	41
Other applications.....	42
Future directions of ancient DNA studies	42
Ancient human DNA studies.....	43
The problem of authenticity in ancient DNA studies	43
Application of ancient human DNA studies.....	43
Investigation of ancient human mitochondrial DNA diversity in Europe	46
Limitations of ancient and modern-day human DNA	48
PRESENTATION OF THE PROJECT.....	50
General aims of the study	50
Samples and organisation of the thesis.....	50
Methodology.....	53
HVR-I sequencing	53
Typing of 22 coding region SNPs (GenoCore22 reaction)	54
Comparative genetic database	57
Principal Component Analysis	58
Classical multidimensional scaling	58
Haplotype sharing.....	58
Coalescence simulation: Bayesian Serial SimCoal and Approximate Bayesian Computation.....	59
Expected outcomes	59
REFERENCES	60

Chapter One - Ancient Mitochondrial DNA Unravels Complex Population History in North East Europe 80

ABSTRACT	81
INTRODUCTION.....	82
Prehistoric, historical and cultural background of north east Europe	82

Reconstructing the population history in north east Europe using modern-day mitochondrial data	83
Reconstructing the population history in north east Europe using ancient mitochondrial data	84
MATERIALS AND METHODS	86
Sample description and archaeological context	86
Sample preparation and DNA extraction	87
Hypervariable-Region I sequencing	88
Typing of mtDNA coding region single nucleotide polymorphisms (SNPs) by GenoCoRe22 multiplex PCR	89
Cloning	89
Quantitative Real-Time PCR.....	90
Authentication of the mtDNA data.....	91
Populations used in comparative analyses	93
Principal Component Analysis	94
Genetic distance mapping.....	94
Haplotype sharing analysis.....	95
Coalescent simulations	95
RESULTS.....	97
Amplification success and authentication of the ancient DNA data	97
Haplogroup distribution in modern-day populations of north Eurasia.....	100
Mesolithic Uznyi Oleni Ostrov/Popovo compared to modern-day Eurasian populations	100
Bronze Age Bolshoy Oleni Ostrov compared to modern-day Eurasian populations	105
18 th century A.D. Chalmny-Varre compared to modern-day Eurasian populations	106
Comparison among ancient Eurasian populations	107
Testing population history hypothesis using Bayesian Serial SimCoal	108
DISCUSSION.....	111
Genetic discontinuity between prehistoric populations and Saami	111
Siberian influence in the mtDNA gene pool of North East Europeans	112

Genetic link with extant populations of the Volga-Ural	112
Singularity of Uznyi Oleni Ostrov within the Mesolithic diversity	114
Importance of haplogroup U in the Palaeolithic/Mesolithic substratum.....	115
Limited genetic impact of ancient north east European foragers on modern-day populations.....	116
LIST OF SUPPLEMENTARY MATERIALS	117
ACKNOWLEDGEMENTS	118
REFERENCES	118
SUPPLEMENTARY MATERIALS	124

Chapter Two - Mitochondrial Genome Sequencing in Mesolithic North East Europe Unearths a New Sub-clade Within the Broadly Distributed Human Haplogroup C1..... 147

ABSTRACT	148
INTRODUCTION	149
Current phylogeography of the human mitochondrial haplogroup C	149
Origins for the sub-clade C1 in Europe.....	152
Complete mitochondrial genome sequencing in ancient and present-day populations	153
MATERIAL and METHODS	154
DNA extraction	154
Enrichment of ancient human mitochondrial DNA.....	155
Resequencing using the MitoChip v.2.0 (Affymetrix®).....	156
SNP confirmation by direct sequencing and minisequencing.....	157
Authentication of the ancient mtDNA sequence	158
Phylogeny of the C1 clade.....	159
RESULTS/DISCUSSION	159
Use of MitoChip v.2.0 for resequencing of ancient mitochondrial genomes.....	159
Resequencing base call rate.....	162
Phylogeny of the mitochondrial C1 lineage in Mesolithic Uznyi Oleni Ostrov	163
Under-sampling of whole mitochondrial genomes in Eurasia	166

Effect of post-Mesolithic population dynamics	167
A proposed shared genetic history for the Icelandic-specific C1e and the Mesolithic C1f European sub-clades	169
CONCLUSION	170
LIST OF SUPPLEMENTARY MATERIALS	170
ACKNOWLEDGMENTS	170
REFERENCES	171
SUPPLEMENTARY MATERIALS	176

Chapter Three - The Mitochondrial Gene Pool of Scythians of the Rostov Area, Russia: A Melting Pot of Eurasian Influences 181

ABSTRACT	182
INTRODUCTION	183
The Scythians	183
The Bronze Age in the central Eurasian Steppe	183
The Iron Age in the central Eurasian Steppe	184
The origins of Scythians	185
Cultural and genetic homogeneity among ancient nomads of the Eurasian Steppe	185
Genetic diversity of present-day Eurasian populations	186
Ancient DNA from central Eurasia	188
MATERIAL AND METHODS	189
Sample description and archaeological context	189
Sample preparation and DNA extraction	190
Hypervariable-Region I sequencing and coding region GenoCore22 typing	190
Cloning	190
Quantitative Real-Time PCR	190
Authentication of the mtDNA data	191
Ancient populations used in comparative analyses	191
Present-day populations used in comparative analyses	193
Map of haplogroup frequencies	194

Principal Component Analysis (PCA).....	195
Haplotype-based analyses of the mtDNA data.....	195
Fixation index (F_{ST}) calculations and Analysis of the Molecular Variance (AMOVA)	195
Classical Multi-Dimensional Scaling (MDS).....	196
Haplotype-sharing analysis	196
RESULTS.....	197
Success rate for the amplification of authenticated ancient mtDNA	197
Problems associated with the independent replication of one ancient mtDNA haplotype	197
Scythian sample set used in this study	199
Mitochondrial haplogroup structure of the Scythians and comparison with modern-day populations of Eurasia.....	200
Mitochondrial haplogroup structure of Iron Age populations of Eurasia	205
Haplotype-based analyses	205
Informative haplotypes of Scythians shared with Central Asians.....	207
Informative haplotypes of Scythians absent from modern-day Central Asians.....	209
Mitochondrial homogeneity among ancient populations of central Eurasia.....	212
Mitochondrial continuity in eastern Siberian populations.....	213
DISCUSSION.....	215
Mitochondrial makeup of Scythians.....	215
Western mtDNA substratum in Bronze Age nomads of central Eurasia	215
Genetic input from the East into ancient nomadic populations of central Eurasia	216
Homogeneity among Iron Age populations of the central Eurasian Steppe	218
‘Western’ genetic influence in the Bronze Age Tarim Basin.....	220
CONCLUSION	221
LIST OF SUPPLEMENTARY MATERIALS	222
ACKNOWLEDGMENTS	223
REFERENCES	223
SUPPLEMENTARY MATERIALS	230

Chapter Four - Local Mitochondrial Continuity In central Sardinia: Ancient DNA Evidence From The Bronze Age..... 250

ABSTRACT	251
INTRODUCTION	252
Sardinians, European genetic outliers	252
Geographical isolation of Sardinia and initial settlement.....	252
Archaeology and history of Sardinia.....	252
Genetic differentiation between modern-day Sardinians and Europeans.....	253
Mitochondrial differentiation among Sardinians.....	254
Impact of long-term demographic processes on the Sardinian gene pool.....	255
Biology of prehistoric Sardinians: cranial morphology and mtDNA.....	256
MATERIAL AND METHODS	258
Archaeological context.....	258
Ancient DNA extraction.....	259
Hypervariable-Region I sequencing and coding region GenoCore22 typing	260
Cloning	260
Comparative mtDNA dataset of modern-day Sardinians.....	260
Coalescent simulations	260
Phylogenetic network	262
RESULTS/DISCUSSION	262
Amplification success and authentication	262
Comparison with modern-day Sardinian haplogroups	265
Comparison with modern-day Sardinian haplotypes	267
Comparison with other ancient Sardinian populations.....	269
Test for mtDNA continuity between ancient and modern-day central Sardinians	274
CONCLUSION	277
LIST OF SUPPLEMENTARY MATERIALS	277
ACKNOWLEDGEMENTS	278
REFERENCES	278

SUPPLEMENTARY MATERIALS	283
General discussion - Conclusion	287
METHODOLOGY OF ANCIENT HUMAN DNA STUDY	289
Ancient DNA amplification success rates	289
Authenticity of ancient mtDNA data.....	290
Replication of ancient genetic data.....	293
Contextual definition of the archaeological sites sampled for ancient DNA	294
ANALYSIS OF ANCIENT GENETIC DATA	295
Analyses based on haplogroup frequencies.....	296
Analyses based on haplotypic data.....	298
CONTRIBUTION OF ANCIENT DNA TO THE RECONSTRUCTION OF HUMAN POPULATION HISTORY.....	300
Genetic diversity of European Palaeolithic/Mesolithic populations and absence of genetic continuity with Neolithic and present-day populations of Europe	303
Mitochondrial influence of eastern Eurasia in eastern Europe.....	306
Origins of European genetic outliers: Saami and Sardinians	307
The power of ancient DNA	308
CONCLUSION	310
Significance and contribution to knowledge	310
Problems encountered	311
Future direction	311
REFERENCES	317

Abstract

The distribution of human genetic variability is the result of thousand years of human evolutionary and population history. Geographical variation in the non-recombining maternally inherited mitochondrial DNA has been studied in a wide array of modern populations in order to reconstruct the migrations that have participated in the spread of our ancestors on the planet. However, population genetic processes (e.g., replacement, genetic drift) can significantly bias the reconstruction and timing of past migratory and demographic events inferred from the analysis of modern-day marker distributions. This can lead to erroneous interpretations of ancient human population history, a problem that potentially could be circumvented by the direct assessment of genetic diversity in ancient humans. Despite important methodological problems associated with contamination and post-mortem degradation of ancient DNA, mitochondrial data have been previously obtained for a few spatially and temporally diverse European populations. Mitochondrial data revealed additional levels of complexity in the population history of Europeans that had remained unknown from the study of modern populations. This justifies the relevance of broadening the sampling of ancient mitochondrial DNA in both time and space.

This study aims at filling gaps in the knowledge of the genetic history of eastern Europeans and of European genetic outliers, the Saami and the Sardinians. This study presents a significant extension to the knowledge of past human mitochondrial diversity. Ancient remains temporally-sampled from three groups of European populations have been examined: north east Europeans (200 – 8,000 years before present; N = 76), Iron Age Scythians of the Rostov area, Russia (2,300 – 2,600 years before present; N = 16), Bronze Age individuals of central Sardinia, Italy (3,200 – 3,400 years before present; N = 16). The genetic characterisation of these populations principally relied on sequencing of the mitochondrial control region and typing of single nucleotide polymorphisms in the coding region.

Changes in mitochondrial DNA structure were tracked through time by comparing ancient and modern populations of Eurasia. Analysis of haplogroup data included principal component analysis, multidimensional scaling, fixation index computation and genetic distance mapping. Haplotypic data were compared by haplotype sharing analysis, phylogenetic networks, Analysis of the Molecular Variance and coalescent simulations. The sequencing of a whole mitochondrial genome in a north east European Mesolithic individual lead to defining a new branch within the human mitochondrial tree.

This work presents direct evidence that Mesolithic eastern Europeans belonged to the same Palaeolithic/Mesolithic genetic background as central and northern Europeans. It was also shown that prehistoric eastern Europeans were the recipients of multiple migrations from the East in prehistory that had not been previously detected and/or timed on the basis of modern mtDNA data. Ancient DNA also provided insights in the genetic history of European genetic outliers; the Saami, whose ancestral population still remain unidentified, and the Sardinians, whose genetic differentiation is proposed to be the result of mating isolation since at least the Bronze Age. This study demonstrates the power of aDNA to reveal previously unknown population processes in the genetic history of modern Eurasians.

Thesis declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Clio Der Sarkissian and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Clio Der Sarkissian

Date: 03/06/11.

Acknowledgments

This work would not have been possible without the financial, technical and intellectual support of The Genographic Project, in collaboration with the National Geographic Society, IBM and the Waitt Family Foundation.

I am very grateful to my principal supervisor, Alan Cooper, for giving me the opportunity to work at the ACAD. I feel very privileged to have been granted access to valuable ancient human remains and to the outstanding work conditions of the ancient DNA laboratory. I would like to thank him for his support and for showing me what it means to be passionate and enthusiastic about science.

I would like to show my gratitude and admiration to my supervisor, Wolfgang Haak, who has taught me so much about ancient human populations and clean ancient DNA laboratory work. I would like to thank him for his wise advice, but also for his patience and understanding in the course of these three years. Danke vielmals, Wolfgang!

I would like to thank my postgraduate coordinators, Robert Reid and John Jennings for their crucial assistance with administrative issues.

I am deeply indebted to all the collaborators who participated in this work. In particular, I would like to thank Oleg Balanovsky, Elena Balanovska, Valery Zaporozhchenko, Guido Brandt, Kurt Alt, Andrew Clarke, David Soria, Carles Lalueza-Fox, Oscar Ramirez, Robin Skeates, Giuseppina Gradoli, Antonio Torroni, Anna Oliveri, Maria Pala, David Caramelli, Alessandra Modi and Martina Lari for their very precious help.

I am particularly grateful to Jeremy Austin for all his help and for the critical review of my work.

I am heartily thankful to Maria Lekis for her administrative help, her friendship and for making me feel at home in Adelaide.

I thank all my past and present fellow co-workers of the Darling Building and the ACAD for their technical and intellectual support. The ACAD has been a pleasant and friendly (and clean!) work environment thanks to all of them. I will also remember fondly of the ACAD events and I feel very fortunate to have developed friendships with so many of my colleagues.

I would like to address very special thanks to my friends in Adelaide: Bastien, Camille, Damien, Doreen, Emma, Gaynor, Grant, Jessica, Julien, Kimiko, Manue, Marjorie, Valentin, and Virginie who have been a real family to me.

I would like to thank Claire C., Claire M., Fanny, Magali, and Thomas for their patience, their support, their advice, and for remaining my friends despite the distance.

Thanks also to my aunt and cousins for their great support during the very last stages of the writing of this thesis.

I owe my deepest gratitude to my family and in particular to my parents Jutta and Yves and my brother Antoine.

À Claire,

General Introduction

Abbreviations: A, adenine; aDNA, ancient DNA; B.C. , Before Christ; bp, base pair; C, cytosine; CRS, Cambridge reference sequence; DNA; deoxyribonucleic acid; ddNTP, dideoxyribonucleoside triphosphate; G, guanine; LBK, Linienbadnkeramik culture; LGM, Last Glacial Maximum; MDS, multidimensionnal scaling ; MRCA, most recent common ancestor; mtDNA, mitochondrial DNA; Ne, effective population size; np, nucleotide position; PCR, Polymerase chain reaction; PTB, N-phenacylthiazolium bromide; rCRS, revised Cambridge reference sequence; SBE, Single Base Extension; T, thymine; TMRCA, time of the most recent common ancestor; UNG, Uracil-N-glycosylase; UV, ultraviolet; yBP, years Before Present.

RECONSTRUCTING PAST HUMAN POPULATION HISTORY USING MODERN MITOCHONDRIAL DNA

Mitochondrial DNA (mtDNA) has emerged as a marker of choice to investigate the genetic history of human populations. In this section, I review the properties of mtDNA, how mtDNA diversity in modern populations has been used to retrace important past human migrations, and finally, I present the drawbacks of modern mtDNA when applied to the reconstruction of the genetic history of humans.

Mitochondrial DNA: presentation

Mitochondria are organelles found in the cytoplasm of eukaryotic cells and play a crucial role in the respiration of the cell (Bandelt et al., 2006a). Mitochondria are thought to have originated as free-living bacteria that parasited proto-eukaryotic cells ~1.5 billion years ago and have since remained in an endosymbiotic relationship inside eukaryotic cells (Margulis, 1981). The mitochondria preserves remnants of the original bacterial genome coding for key aspects of the mitochondrial machinery, but over the course of evolution, most mitochondrial genes have been transferred to the nucleus. The extent of these nuclear insertions was estimated to represent at least 400,000 base pairs (bp) in the human genome (Bensasson et al., 2001). The number of mitochondria varies considerably according to cell type, but in humans it is estimated to be around 3,000 per cell on average, with around 2 genomes per mitochondria. Within most animals, including humans, the mitochondrial genome follows a strictly maternal inheritance, i.e., from mother to offspring, which presumably relates to mechanisms that eliminate paternal mtDNA within the zygote (Schwartz & Vissing, 2002).

The first complete human mitochondrial genome sequence (~16,569 bp) was obtained in 1981 (Anderson et al., 1981). This sequence was subsequently used as a reference sequence for the study of human mtDNA and was termed the 'Cambridge reference sequence' (CRS). A modified version of the CRS, in which eleven sequencing errors were corrected, was published in 1999 (Andrews et al., 1999) and renamed 'revised Cambridge reference sequence' (rCRS). The human mitochondrial genome is composed of a coding and a non-coding region. The coding region contains 37 genes encoding for 13 proteins, 22 transfer RNAs and two ribosomal RNAs, which are involved in the synthesis of proteins that participate in cellular respiration (Anderson et al., 1981). The non-coding region, also called 'control region', contains 'hypervariable

regions I and II' (HVR-I and HVR-II). The hypervariable regions owe their name to the observation of significantly more genetic variation in these regions than in the coding region of the mtDNA genome (Stoneking et al., 1991). HVR-I (440 bp) ranges from nucleotide positions (np) 16129 to 16569 (according to the rCRS) and HVR-II (574 bp) ranges from np 00001 to 00574.

Studying mitochondrial variation

Genetic variation

Genetic variation is generated by a process called 'mutation', which describes the substitution of a base by another, or the loss (deletion) or addition of a base (insertion), leading to detectable changes between DNA sequences (e.g., reviewed in Jobling, Hurles & Tyler-Smith, 2004). There are two types of substitutions: transitions and transversions. Transitions are substitutions from one purine base to the other (i.e., A \leftrightarrow G) or from one pyrimidine base to the other (i.e., C \leftrightarrow T). Transversions are any other kind of substitution (i.e., changes from a purine to a pyrimidine), and are generally observed to occur far less often than transitions, potentially due to the conformational disruption caused to the double stranded helical structure.

The processes by which substitutions appear in DNA sequences have been described by a range of mathematical models that take into account different rates for transitions and transversions, nucleotide changes, and the base composition of the DNA sequence. These models vary in their levels of complexity, from the most simple, the Jukes-Cantor model (all rates and all base frequencies equal; Jukes-Cantor, 1969) to the General Time Reversible model (different rates for all nucleotide changes and all base frequencies different; Tavaré, 1986).

Phylogenetics and phylogeography

The history of mutational events can be reconstructed and visualised by an evolutionary tree linking the DNA sequences observed in a population or a species. This evolutionary reconstruction is termed a phylogenetic tree (see e.g., Hall, 2001). If mutations are assumed to accumulate at a relatively constant rate over time, the number of mutational events that are necessary to link two sequences is related to the time that has passed since these sequences last shared a common ancestor. In a phylogenetic tree, all the sequences arising from the same common ancestor are

designated as belonging to the same clade (see e.g., Hall, 2001). The geographical distribution of clades within a phylogeny is termed phylogeography, and records the evolutionary history and movement of a population or species (Avise et al., 1987; Templeton et al., 1995).

Dating using molecular data, and its limits

The ‘molecular clock hypothesis’ proposes that changes in a DNA sequence caused by nucleotide substitution appear at a rate that is roughly constant over time and among lineages of a phylogenetic tree (Zuckerkandl & Pauling, 1965; Jukes & Cantor, 1969). An average mutation rate can be estimated provided that the genetic diversity within a phylogenetic group is assessed and that an event within the tree - serving as calibration point - is independently dated. Calibration points are generally estimated from the fossil record, and used to define the minimal time of divergence between two taxa. Information from the archaeological (e.g., date of earliest archaeological evidence for human presence) or biogeographical record (e.g., date of apparition of land bridges between continents) can also be used as calibration points. Once a mutation rate is calculated, it can be used to date other divergence events within the phylogeny, and the time of the most recent common ancestor (TMRCA) for the group (e.g., reviewed in Jobling, Hurles & Tyler-Smith, 2004).

Empirical datasets rapidly provided evidence of deviation from the simple model of the molecular clock (constant rate through time). Substitution rates have been found to be variable between chromosomal positions in the genome (e.g., Sharp et al., 1989), sites in a sequence (Wakeley, 1993) and organisms (e.g., Douzery et al., 2003).

Rate heterogeneity among sites was taken into consideration in models of sequence evolution. For example, it is recognised that third codon positions in protein-coding genes are freer to vary than the first or second codon. This is explained by redundancy of the genetic code that allows the same amino-acid to be encoded by several DNA codons differing by the nucleotide at the third codon position (redundancy of the genetic code; Goldman & Yang, 1994). Mutational ‘hotspots’ represent an extreme case of rate heterogeneity where small proportion of sites evolving at a rate significantly higher than the majority of the sites in a sequence. A well-studied example of mutational hot spots is those found in human mtDNA HVR,

which have been identified by diverse methods (Meyer, 1999; Bandelt et al., 2006b; Rosset et al., 2008).

Models have been developed that allow for rate heterogeneity among sites in a DNA sequence, generally by defining a set of rate classes that each position can fall into (e.g., Hasegawa, 1985; Nei & Gojobori, 1986; Yang, 1993; Tamura & Nei, 1993; Yang, 1996). These models have been widely accepted and implemented in commonly used programs for phylogenetic reconstruction.

The substitution rate in human mtDNA has been investigated in a wide range of studies that have used diverse datasets and methods, representing different evolutionary timescales: within family trees (i.e., pedigrees; e.g, Parsons et al., 1997; Sigurðardóttir et al., 2000; Heyer et al., 2001) or across the human phylogeny (e.g, for HVR in Forster et al., 1996; the coding region in Mishmar et al., 2003; and synonymous substitutions in protein-coding genes Kivisild et al., 2006). Substitution rates were also estimated using different calibration points such as the divergence between humans and chimpanzees (inter-species; e.g., Mishmar et al., 2003; Kivisild et al., 2006), or dates for human expansions associated with particular clades and estimated on the basis of archaeological/climatic evidence (intra-specific), e.g., the post-glacial expansion of European haplogroups in Europe (11,000 – 25,000 years Before Present, yBP) and the colonisation of Australia and Melanesia (~40,000 - 45,000 yBP; Endicott & Ho, 2008). The problem is that the rates reported in these studies vary significantly. In particular, rates calculated from pedigrees have always been found to be significantly higher than those calculated using deeper evolutionary timescales (e.g., human-chimpanzee divergence). Discrepancies in the estimates of the human mtDNA substitution rate were shown to create major differences in estimates of the timing of demographic events. This was exemplified by the comparison of dates for the colonisation of the Americas that were obtained using the same datasets but different substitution rates (Ho & Endicott, 2008). Coalescent age of American-specific clades indicated dates for the colonisation of the America ~20,730 yBP using the rate proposed by Mishmar et al., 2003, but only ~13,960 yBP with the rate published in Endicott & Ho, 2008.

Problems associated with the estimation of the human mtDNA substitution rate have been reviewed in Endicott et al., 2009a, and include differences in the substitution rate among species when using human-chimpanzee divergence as a calibration, and error on the date used for calibration, e.g., divergence between

humans and chimpanzees. The dating approach that seems the most accepted at present is the one published in Soares et al., 2009 that took into consideration a large dataset of mtDNA complete genome sequences (more than 2,000), a revised date for the human-chimpanzee divergence and a modest effect of purifying selection in the mtDNA coding region (i.e., elimination of slightly deleterious mutations). However, the problems associated with the estimation of the substitution rate and the dating of divergence events using modern mtDNA data should be kept in mind when considering dates that were calculated on the basis of molecular data.

Population genetics

The study of genetic diversity in a population and its changes in time is the subject of population genetics. Population history is reconstructed by estimating the contribution of processes involved in the change in frequencies of genetic variants through time. These processes are genetic drift, selection and migration.

(1) Genetic drift

The random variation in frequencies of genetic variants in a population due to their random sampling from one generation to another is called 'genetic drift' (Wright, 1931). Genetic drift can cause the reduction of the genetic diversity in population either by fixation (i.e., the increase in frequency of a genetic variant until it becomes fixed in the population) or by elimination of a genetic variant (i.e., decrease in frequency of a genetic variant until it becomes extinct). Genetic drift was described in the Wright-Fisher model (Fisher, 1930; Wright, 1931) that estimates the probability of obtaining a genetic variant for one generation based on its frequency in the previous generation. The Wright-Fisher model is based on several assumptions: non-overlapping generations, constant population sizes through time and random mating (for a review, see e.g., Charlesworth, 2009).

Real populations are usually not idealised Wright-Fisher populations. The concept of an effective population size (N_e) was introduced in order to compare the effect of genetic drift on populations (Wright, 1931). When studying a real population characterised by a census size, N_e is the size of an idealised Wright-Fisher population that undergoes the same amount of genetic drift as the population under study. Hence, the N_e of a population allows the impact of genetic drift on this population to be estimated, and depends on the size of the population. For example, the effect of genetic drift is larger in populations characterised by smaller N_e .

‘Bottlenecks’ and ‘founder effects’ are processes that also have major impacts on genetic drift. A bottleneck describes the reduction of genetic diversity in a population due the loss of genetic variants during a reduction of its size (e.g., caused by shrinking resources, habitat, etc). A founder effect is observed during a colonisation event, i.e., the movement of a (small) subset of a population into a previously unoccupied territory. This means that the colonising population will carry only a portion of the original genetic diversity of the source population. As a result, the new founder population has a reduced genetic diversity, and the ‘founder effect’ is a special case of genetic drift.

(2) Selection

The concept of selection has been defined by Charles Darwin (1859) and then enunciated in Fisher’s theorem of natural selection (Fisher, 1930). Selection describes how individuals in a population can have different contributions to the next generation as a function of their ability to survive (until they can reproduce) and reproduce. For example, the carriers of a genetic variant that would provide a selective advantage, or an ‘adaptation’ to specific environmental conditions, will tend to survive until they reach reproductive age in larger numbers, thus giving rise to more numerous descendents. The offspring inherits the advantageous genetic variant and as a consequence, the genetic variant represents a larger portion of the gene pool in this generation. If these environmental conditions are constant through time, this process is repeated in each generation. Eventually, selection of this genetic variant leads to an increase in the frequency of traits that are advantageous with regard to the given environment. Selection is materialised through diverse forms that can be sexual selection, fertility, fecundity or an increase in viability or mortality. The effect of selection is larger for variants involved in a function (e.g., in genes or regulatory sequences). DNA associated with no particular function is indeed freer to vary, or is selectively less constrained, because any genetic variation would have little or no consequence on the ability of the organism to survive or reproduce. In population genetics studies that aim at reconstructing past population history and not at examining selection itself, the chosen genetic markers are required to be ‘neutral’, i.e., free from selection like mtDNA HVR.

(3) Migration

Migration is a process in which the frequencies of genetic variants change through time under the action of sub-populations (or demes) within an encompassing

‘meta-population’. Individuals from a ‘source’ population (origin of the migration) add their genetic variants to a ‘sink’ population (destination of the migration), thus altering the frequencies of genetic variants in the sink population.

The genetic impact of the migration depends, first, on the proportion of the sink population represented by newcomers from the source population, and second, on the level of genetic differentiation (i.e., percentage of shared genetic variants) between the two populations. A migration has an evolutionary impact (gene flow) on the sink population only if newcomers contribute to the gene pool of the subsequent generations (i.e., reproduce within the sink population). Gene flow is a factor that prevents genetic differentiation between populations.

The coalescent theory and coalescent simulations

In order to test hypotheses about the impact of evolutionary forces (e.g., genetic drift) in the course of the demographic history of the population of interest, a statistical framework taking into account the mutational processes that gave rise to the observed diversity is needed. A statistical framework that allows lineage coalescence, i.e., merging of lineages backwards through time, and mutational history to be taken into consideration in a genealogy is the coalescent model (reviewed in Rosenberg & Nordborg, 2002). The coalescent model was first described in the 1980s and treated by Kingman (Kingman, 1982; Hudson, 1983; Tajima, 1983; Kingman, 2000) and is a stochastic model that extends the classical population genetics models used to analyse DNA data.

The principle of the coalescent model can be described as follows. In the absence of selection, at each generation of a genealogy, the sampled lineages can be viewed as randomly picking their parent lineages, going backwards in time. When the same parent lineage is picked by two lineages, the lineages are said to coalesce. When all lineages coalesce into a single lineage, this lineage represents the Most Recent Common Ancestor (MRCA) of the sample under study. The number of lineages that pick their parent lineages and the size of the population both impact the rate at which the lineages coalesce: the more lineages, the faster the rate and the more the number of parents to choose from, the slower the rate. The rate of coalescence can also be impacted by the age structure, skewed sex ratios and reproductive success, whereas the shape of the genealogical trees can be altered by changes in population size and population structure (Nordborg, 2001). The coalescent allows the evolution of the

populations to be simulated backward in time until all lineages coalesce, then, mutations are added along the branches of the newly generated tree. On the basis of this tree, the population parameters of the stochastic genealogical process are estimated. Coalescent simulations in model-based approaches are the more widespread use of the coalescent theory (Hudson, 1990).

The objective of coalescent simulations is to determine, which demographic model among a range of pre-defined models best explains the observed genetic data or whether a proposed demographic model can be rejected considering the observed data. The first step of coalescent simulations is to define the demographic models to be tested, e.g., constant population size, population expansion or migration. Coalescent simulations will generate a large number of different genealogical trees based on random inclusion of mutations and stochastic repetitions of the evolutionary processes as defined by the demographic model. For each tree, population summary statistics (i.e., describing the genetic diversity in a population) are calculated resulting in a distribution of simulated statistics. These distributions are compared to the observed population statistics calculated from the population(s) under study. When different demographic models are simulated, the 'fit' of each model to the observed data can be assessed in order to determine which model is the more likely to explain the observed pattern from the population under study. The program SimCoal 2.0 (Laval & Excoffier, 2004) can be used to perform coalescent simulations using genetic data sampled from modern populations. The program Bayesian Serial SimCoal (Anderson et al., 2005) was later adapted from SimCoal 1.0 (Excoffier, Novembre & Schneider, 2000) in order to allow genetic data to be sampled in time, as for ancient DNA sequences.

Bayesian Serial SimCoal allows for population parameters of the tested demographic models (e.g., growth rate of an expanding population or proportion of migrants into a sink population) to be drawn from a prior distribution in cases where no a priori knowledge of these parameters is available. The output of coalescent simulations, including those using Bayesian Serial SimCoal, can be analysed within an Approximate Bayesian Computation statistical framework (ABC; Beaumont et al., 2002). The ABC algorithm calculates the Euclidean distance between the observed and simulated population statistics (e.g., haplotype diversity, fixation indexes) that have been generated for all population parameters drawn from the distribution. Parameters yielding small Euclidean distances between simulated and observed

statistics are more likely to characterise the ‘true’ population history. A distribution of the most likely population parameters can then be determined. This approach can be repeated in order to refine the estimation of population parameters. However, there are limits to the application of these approaches to ancient DNA data. While the coalescent simulation programs are able to estimate population parameter distributions correctly from the simulated genealogies according to the demographic model tested, the same population parameters estimated from an ancient sample set are probably not representative of the true value for the population under study due to the stochasticity of sampling. Another critic of the model-based approach is that only a subset of models, not representative of the real model can be tested (Templeton et al., 2009; Templeton et al., 2010).

Reconstructing past genetic history of Europeans using mitochondrial data

As the mtDNA genome is haploid, non-recombining and exclusively maternally-inherited it has emerged as a marker of choice for the investigation of hominin evolutionary history (i.e., of humans, chimpanzees and their extinct cousins; Richards & Macaulay, 2001). Mitochondrial DNA has been widely used to reconstruct human phylogeny, phylogeography and population history (e.g., migrations). In particular, the highly polymorphic HVR-I and HVR-II portions of the mtDNA genome have been found to be very informative and have been characterised to a great extent in human populations (Stoneking et al., 1991). The global sampling of current human mtDNA diversity (both at the HVR and complete mtDNA genome levels) means that the human mtDNA phylogeny has been extensively described compared to other species. Current patterns of geographical distribution of human mtDNA diversity have been used to reconstruct the main migrations in the spread of humans around the globe (Richards, 2003).

African origins of humans

The initial studies of geographical structure in human mtDNA diversity relied on the analysis of restriction fragment length polymorphisms (RFLP) in globally sampled individuals (Brown et al., 1980). This study calculated a TMRCA for humans of ~180,000 yBP, and this controversial result was later supported by a landmark study dubbed ‘mitochondrial Eve’ (Cann et al., 1987). The Cann et al. 1987 analysis used higher molecular resolution restriction mapping and larger sample-sizes to

estimate that the human mtDNA most recent common ancestor lived around 140,000 to 290,000 yBP. An important result of this study was that the mitochondrial Eve was most likely African. This hypothesis was based on the fact that the root of the tree constructed by parsimony phylogenetic analysis separated an African-specific branch from the rest of the tree, which encompassed lineages of all ethnicities. The African origin was also supported by the fact that the largest inter-population differences were observed between Africans and other populations. Africa was later confirmed as the most likely source of the human mitochondrial variability in a study that used sequences of a chimpanzee (*Pan troglodytes*), the closest living relatives of humans, to root the human mitochondrial tree constructed using HVR sequences (Vigilant et al., 1991). A time of the most recent common ancestor (TMRCA) for humans was calculated ~166,000 - 249,000 yBP, based on the divergence time between humans and chimpanzees of 4 - 6 million yBP as estimated from the fossil record.

Another important result of the Cann et al., 1987 study was the calculation of TMRCA for all non-African humans, also referred to as the 'Recent Out-of-Africa' event, ~62,000 – 225,000 years BP. This date estimate tends to support a relatively short timescale for the common origin of all humans on the planet and for their spread out of the African continent. The 'Out-of Africa' model proposes that humans originated in Africa, most likely in East Africa, where they evolved before a subset of the human population left Africa and colonised the rest of the world. The 'Out-of Africa' model was formally established in Stringer & Andrews, 1988. This model contrasts with the 'Multi-regional model', which proposes that humans independently evolved from archaic humans into anatomically modern humans at a local scale all around the world (Wolpoff, 1988). It is also important to note that these early genetic studies (Brown et al., 1980; Cann et al., 1987; Vigilant et al., 1991) verify previous conclusions drawn from the analysis of protein variability showing that the genetic diversity observed between ethnicities is smaller than the genetic diversity observed within ethnicities, thus implying that most of the genetic variation is shared among populations (e.g., Mourant et al., 1978; Lewontin, 1972).

As sequencing technology became more widely available, the study of mtDNA variation has increasingly relied on HVR sequencing, sometimes combined with RFLP or complete mtDNA genomes (e.g., Macaulay et al., 1999; Wallace et al., 1999; Ingman et al., 2000; Maca-Meyer et al., 2001; Herrnstadt et al., 2002; Mishmar et al., 2003; Metspalu et al., 2004). All studies of the distribution of human mtDNA

variability at the global level have so far supported the African of humans and no genetic evidence has been found for the alternative 'Multi-regional model'. The increasing number of mtDNA studies from human populations around the world meanwhile led to a refined and detailed reconstruction of the human mtDNA phylogeny. This phylogeny confirmed the split between the 'African' and the 'non-African' groups. Clades within the human mtDNA tree, called 'haplogroups', were named according to nomenclature first introduced in a study that defined four mtDNA haplogroups A, B, C and D found in present-day Native Americans on the basis of RFLPs (Torroni et al., 1993). The mtDNA lineages thought to have originated in Africa were called 'L', and all other lineages thought to have emerged after humans initially left Africa were classified within 'L3' into two non-African macro-haplogroups 'M' and 'N' (Chen et al., 1995; Passarino et al., 1996; Macaulay et al., 1999).

Out-of-Africa

Numerous genetic studies have investigated the timing of the spread of humans out of Africa. Calculation of the coalescent ages of macro-haplogroups M and N using the Mishmar rate (Mishmar et al., 2003) yielded ages of ~60,000 -70,000 yBP and favoured a relatively slow single expansion to the Near East and India, followed by a rapid migration to South East Asia, and eventually Australia, along a coastal route (Macaulay et al., 2005). Another date of ~40,000 and 50,000 yBP for the 'Out-of-Africa' was calculated using a mutation rate based on synonymous transitions (Kivisild et al., 2006) and suggested a rapid spread of humans out of Africa, as the first archaeological evidence for human presence in Australia and New Guinea are dates ~45,000 yBP (reviewed in O'Connell & Allen, 2004). Recent recalculation of the coalescent ages confirmed the dates calculated by Macaulay et al., 2005 with L3 dated ~70,000 yBP, and M and N, dated ~50,000 – 70,000 yBP (Soares et al., 2009).

Sampling of mtDNA data at the global scale has revealed pronounced patterns of geographical distribution in haplogroups belonging to the non-African macro-haplogroups M and N (Torroni et al., 1993; Wallace et al., 1999; Macaulay et al., 1999; Ingman et al., 2000; Maca-Meyer et al., 2001; Herrnstadt et al., 2002; Mishmar et al., 2003; Metspalu et al., 2004). Macro-haplogroup M is composed of haplogroups C, D, E, G, Q and Z, which are currently found in Asia, Oceania and the Americas. Macro-haplogroup N is structured in two branches. The first branch gives rise to haplogroup R, which is itself split into haplogroups B, F, J, P, T, as well as haplogroup

R0 (including HV, H, and V) and haplogroup UK (including haplogroups U and K). The non-R clades within N, termed N*, are haplogroups A, N1, O, S, X, and Y, with haplogroup N1 containing haplogroups N1a, N1b, and I. Haplogroup N includes all the haplogroups that are the most frequent in Europe: haplogroups HV, H, I, J, K, T, U, V, W, and X. The other non-European specific haplogroups are found in present-day populations of Asia, Oceania and the Americas. The geographical patterns of distribution for the non-African haplogroups were used to reconstruct the migrations of humans out of Africa. The proposed scenario for the deep origins of European mtDNA lineages is that precursors of most European lineages were part of a migration from Africa to the Levant (i.e., the area bordering on the eastern Mediterranean Sea from Turkey to Egypt) ~50,000 – 70,000 yBP where they remained until taking part in the colonisation of Europe (reviewed in Jobling, Hurles and Tyler-Smith, 2004; Richards et al., 2006).

Genetic origins of Europeans

The archaeological record dates the first colonisation of western Europe by anatomically modern humans ~40,000 - 42,000 yBP, with earlier dates found for southern Europe (Mellars, 2006). The first anatomically modern human populations of Europe were characterised by a nomadic lifestyle and a reliance on fishing, hunting, and gathering as food sources (reviewed in Soares, 2010). The closest relatives of humans, the Neanderthals, who showed human-like but robust anatomical features, had reached Europe earlier, ~350,000 - 600,000 yBP. Anatomically modern humans and Neanderthals hence cohabited for ~10,000 years before Neanderthals went extinct ~30,000 yBP (Bischoff et al., 2003; Finlayson et al., 2006; Harvati, 2007).

Investigation of the genetic origins of Europeans has focused on the question of the relative contribution to the gene pool of present-day Europeans of three main events in prehistory (e.g., reviewed in Soares et al., 2010): the initial Upper Palaeolithic settlement of Europe (~40,000 yBP) by hunter-gatherers, their recolonisation of Europe from southern European refugia ~10,000 - 15,000 yBP after the Last Glacial Maximum (LGM, ~19,500 – 25,000 yBP), and the potential arrival of Neolithic early farmers from the Near East during the Neolithic transition (i.e., the transition from a foraging to an agricultural lifestyle; ~10,000 yBP).

Early studies of autosomal markers observed a clear South East to North West gradient in the distribution of the genetic data in Europe and the Near East

(Ammerman & Cavalli-Sforza, 1984; Sokal et al., 1991, Chikhi et al., 1998). They interpreted this gradient as a signal of an important migration of early farmers from the Near East into Europe during the Neolithic transition. These results supported a large contribution of descendants of Neolithic farmers in the present-day European population and, hence, the model of ‘demic diffusion’, according to which the Neolithic transition involved significant migrations from the Near East (Ammerman & Cavalli-Sforza, 1984). However, it is difficult to discriminate the genetic effects of the Upper Palaeolithic colonisation, the post-glacial recolonisation and the Neolithic transition in Europe since they followed similar South East trajectories into Europe. Superimposed migratory events could have created a South East to North West gradient in the distribution of the genetic diversity. As no date could be associated with the gradient observed for the autosomal markers, it was proposed that they in fact represent the initial colonisation of Europe during the Upper Palaeolithic rather than the Neolithic expansion (Richards et al., 1996).

Mitochondrial DNA was used to address this issue because it theoretically allows the dating of sub-sets of genetic diversity through the calculation of coalescent age of haplogroups, or sub-haplogroups (Richards et al., 1996). Studies of the human mtDNA diversity defined nine main European haplogroups: haplogroups H, I, J, K, T, U, V, W, and X (Richards et al., 1996; Richards et al., 1998; Torroni et al., 1998; Richards et al., 2000; Torroni et al., 2001). In their investigation of the impact of the Neolithic transition on the modern-day European gene pool, Richards et al., 2000 used a ‘founder analysis’ to identify founder types in mtDNA haplogroups by comparing the genetic diversity in a source population (Near East) to the derived diversity in the sink population (Europe). In combination with coalescent age calculations, the founder analysis identified sub-clades likely to have reached Europe during the initial colonisation of Europe during the Upper Palaeolithic (haplogroup U in the Early Upper Palaeolithic: ~45,000 – 55,000 yBP, haplogroups HV, I and U4 in the Middle Upper Palaeolithic: ~17,000 - 38,000 yBP, haplogroups H, K and T2 in the Late Upper Palaeolithic: ~8,000 - 17,000 yBP), and those likely to derive from the genetic input of Neolithic migrants from the Near East (haplogroups J1a and T, ~6,000 – 13,000 yBP; Richards et al., 2000). Based on these estimates, Richards et al., 2000 concluded that ~80% of the present-day European mtDNA had arisen from Upper-Palaeolithic hunter-gatherers versus ~20% from migrating Near East farmers of the Neolithic (Richards et al., 2000). These results supported the model of ‘cultural diffusion’ for

the Neolithic, i.e., transfer of agricultural lifestyle and technologies involving little migration from the Near East.

Sampling of the European mtDNA diversity also detected genetic signals that were associated with the recolonisation of Europe after the LGM (19,500 – 25,000 yBP). During the LGM, the density of human occupation increased in southern refugia in south west Europe, along the Mediterranean Sea, in the Balkans, the Levant and the east European Plain. European populations are thought to have then re-expanded as climatic conditions improved after the LGM (Dolukhanov, 1993; Gamble et al, 2004; Gamble et al., 2005; reviewed in Soares et al., 2009). The distribution of haplogroups H1, H3, U5b1, and V along a South West to North East gradient was proposed to be the result of post-glacial re-expansion from the Franco-Iberian refugium to the rest of Europe, which was dated to the Post-Glacial ~10,000 yBP (Torroni et al., 1998; Torroni et al., 2001; Achilli et al., 2004; Tambets et al., 2004; Pereira et al., 2005).

Coalescent ages for European haplogroups presented in these studies were re-estimated with a higher accuracy using larger datasets of complete mtDNA genomes and techniques accounting for some biases (e.g., effect of purifying selection in Soares et al., 2009). Recalculation of coalescent ages of European haplogroups confirmed that haplogroup U, and in particular U5 and U8, were the oldest haplogroups in Europe (Soares et al. 2009; Soares et al. 2010). The Late Glacial recolonisation of most of western and central Europe from the Franco-Iberian refugium ~10,000 – 15,000 yBP was not supported by recalculation of coalescent ages, which rather supported an earlier recolonisation shortly after the LGM. Recolonisation of eastern Europe from refugia located in eastern Europe immediately after the LGM was proposed from archaeological data (Dolukhanov, 1993) and was supported by recalculation of coalescent ages, of U4 notably (Soares et al; 2009). Recalculation of coalescent ages also confirmed the younger ages of the European clades T1 (~8,000 – 21,700 yBP) and J1a (6,100 – 17,600 yBP), in accordance with the idea that they might have reached Europe during the Neolithic expansion (Soares et al., 2009). However, the issues associated with the molecular-based dating approaches discussed earlier (see section ‘Dating using molecular data, and its limits’) still make these assumptions debatable.

Limits of modern DNA studies

The reconstruction of the genetic history of Europeans on the basis of coalescent ages for European mtDNA haplogroups is problematic in many regards (e.g., Barbujani & Goldstein, 2004). First, some of these coalescent ages estimates have been suggested to be inaccurate (see section ‘Dating using molecular data, and its limits’ above and e.g., Endicott & Ho, 2008; Soares et al., 2009). Coalescent ages have also often been misleadingly interpreted as representing colonisation ages. This frequent misconception was most famously illustrated by Barbujani et al., 1998, detailing the example of a potential colonisation of Mars by Europeans carrying mtDNA lineages whose coalescent ages date to the Palaeolithic. The descendants of these colonisers will carry lineages that coalesce in the Palaeolithic as well, which obviously does not represent the date of the colonisation of Mars by their ancestors (Barbujani et al., 1998). Finally, the reconstruction of past genetic history on the basis of modern genetic data is impaired by population genetic processes, such as e.g., population replacement, lineage extinction, and genetic drift, that can significantly bias the reconstruction and timing of past migratory and demographic events inferred from the analysis of modern-day mtDNA distributions. For example, lineage extinctions or migrations could have minimised the detectable genetic impact of early farmers in the modern-day gene pool of Europeans. The hypotheses that have been made on the basis of the study of the present-day mtDNA diversity can only be tested by the direct retrieval of mtDNA data from the Mesolithic hunter-gatherers and Neolithic early farmers (see review by Richards, 2003).

RECONSTRUCTING PAST HUMAN POPULATION HISTORY USING ANCIENT DNA

Ancient DNA (aDNA) has been used in humans and other organisms to test hypotheses emerging from the interpretation of the genetic diversity in modern populations by providing a snapshot of the past genetic diversity. In this section, I present a history of ancient DNA studies, the methodological challenges associated with studying aDNA and how they can be overcome. I also review previous application of aDNA and gaps in the knowledge of the human genetic history as reconstructed by previous modern and ancient DNA studies.

First ancient DNA studies

Early attempts to recover DNA molecules from long-dead specimens were fuelled by the assumption that DNA could survive through time under certain preservation conditions. Reports of the first DNA amplifications from ancient specimens were published 25 years ago (Higuchi et al., 1984; Pääbo et al., 1985). Higuchi et al. (1984) described the first amplification, through bacterial cloning, of aDNA from a hundred year old stuffed quagga (*Equus quagga*), an extinct member of the horse family, subspecies of the plains zebra. This finding was soon followed by DNA amplification from a human Egyptian mummy (5th century Before Christ, BC; Pääbo, 1985). These studies showed that DNA found in ancient remains was degraded and present at low-copy numbers, thus making the molecular methods available at that time very limiting.

The methodological limitations of cloning were overcome with the development of the Polymerase Chain Reaction (PCR, Mullis & Faloona, 1987). PCR allows specific fragments of DNA present in an extract to be specifically amplified. The sensitivity of PCR enabling the amplification of DNA from one single DNA molecule within an extract allowed its application to aDNA. Initially, aDNA studies were limited to relatively recent samples dated between hundreds to thousands of years, but soon turned into a race to retrieve DNA from the oldest possible sample. As a result, the retrieval of short DNA fragments was claimed for insects (e.g., DeSalle et al., 1992) and plants (Poinar et al., 1993) trapped in Dominican amber dated 25 to 40 million years, as well as from bones of an 80 million year old dinosaur (Woodward et al., 1994). The age for the most ancient DNA sequence was reached with the culture of bacteria trapped in salt crystals formed 250 million years ago (Vreeland et al., 2000).

However, critical re-analyses of nearly all of these aDNA sequences revealed that in fact they originated from contamination by modern human or microbial DNA (Zischler et al., 1995; DeSalle et al., 1994), highlighting the central nature of aDNA, which is post-mortem (i.e., after the death of the cell) degradation.

The nature of ancient DNA

Early work showed that the study of aDNA is challenging due to the fact that ancient remains are not as good a source of DNA as fresh specimens due to post-mortem damage.

Post-mortem degradation of DNA

The various processes involved in post-mortem damage of DNA are described below. After the death of the cells in an organism, the inactivation of the cellular DNA repair pathways exposes DNA molecules to degradation. DNA damage leads to a reduction in the amount of DNA that can be retrieved using molecular techniques such as PCR, but also can result in the actual alteration of the sequence information encoded by DNA molecules. Ancient DNA molecules are characterised by the following types of damage: (1) fragmentation, (2) blocking lesions, (3) base modification (Pääbo et al., 2000).

(1) Fragmentation

Fragmentation of DNA is caused by both biological and chemical processes. Bacteria, fungi and insects as well as enzymes such as lysosomal nucleases participate in the post-mortem fragmentation of DNA. Nuclease activity creates single-stranded breaks (nicks) in the DNA chain, which leads to the fragmentation of DNA molecules in the 100 - 500 bp size range.

Chemical processes, such as the hydrolysis of phosphodiester and glycosidic bonds, are also involved in the post-mortem fragmentation of DNA (Pääbo et al., 1989). Hydrolysis of the phosphodiester bond produces single-stranded nicks in the phosphate-sugar backbone, whereas hydrolysis of the glycosidic bonds, i.e., between the sugar back-bone and the nitrous base, generates abasic (apurinic) sites that are eventually responsible for strand breakage (by depurination).

The size of the DNA molecules present in aDNA extracts has been shown to be significantly smaller than in DNA extracts obtained from fresh material. This is the result of post-mortem DNA fragmentation. In particular, the copy-number of targeted aDNA fragments has been demonstrated to decrease exponentially as their size increases (Pääbo et al., 2004; Noonan et al., 2005; Malmström et al., 2007; Adler et al., 2010). The consequence of severe fragmentation of aDNA molecules is that long DNA fragments are extremely difficult to amplify by PCR.

(2) Blocking lesions

The size of fragments amplifiable by PCR is also limited by blocking lesions, which prevent strand elongation by the polymerase. Blocking lesions are found in the form of base or ribose fragmentation caused by hydrolysis or oxidation. The oxidative action of free radicals (e.g., peroxide, hydrogen peroxide and hydroxyl radicals) created by ionizing radiation was shown to generate hydantoin derivatives of pyrimidines that

block DNA elongation (Poinar et al., 1996). Blocking lesions were estimated to affect around 40% of the aDNA molecules in an extract (Heyn et al., 2010). Another form of blocking lesion is represented by cross-links. DNA-DNA or DNA-protein cross-links are the products of the 'Maillard' reaction, which describes the condensation between sugar (in DNA molecules) and primary amino-acid residues (in proteins and DNA molecules). N-phenacylthiazolium bromide (PTB) has previously been used in order to break intermolecular cross-links in aDNA extracts but such treatment is rarely applied nowadays.

Besides preventing DNA elongation of long DNA molecules, blocking lesions and DNA fragmentation also promote 'jumping-PCR'. Jumping PCR occurs when DNA elongation starts from one DNA strand, is then interrupted by a strand break or by a blocking lesion, and eventually resumes (potentially in the next cycle) using another DNA molecule as a template. Because the successive DNA molecules from which DNA has been amplified are not necessarily similar, jumping PCR leads to the generation of 'chimeric' sequences (reviewed in Pääbo et al., 2004; Willerslev & Cooper, 2005).

(3) Base modification

After the death of the cell, nitrous bases in aDNA molecules are subjected to hydrolysis, in particular leading to the deamination of cytosines into uracils, 5-methyl-cytosines into thymines, adenines into hypoxanthines, and guanines to xanthines. During the PCR, these modified bases are not accurately recognised by the polymerase, leading to nucleotide misincorporation and eventually to erroneous DNA sequences. As a result, errors arising from the amplification of aDNA molecules combine errors due to post-mortem base modification and the intrinsic error rate of the polymerase. The most common form of base modification in aDNA molecules has been shown to be caused by the hydrolytic deamination of cytosine into uracil. Uracil (U) is recognised as a thymine (T) by the DNA polymerase, which incorporates an adenine residue (A). Therefore, deamination of cytosines (C) leads to artificial CG → TA transitions (Lindahl, 1993; Hansen et al., 2001; Stiller et al., 2006; Gilbert et al., 2007; Brotherton et al., 2007; Briggs et al., 2007). To address this issue, treatment of aDNA extracts by Uracil-N-glycosylase (UNG) prior to PCR is sometimes used to remove uracil residues produced by the deamination of cytosines.

The main problem arising from base modification is that it generates artificial substitutions, which are the type of mutation most commonly found in organisms

because of their weaker evolutionary. Another problem caused by base modification in aDNA is that some sites in the mtDNA genome shown to be particularly prone to this type of damage ('hot spots') are also phylogenetically informative within the species under study (Gilbert et al., 2003). Base modification and subsequent nucleotide misincorporation could in theory lead to erroneous phylogenetic interpretation.

Contamination

The most important problem within aDNA research is contamination with intact exogenous modern DNA at any stage of the experiment (Handt et al., 1994). This was highlighted by studies that claimed to retrieve very ancient DNA sequences in early studies, which later proved to be false positives derived from contamination (see above; Pääbo, 2004). Intact DNA molecules originating from a modern organism, even in minute amounts, constitute a much better (less degraded) template for PCR amplification than the low-concentration, damaged DNA extracted from an ancient sample. As a result of the preferential amplification of modern over ancient DNA, the population of amplicons is likely to be made of a larger proportion of fragments arising from contaminants than from endogenous DNA. In many cases, a contaminant can give rise to all the amplicons, leading to erroneous results.

Contamination by modern DNA is a constant and ubiquitous hazard and may occur at any time during the excavation of the remains, storage in museums, handling during anthropological examination, sampling, and also during experiments in the aDNA laboratory. Sources of contamination include living organisms, other samples or aDNA extracts processed in parallel (cross-contamination) or previously PCR-amplified products (carry-over, reviewed in e.g., Sampietro et al., 2006). Procedures aiming at monitoring and reducing the risk of contamination within aDNA laboratories have been developed and refined over the past two decades (Pääbo et al., 2000; Cooper & Poinar, 2000; Cooper & Willerslev, 2005; Gilbert et al., 2005a; Gilbert et al., 2005b). Recently, studies have also proposed guidelines in order to reduce the risk of contamination during the treatment of the samples during and after their excavation (e.g., washing, brushing of the remains should be avoided; Sampietro et al., 2006; Pruvost et al., 2007). Contamination by modern DNA is particularly problematic when studying ancient human remains. Humans are inevitably involved in the excavation of the ancient remains, their morphological and genetic analysis, as well as in the production and delivery of the laboratory reagents and materials used to process the

samples. As a result, the risk of contamination with modern human DNA is omnipresent in ancient human DNA studies. In addition, human DNA contamination can be similar, or identical to individuals under study making it impossible to detect. In comparison, studies of other species have been suggested to be less prone to contamination. However, this ignores the problem that previously amplified amplicons (sometimes slightly damaged by laboratory bleach and UV treatments) are a significant contamination risk on any aDNA study.

These factors led to the establishment of a range of laboratory procedures to reduce the risk of contamination by modern DNA and to the authentication of aDNA results (Pääbo et al., 2000; Cooper & Poinar, 2000; Cooper & Willerslev, 2005; Gilbert et al., 2005a; Gilbert et al., 2005b). However, it is important to note that even when an aDNA result passes all authentication criteria it should only be considered to have failed to have been disproven, rather than having been proven.

Authenticating ancient DNA data

A set of rules and measures taken to reduce the impact of post-mortem damage and contamination by modern DNA have been defined to validate aDNA results (Handt et al., 1996; Poinar et al., 1996; Krings et al., 1997; Cooper & Poinar, 2000). The authenticity of aDNA data is commonly assessed based on the following criteria:

(1) In order to reduce contamination of ancient samples by modern DNA, direct contact between ancient remains and living organisms should be avoided. Researchers (i.e., archaeologists, anthropologists, biologists) working on the excavation site, in museums or in laboratories, should wear protective clothing including gloves, face masks, face shields, full-body suits, gum boots when in presence of the samples. In addition, all the supplies used to handle ancient samples should be decontaminated using oxidants, such as sodium hydroxide (commercial bleach) before use.

(2) Post-excavation treatment of ancient remains, such as washing or brushing, that could affect the porosity of samples like teeth or bones should be avoided in order to limit the contamination of the interior of ancient specimens. Ideally, ancient samples should be excavated with the soil in which they are embedded and should not be washed before the samples are processed in a laboratory dedicated to aDNA work. This precaution is thought to reduce the risk of contamination from the outer surface of ancient samples (Sampietro et al., 2006; Pruvost et al., 2007).

(3) Pre-PCR work on ancient specimens should be carried-out in a laboratory exclusively dedicated to aDNA work. Precautions should be taken to minimise contamination by modern DNA, such as, physical isolation of the aDNA-dedicated laboratory from any molecular biology work (Lindahl, 1993), positive air pressure conditions in order to reduce contamination from the environment, routine decontamination of the laboratory surfaces and instruments by exposure to ultraviolet radiation (UV) and thorough cleaning using DNA oxidants such as bleach, as well as Decon (Decon labs) and Ethanol. In addition, all personnel entering the facility should wear the protective clothing described in (1) at all times.

(4) Prior to aDNA extraction, the outer surface of ancient samples should be decontaminated. Various decontamination methods have previously been proposed (Kolman & Tuross, 2000; Caramelli et al., 2003; Vernesi et al., 2004). In particular, ancient samples can be subjected to UV radiation and their outer surface cleaned using bleach and eventually mechanically removed using cutting discs and/or sandblasting units.

(5) It is of prime importance that aDNA extraction from two independent samples for a same individual are replicated in order to identify contamination that may have occurred during aDNA experiments, as this is one of the few robust ways to support an aDNA result (Pääbo et al., 2004; Gilbert et al., 2005a).

(6) The fact that the distribution of DNA molecule sizes in aDNA extracts is skewed towards small fragments can be used to support the authenticity of aDNA sequences. Quantitative real time PCR (qPCR) assays can be used to determine copy-numbers of DNA targets of varying fragment lengths within an aDNA extract. Yields of DNA amplification significantly larger for smaller amplicons than for larger amplicons can be interpreted in support of low levels of contamination by longer modern DNA molecules (Malmström et al., 2007).

(7) Due to the exponential decay of aDNA templates (Pääbo et al., 2004; Noonan et al., 2005; Malmström et al., 2007; Adler et al., 2010) PCR amplification success rates increase as the size of the targeted fragment decreases. Consequently, primers should be designed to amplify short overlapping fragments. To reduce the impact of contamination of PCRs by modern DNA from the environment, primers should be designed to amplify DNA from the species under study as specifically as possible. Of note, modern DNA contaminants from humans, pigs, cows and chickens (Leonard et al., 2007) have been found in reagents used in aDNA laboratories.

(8) In order to monitor modern DNA contamination arising from materials and reagents used during DNA extraction and DNA amplification, extraction and PCR blank/mock controls (carried out in absence of ancient sample or extract) should be performed at the same time as ancient samples are processed (Pääbo, 1989). However, in some instances the absence of DNA products in PCR blank controls does not necessarily indicate the absence of contamination in the PCRs containing aDNA extracts. This phenomenon, termed the carrier effect, has been suggested to relate to non-specific binding of low numbers of contaminant templates by molecules present in aDNA extracts, such as sugars, proteins and DNA or charged areas of plastic tube walls. The absence of such molecules from traditional negative controls potentially allows the few contaminant templates to be amplified by PCR (Cooper, 1992; Handt et al., 1994; Leonard, 2006). In order to detect the carrier effect, ancient remains belonging to another species than the species of interest can be used as negative controls. In these negative controls, the biochemical conditions of the negative controls are similar to those of the aDNA extracts (Pääbo et al., 2004). This approach is difficult to apply to ancient human DNA work however, as animal remains have generally been handled extensively by archaeologists, so that these negative controls may also yield amplicons with human primers in most cases, independently from possible contamination of the reagents or the laboratory.

(9) Cloning amplicons and sequencing many clones can be used to identify erroneous mutations arising from contamination and base misincorporation (Krings et al., 1997). In the case where a molecule characterised by a modified base is used as a template during the late stages of the PCR, artificial sequences arising from this damaged template will represent a small proportion of the amplified molecules, and hence, of the clone sequences. These rare mutations within clone sequences are called singleton substitutions, and can be identified easily and removed from the consensus sequence. Mutations representing the endogenous phylogenetic signal should appear in the majority of the clones in a consistent fashion. Statistical methods, based on the c-statistic (Helgason et al., 2007) and Bayesian phylogenetics (Ho et al., 2007) were also developed in order to detect artificial mutations arising from DNA damage and to determine endogenous haplotypes. The use of cloning to authenticate aDNA sequences has recently been challenged (Pruvost et al., 2008; Winters et al., 2011). In low DNA template conditions, damaged molecules may serve as templates during the early stages of the PCR amplification. As a consequence, damaged sequences may

represent the overwhelming majority of the clones sequenced (Hofreiter et al., 2001a). In these conditions, the results of cloning may be misleading and repeat PCR amplifications multiple times followed by direct sequencing may be more efficient at retrieving the endogenous DNA sequence (Pruvost et al., 2008).

(10) Multiple independent repetitions of PCR amplification should be used in order to address the limits of cloning in identifying artificial mutations (see (9)), but also to monitor and rule out sporadic contamination of PCR reactions.

(11) The impact of miscoding lesions due to cytosine deamination can potentially be reduced by removing Uracil residues from aDNA sequences by Uracil N-glycosylase (UNG) treatment that leads to the fragmentation of DNA molecules containing Uracil residues (Pääbo, 1989; Hofreiter et al., 2001a). An alternative approach that does not lead to a reduction in the amount of aDNA templates, specifically ‘repairs’ deaminated cytosines through ‘short patch base excision repair’ (Mitchell et al., 2005). The latter is not commonly used as it is complex to carry out, and has not been shown to lead to large scale improvements.

(12) In ancient human studies, potential contamination should be monitored by typing all workers involved (archaeologists, anthropologists and biologists) and by comparing sequence profiles with data from the ancient specimens.

(13) If logistically possible, a fraction of aDNA results should be replicated in an independent laboratory (Pääbo et al., 2004, Willerslev, 2004). Replication of aDNA extraction and sequencing from an independent sample in a separate aDNA laboratory allows the detection of laboratory-specific contaminations in both laboratories. Ideally, samples used in independent replications are best sent directly from the museum collection or the excavation site, rather than from the aDNA laboratory requesting the independent replication.

(14) Assessment of the biochemical preservation of an ancient sample can provide support to the authenticity of aDNA data generated from this specimen. The biomolecular preservation has been used as a proxy for the extent of diagenetic alteration in an ancient sample, and probability of DNA survival. A common approach is to estimate the total amount, composition and racemisation of amino acid (reviewed in Pääbo et al., 2004). The latter describes the structural conversion of an amino acid from one racemic form to another (e.g., conversion of the L to D forms in aspartic acid; Poinar et al., 1996), which are indicative of the extent of biochemical degradation of the sample. Ancient remains containing small amounts of amino acids or significantly

racemised amino acids have been shown to contain low amounts of amplifiable DNA molecules. Other methods that have been put forward include estimation of the ratio of peptide fragments to single amino acids via mass spectrometry (Poinar et al., 1999), direct assessment of bone histology (Bailey et al., 1996; Barnes et al., 2000; Colson et al., 1997; Jans et al., 2004), measurement of porosity and density in bone (Nielsen-Marsh et al., 2000) and transmission electron microscopy (Koon et al., 2003). These methods can be used to rapidly screen available samples to identify remains that are most likely to allow DNA amplification (Pääbo et al., 2000). However, the predictive ability of racemisation levels in ancient samples has been questioned (Collins et al., 2009), and it is not commonly used.

(15) Unusual phylogenetic signals can be indicative of erroneous aDNA sequences due to contamination, jumping PCR or miscoding-lesions. In the case of ancient human mtDNA studies, the phylogenetic consistency of the HVR can be assessed in some cases. For example, SNPs diagnostic of Asian or European haplogroups, with respect to the origin of the specimen. Phylogenetic consistency between HVR and coding region sequences provide a further important test of the authenticity of aDNA results.

From a practical perspective, it is possible that some of these procedures cannot be applied due the particular context of the burial - e.g., lack of multiple independent samples from a given individual for independent replication - or of the excavation, e.g., lack of comparative sequence profiles from all archaeologists or anthropologists in the case of ancient human DNA work. It should be kept in mind that while errors arising from miscoding lesions can technically be identified, the absence of contamination from exogenous sources can never be absolutely ruled out.

Survival of DNA

The relative rates at which various types of DNA damage accumulate are still imprecisely known, and depend on the post-mortem history of the remains. It is clear that there is no direct correlation between DNA preservation and age of the samples (Pääbo, 1989; Höss et al., 1996; Poinar et al., 1996). Studies have attempted to describe DNA diagenesis, i.e., post mortem changes, under different preservation conditions (e.g., Burger et al., 1999) in order to identify conditions that are optimal for DNA preservation. However, such tests cannot be performed across a wide range of

realistic conditions, and include multiple environmental factors. The conditions of the environment - as well as the micro-environment (in the direct vicinity of the samples; Hagelberg et al., 1991) - in which the samples have been preserved, are thought to have a significant impact on DNA preservation. Amongst other things, microbial activity has been shown to be involved in DNA damage (e.g., Pääbo, 1989; Burger et al., 1999) and as a consequence conditions detrimental to bacterial communities are believed to be favourable to aDNA survival. Cold environments with conditions that minimise the impact of water activity (involved in hydrolysis) and oxygen (involved in oxidation), such as permafrost, appear to provide optimal conditions for DNA preservation. Kinetic calculations based on in vitro tests, predict that small fragments of DNA (100 – 500 bp) are expected to survive no more than 10,000 years in areas characterised by a temperate climate, whereas they can reach a maximum of 100,000 years in colder environments (Pääbo et al., 2000). Accordingly, the oldest samples that have yielded replicated aDNA sequences were found in ice cores (450,000 – 800,000 year old plants and insects; Willerslev et al., 2007) and caves (400,000 year old cave bear; Valdiosera et al., 2006). Likewise, ancient DNA sequences have been analysed from constantly frozen soils, for example: 50,000 – 65,000 year old bison mtDNA (Shapiro et al., 2004), and 300,000 – 400,000 year old plant chloroplast DNA (Willerslev et al., 2003). The excellent preservation conditions of permafrost even allowed the complete genome of ~20,000 year old woolly mammoth specimens to be sequenced (Miller et al., 2008), although recent advances in genomic sequencing have now permitted this with non-frozen hominid remains (Neanderthals in Green et al., 2010; Denisovans in Reich et al., 2010). The aDNA amplification strategy should be adapted to the age and the particular environmental conditions in which the samples have been preserved, in particular in terms of the size of the fragments targeted. In effect, permafrost samples are far more likely to yield larger amounts of longer amplicons than samples originating from a temperate environment of the same age.

Applications of ancient DNA

A wide range of sample types have yielded aDNA including: bones, teeth, hair (Hagelberg et al., 1989; Hagelberg et al., 1991; Gilbert et al., 2004a), feces (coprolites; Poinar et al., 1998), leather (Burger et al., 1999; Burger et al., 2000; Vuissoz et al., 2004), insects (Thomsen et al., 2009), feathers (Rawlence et al., 2009), egg shells

(Oskam et al., 2010), plants (e.g., Willerslev et al., 2003), ice cores (Willerslev et al., 2007), sediments (Willerslev et al., 2003). Ancient DNA has been recovered from freshly excavated samples and museum specimens including mounted animals (Higuchi et al., 1984), and alcohol-preserved tissues. Within vertebrate hard tissues, the best sources of aDNA are generally considered to be compact bones and teeth, and they also are the types of samples the most commonly preserved in the archaeological record. In particular, DNA in teeth is thought to benefit from the protection of the enamel.

Ancient DNA has been used to analyse a wide range of species, over broad geographical and temporal scales to address a wide range of biological questions:

Phylogeny and phylogeography of extant and extinct species

Ancient DNA has often been used to determine the phylogenetic or taxonomic position of extinct species – which has often been previously assessed on the basis of morphological or fossil data such as quaggas (Higuchi et al., 1984), mammoths (Debruyne et al., 2003), moas (Cooper et al., 1992; Cooper et al., 1995); cave bears, (Loreille et al., 2001), casseroles, Neanderthals (Krings et al., 1997; for a more complete list see reviews, e.g., Hofreiter et al., 2001b; Pääbo et al., 2004; Ramakrishnan et al., 2009). Ancient phylogenetic studies initially used short sequences from particular informative regions of the mtDNA genome, such as HVR or the cytochrome b but advances in sequencing technology have allowed the characterisation of complete mtDNA genomes (e.g., moas, format as above Cooper et al., 2001; mammoths, Krause et al., 2006; Neanderthals, Green et al., 2008; aurochs, Edwards et al., 2010) and even complete genomes: mammoths (Miller et al., 2008); Neanderthals (Green et al., 2010), humans (Rasmussen et al., 2010), Denisovans (Reich et al., 2010).

In addition to extinct species, ancient DNA has been used to reconstruct population level processes within living taxa, for example brown bears (Barnes et al., 2002), voles (Hadly et al., 2004), kiwis (Shepherd & Lambert, 2008; for a more complete list see reviews, e.g., Hofreiter et al., 2001b; Pääbo et al., 2004; Ramakrishnan et al., 2009).

Domestication processes

Ancient DNA has been used to study the processes involved in animal and plant domestication - i.e., the human-driven selection of traits in wild ancestral populations – which was initiated during the transition from foraging to agricultural human lifestyles, termed the ‘Neolithic transition’. Domestication processes have been investigated by aDNA for various plant and animal species including (see review in Zeder et al., 2006 and Zeder et al., 2008): cattle (Troy et al., 2001; Beja-Pereira et al., 2003; Edwards et al., 2010), dogs (Savolainen et al., 2002), horses (Vila et al., 2001), pigs (Larson et al., 2010), and South-American maize (Jaenicke-Despres et al., 2003).

Evolution of biological functions

The genetic basis of natural selection acting on functional genes has also been investigated using aDNA, for example in Neanderthals (Lalueza-Fox et al., 2007) and in mammoths (Campbell et al., 2010). Species-specific substitutions were identified in the sequences of the genes coding for the pigmentation-regulating melanocortin receptor 1 of Neanderthals and the haemoglobin of mammoths. The activity of the proteins was tested *in vitro* and functional analysis suggested that Neanderthals may have had varied pigmentation levels while mammoth haemoglobin exhibited adaptations to the cold conditions of the Ice Age in the northernmost latitudes. Other examples include the investigation of coat coloration in horses (Ludwig et al., 2009), lactase persistence (Burger et al., 2007) and ABO blood groups in humans (Hummel et al., 2002). These studies demonstrate the potential of aDNA to provide information about phenotypes and physiology in extinct species or past populations.

Temporal population dynamics

Changes in the demographic history of ancient populations through time can be investigated through the retrieval of aDNA from temporally sampled specimens. Evolution of population size can be estimated by using the program Bayesian Skyline Plot (Drummond et al., 2005) implemented in the phylogeny reconstruction program BEAST (Drummond & Rambaut, 2007). For example, this approach was applied to ancient mtDNA obtained bison (Shapiro et al., 2004) and musk oxen (Campos et al., 2010). Another program based on the coalescent model allowing demographic models to be tested using temporally-sampled genetic data (Bayesian Serial SimCoal; Anderson et al., 2005) has been applied to a range of aDNA datasets to investigate a range of

questions related to the reconstruction of population history (Ramakrishnan et al., 2009). In the future, results could potentially be compared with knowledge about changes in climatic conditions and human distribution in order to determine their relative contribution to population extinction, as suggested, for example, for mammoths (Nogués-Bravo et al., 2008).

Other applications

Other applications of ancient DNA include the reconstruction of palaeoclimates (from plant material; Willerslev et al., 1999) and paleodiets (from coprolites; Poinar et al. 1998; Wood et al., 2011). The study of palaeopathologies through the retrieval of bacterial and parasitic DNA is a promising field of application for aDNA (Haensch et al., 2010). However, early studies (e.g., *Yersinia pestis* in medieval specimens; Raoult et al., 2000) lacked convincing arguments to demonstrate the reliability of the bacterial data reported, in particular the absence of contamination from closely related bacteria or strains (Gilbert et al., 2004b).

Future directions of ancient DNA studies

Phylogenies and past population history will be constructed with increasingly detailed resolution as more aDNA data will become available. With the development of techniques aiming to apply high throughput sequencing technologies to ancient DNA (e.g., library construction and target enrichment strategies), it is likely that the amount of aDNA data will increase dramatically in the coming years. Improvements in aDNA methodologies are also set to broaden the range of applications of aDNA.

Similarly, after facing many methodological problems mostly associated with detecting and removing contamination, the number of reliable DNA sequences obtained from ancient human remains has considerably increased in recent years. Considering the large amount of human samples collected and available from the archaeological record, a wide array of questions about human evolutionary history are set to be addressed by multi-disciplinary studies combining aDNA with anthropology, archaeology, and linguistics.

Ancient human DNA studies

The problem of authenticity in ancient DNA studies

The authenticity of sequences from many human aDNA studies, especially at the beginning of aDNA research has failed to convince the scientific community (Bandelt, 2005). This can be explained by the omnipresent problem of contamination by modern human DNA and a lack of appropriate authentication strategies in these studies. A famous example suggesting contamination of ancient samples is that of ‘Ötzi’, a ~5,300 year old Chalcolithic individual found mummified in glacial ice in the Italian Tyrol. Initial ancient mtDNA sequences suggested that this individual belonged to haplogroup K (Handt et al., 1994; Rollo et al., 2006). The reported haplotype was later found to be erroneous, likely a result of contamination (Ermini et al., 2008; Endicott et al., 2009b), despite independent replication of the results between leading aDNA laboratories. Another famous example of highly criticised ancient human DNA results are those obtained from ~60,000 year old remains from Lake Mungo, Australia (Adcock et al., 2001). These examples show the difficulties involved in authenticating ancient human DNA results, especially in cases where researchers are likely to be genetically closely related to the individuals they study (e.g., European researchers studying ancient European remains). While some spectacular claims (Hawass et al., 2010) are still being made on the basis of ancient human DNA results, and are fiercely criticized (Marchant, 2011), more caution and more stringent procedures are now being applied to the study of ancient human DNA. The authentication criteria used in aDNA studies are an important guide to allow reviewers and readers to assess the validity of the presented aDNA results.

Application of ancient human DNA studies

(1) Hominin evolution

Ancient DNA has provided completely unexpected insights into the evolution of anatomically modern humans and their phylogenetic relationships with extinct close relatives, such as Neanderthals and the recently discovered Denisova hominin.

Skeletal remains showing the human-like, but robust, anatomical features of Neanderthals were first discovered in the Neander valley (‘Neanderthal’ in old German) in 1856. Additional discoveries in the skeletal record indicate that Neanderthals lived over a geographical range covering western Europe to central

Europe, the Zagros mountains (Kurdistan), and the Altai, between ~28,000 and ~200,000 yBP. Neanderthals then disappeared from the archaeological record and many hypotheses have been proposed to explain their extinction, e.g., climate change or out-competition by humans. In Europe, anatomically modern humans and Neanderthals overlapped for ~10,000 years, since anatomically modern humans are thought to have reached Europe around ~40,000 yBP (Bischoff et al., 2003; Harvati 2007). This observation led to questions that have been discussed for more than a century, such as, how genetically closely related were Neanderthals to anatomically modern humans? Did admixture happen between anatomically modern humans and Neanderthals? If admixture happened, what is the genetic contribution of Neanderthal genes to the present-day human gene-pool (Tattersall et al., 1999)? Ancient DNA provided direct evidence to fill some of the gaps in the archaeological and anthropological investigation of these questions. First mtDNA from a Neanderthal specimen showed that Neanderthals fell outside the mtDNA variability of present-day humans (Krings et al., 1997). Subsequent analyses of Neanderthal mtDNA in additional specimens confirmed these results (Krings et al., 2000; Ovchinnikov et al., 2000; Schmitz et al., 2002; Lalueza-Fox et al., 2005; Orlando et al., 2006; Krause et al., 2007; Green et al., 2008; Briggs et al., 2009; Lalueza-Fox et al., 2011). Serial coalescent-based statistical analyses of the genetic discontinuity between Neanderthals, on the one hand, and on the other hand, Palaeolithic (Caramelli et al., 2008) and present-day humans, under a wide range of demographic models supported little admixture between anatomically modern humans and Neanderthals (Belle et al., 2009). However, the sequencing of the complete genome of a Neanderthal specimen revealed that Neanderthals shared parts of their genome with all non-African humans, indicating some level of interbreeding between anatomically modern humans and Neanderthals, probably in the Middle East as modern humans moved out of Africa (Green et al., 2010). Nuclear gene flow between extinct hominins and anatomically modern humans is also indicated through the analysis of the complete genome of a ~40,000 year old specimen from the Denisova cave in the Altai mountains, Siberia (Reich et al., 2010; see also Krause et al., 2010a).

These studies showed the potential of aDNA to provide unique genetic evidence about hominin evolution. In the future, the investigation of the evolutionary relationships among hominins may be broadened to, for example, European *Homo heidelbergensis* (~400,000 - 600,000 yBP; Schoetensack, 1908), Indonesian *Homo*

floresiensis (~12,000 yBP; Brown et al., 2004) or the remains from the Zhiren cave in China (~100,000 yBP; Liu et al., 2010). Future development of aDNA techniques, may allow the genetic study of these remains that has been limited until today by the age of the samples (e.g., *H. heidelbergensis*) or warm and wet conditions non-optimal for DNA preservation (*H. floresiensis*).

(2) Ancient past human genetic diversity

Ancient DNA studies have reported mtDNA HVR data over a broad temporal and geographical scale. On the temporal scale, aDNA techniques have been applied to human remains ranging in age from the Palaeolithic around 23,000 - 30,000 (Caramelli et al., 2008; Krause et al., 2010b) to a few centuries ago (Rogaev et al., 2009). Studies have reported ancient human DNA over a wide geographical range, including Central Asia (Lalueza-Fox et al., 2004), Europe (Handt et al., 1996; Izagirre et al., 1999; De Benedetto et al., 2000; Haak et al., 2005; Sampietro et al., 2007; Caramelli et al., 2007a), the Adaman Islands (Endicott et al., 2006), northern America (Gilbert et al., 2007), Siberia (Keyser et al., 2009), eastern Asia (Adachi et al., 2009), southern America (Kemp et al., 2009), Iceland (Helgason et al., 2009), Anatolia (Ottoni et al., 2011). Ancient complete mtDNA genomes have also been sequenced, but their number is still limited to very few individuals: a 3,400 - 4,500 year-old Palaeo-Eskimo from Greenland (Gilbert et al., 2008), a 5,100 - 5,350 year-old Tyrolean mummy (Ermini et al., 2008) and a 30,000 year-old Western Russian individual from Kostenki 14 (Krause et al., 2010b).

Some ancient human DNA studies have targeted other markers. Y-chromosome markers helped compare maternal and paternal genetic histories of past human populations (e.g., in south Siberian and Mongolian nomads, Keyser-Tracqui et al., 2003, Keyser et al., 2009; and in Neolithic early farmers, Haak et al., 2005; Haak et al., 2008; Haak et al., 2010; Lacan et al., 2011). Amelogenin typing was applied to determine the sex of ancient human skeletal remains (Meyer et al., 2000). STR profiles generated from graveyards were used to formulate hypotheses about possible familial relationships (e.g., Keyser-Tracqui et al., 2003). Mitochondrial DNA, Y-chromosome data and STR profiles were combined with strontium isotope analysis to identify the oldest nuclear family in a Corded Ware Culture multiple burial in Germany (4,600 yBP; Haak et al., 2008). Human evolution mechanisms, such as selection, have also been investigated by applying aDNA to the typing of nuclear SNPs involved in metabolic functions. This was done to examine the prevalence of

mutations associated with lactase persistence, i.e., the ability to digest milk in adulthood, in ancient populations of early farmers of central Europe (Burger et al., 2007) and of Scandinavian hunter-gatherers (Malmström et al., 2010).

Most ancient human DNA studies have reported genetic data from isolated samples rather than populations. In early studies this was explained by the challenges involved in the generation of aDNA data from large numbers of human individuals, including the contamination risk. The number of individuals in the populations sampled for aDNA is significantly smaller (typically smaller than 25 individuals) than sample sizes typically reported in genetic studies of modern populations.

(3) Genetic characterisation of historical figures

Ancient DNA has also been applied to characterise historical figures. The most famous example is the study of the remains attributed to the family of the last Tsars of Russia, the Romanovs. Results eventually led to the identification of all the members of the family used a set of markers including paternal Y chromosome, maternal mtDNA, and STRs that were compared with genetic data obtained from living maternal and paternal relatives (Knight et al., 2004; Rogaev et al., 2009). Other historical individuals investigated by aDNA include: the French King Louis XVI (Lalueza-Fox et al., 2010) and his son Louis XVII (Jehaes et al., 2001), the medieval Italian poet Petrarch (Caramelli et al., 2007b), the founder of Stockholm, Birger Magnusson (Malmström et al., 2011), and the Pharaoh Tutankhamun (Hawass et al., 2010) although the latter has been highly criticised (Marchant, 2011).

Investigation of ancient human mitochondrial DNA diversity in Europe

Ancient mitochondrial DNA generated from European human populations has been used to investigate genetic continuity/discontinuity through time (in the past and between past and present-day populations). The question to which aDNA has provided the most evidence is that of the Neolithic transition in Europe and its impact on the gene pool of modern Europe. Studies from the ancient data recovered from Upper Palaeolithic (~10,000 - 45,000 yBP; De Benedetto et al., 2000; Caramelli et al., 2008; Bramanti et al., 2009; Krause et al., 2010b), Mesolithic (~6,000 – 12,000 yBP; De Benedetto et al., 2000; Bramanti et al., 2009; Malmström et al., 2009) and Neolithic (~5,000 – 7,500 yBP; Handt et al., 1996; De Benedetto et al., 2000; Haak et al., 2005; Malmström et al., 2009; Sampietro et al., 2007; Haak et al., 2008; Ermini et al., 2008; Haak et al., 2010) populations provide an emerging pattern of the mtDNA diversity

and population dynamics of prehistoric Europe. Mitochondrial data obtained from Palaeolithic and Mesolithic hunter-gatherers from central/eastern Europe (Bramanti et al., 2009) and Scandinavia (Pitted Ware Culture, Malmström et al., 2009) suggested a genetic discontinuity between the Mesolithic and the present-day. When comparing European Mesolithic populations to populations of the Neolithic in Spain (Sampietro et al., 2007), France (Lacan et al., 2011) and Germany (Haak et al., 2010), differences in the genetic structure of hunter-gatherers and early farmers appear, supporting the idea that the Neolithic transition involved a significant human migrations. However, the lack of data for Mesolithic populations of western/southern Europe currently prevents definitive conclusions about genetic continuity between the Mesolithic and the Neolithic in this region. The available of Neolithic mtDNA datasets also confirmed archaeological hypotheses proposing a geographically heterogeneous nature for the Neolithic transition in Europe (Price et al., 2000; Bocquet-Appel et al., 2009). The mtDNA structures observed in Neolithic populations located around the Mediterranean Basin, i.e., in Spain and in southern France, were shown to be different from that of populations of the Linienbandkeramik culture (LBK), central Germany. Neolithic populations of Spain and France showed genetic similarities with modern-day populations of southern Europe, suggesting genetic continuity in this area. In contrast, the LBK population of Germany showed a large genetic discontinuity with present-day central Europeans. This was partly due to the high frequencies of haplogroup N1a in the LBK samples, which constitute the genetic signature of this population, while N1a is found at very low frequencies in modern-day Europe (Haak et al., 2005). This result implies a major role of post-Neolithic population processes (migration, population replacements, and bottlenecks). In this case, the genetic discontinuity between the Neolithic and the present-day in central Europe was demonstrated using a range of statistical methods that are commonly used in modern population genetics (e.g., principal component analysis), or that take into consideration the temporal sampling of the genetic data (serial coalescent simulations with the program Bayesian Serial SimCoal; Anderson et al., 2005). Bayesian Serial Simcoal coalescent analysis was previously applied to the investigation of genetic continuity between pre-classical Etruscans and modern Tuscans (Belle et al., 2006), between Nuragic and present-day Sardinians (Ghirotto et al., 2010), and between Wari and post-Wari populations in the Andes (Kemp et al., 2009). However, classical and

coalescent-based statistical techniques are rarely combined in ancient population genetics studies.

The example of the investigation of the Neolithic transition in Europe shows how aDNA can be used to unravel population mtDNA structures that could not have been reconstructed from the analysis of mtDNA in modern populations, and to statistically test models of population history based on direct evidence from aDNA.

Limitations of ancient and modern-day human DNA

The reconstruction of past human population history on the basis of modern genetic data is limited due to certain drawbacks that can be circumvented by the analysis of aDNA data. While recent studies have dramatically increased the amount of genetic data available for ancient human populations, the genetic characterisation of ancient human populations is still far from being dense enough at a geographical and temporal scale to allow past human genetic history to be reconstructed at great resolution, and will often not be able to resolve complex demographic histories.

Reconstruction of the genetic history of Europeans should extend in time and focus more on peripheral areas such as north eastern Europe that extends from the Black Sea in the South to the Kola peninsula in the North. The recolonisation of eastern Europe after the Ice Age, was proposed to have originated from ‘cryptic’ refugia in northern and eastern Europe (Dolukhanov, 1993). In eastern Europe, the archaeological record suggests a more gradual impact of the early Neolithic, which is thought to have involved less gene flow from early farmers (Zvelebil & Dolukhanov, 1991). The peripheral area of eastern Europe is also of interest because of its geographical location. In the South, eastern Europe opens to the central Eurasian Steppe and to central Asia. In the North, it is bound by the Urals, a ‘buffer zone’ between Europe and Siberia. As a result, eastern Europe can be proposed to have been the recipient of genetic influences from various origins in the East (e.g., Grousset 1970; Kozłowski & Bandi, 1984). However, populations of north east Europe were found to fit within the European mtDNA diversity, despite small differences (e.g., Pliss et al., 2006; Lappalainen et al., 2008; Tambets et al., 2004; Malyarchuk et al., 2008). It is possible that eastern Europe underwent a similar genetic history to southern/western/central Europe, as suggested by similar mtDNA composition. Alternatively, eastern Europe has been subjected to various external influences, which have been diluted by gene flow from the West (e.g., the Slavs from central Europe, the

Vikings from Scandinavia in Medieval times), leading to an homogenisation with the rest of Europe. It seems clear that such complex scenarios cannot be reconstructed on the basis of modern genetic data alone and that ancient DNA can potentially provide resolution to tease these scenarios apart. Despite the fact that some studies have reported aDNA from eastern Europe, the knowledge of prehistoric populations in this area is still limited to a few Palaeolithic/Mesolithic individuals (Bramanti et al., 2009; Krause et al., 2010).

Modern DNA studies have also found their limits in the reconstruction of European genetic outliers such as the Saami of northern Europe (Guglielmino et al., 1990), Basques of the Pyrenees (Bertranpetit et al., 1995), Icelanders (Helgason et al., 2000), Sardinians (Morelli et al., 2000) and Finns (Lahermo et al., 1996). These populations are characterised by low genetic diversity, most likely resulting from geographical, linguistic and/or cultural isolation. Modern DNA studies have identified founder events, and/or genetic drift as causes of the reduction in genetic diversity of these populations, but little is known about the timescale of the genetic history of European genetic outliers. Following the mtDNA diversity of populations potentially ancestral to European genetic outliers, as enabled by aDNA, appears a powerful approach to access this temporal information. However, previous aDNA studies have not provided conclusive elements to the question of the origins of Saami (Bramanti et al., 2009; Malmström et al., 2009), or Sardinians (Caramelli et al., 2007a).

Few ancient human DNA studies have adopted a population approach so far, i.e., targeting a significant number of individuals belonging to a same geographical, temporal, anthropological and/or cultural entity, due to methodological challenges. Furthermore, the number of mtDNA genomes obtained from ancient human remains is still small. Mitochondrial genomes could however provide precious information in terms of genetic relationships and genetic affinity between populations. Finally, it appears that the range of statistical techniques available to compare aDNA among ancient populations or with modern populations is not always fully explored. In particular, methods that have been designed to account for temporal discrepancies in datasets, such as coalescent-based analyses (e.g., Bayesian Serial SimCoal, Anderson et al., 2005), could be more widely applied in order to rule out simple demographic models and explain genetic patterns in ancient human genetic datasets.

PRESENTATION OF THE PROJECT

General aims of the study

The work presented here aims at extending the knowledge of the genetic history of past human populations by overcoming the limits inherent to the study of mtDNA in modern populations and the lack of data available from ancient remains. This study will investigate the genetic history of eastern Europeans - including the population of European outliers - the Saami, in the Mesolithic, the Bronze Age, the Iron Age and in historical/present times, as well as the Sardinians, another genetic outlier.

The first objective is to obtain an aDNA dataset that is informative and reliable. The mtDNA dataset generated here should be characterized well enough to allow a comparison, at the haplogroup and the haplotype levels, with mtDNA data available from previously described ancient and modern populations. The impact of artificial mutations arising from DNA damage and contamination by modern DNA should be minimised.

The second objective of this study is to detect genetic “links” or “affinities” between ancient and modern populations, i.e., similarities in their genetic structures suggesting common origins. Eventually, the comparison of the available mtDNA data from ancient and modern populations will lead to the identification of genetic continuity and/or discontinuity. The goal of this work is to identify and date distinct population events - i.e., migrations, expansions, extinctions - that may have shaped the current mtDNA diversity of European populations, including genetic outliers. In the case of eastern European populations, the identification of genetic discontinuities could reveal migrations from western or eastern Eurasia. Ancient DNA would then provide a timeframe for these migrations. Ancient mtDNA data will be assessed in the light of archaeological and geographical context information of the sampled sites in order to evaluate the impact of cultural and environmental factors on the distribution of mtDNA lineages in space and time.

Samples and organisation of the thesis

A total of 269 samples representing 161 individuals (127 eastern Europeans, 34 Sardinians) were collected for this study. I tried to adopt a population approach, in which aDNA is extracted from a significant number of individuals (here, more than

ten individuals) for the same site. This thesis is organized in four chapters according to the geographical origin of the samples examined (Table 1 and Figure 1):

Chapter One: The mtDNA diversity in north east Europe is investigated for three time-periods: the Mesolithic (~7,000 - 7,500 yBP; Uznyi Oleni Ostrov and Popovo archaeological sites), the Bronze Age (~6,500 yBP; Bolshoy Oleni Ostrov archaeological site) and the 17-18th centuries (Chalmny-Varre Saami graveyard). The objective of this chapter is first to broaden the knowledge of the mtDNA diversity in European Mesolithic populations by extending the geographical range to the North East. The deep ancestry of present-day North East Europeans and Saami genetic outliers is explored by testing mtDNA continuity/discontinuity in this transect through time (from the Mesolithic to the present-day).

Chapter Two: I evaluate the power of combining aDNA and whole mitochondrial genome sequencing in order to reconstruct human mitochondrial phylogenies. This chapter presents the sequencing of the mitochondrial genome of a 7,500 year old individual from the Mesolithic site Uznyi Oleni Ostrov in Karelia, Western Russia (Chapter One). The improved resolution provided by the sequencing of this ancient mitochondrial genome is used to extend the phylogeny of human mtDNA sub-haplogroup to which the Uznyi Oleni Ostrov haplotype belongs. Scenarios for the origins and evolutionary history of this lineage are assessed.

Chapter Three: This chapter presents mtDNA data obtained from Iron Age horse-riding Scythians (2,300 – 2,800 yBP) of the Black Sea area (Rostov-on-Don region, south western Russia). By providing a genetic portrait of the Scythians, this work widens our knowledge from previous archaeological and anthropological studies about this nomadic population. The Scythian sites examined in Chapter Three are located at the eastern extremity of Europe and the western end of the central Eurasian Steppe. The study of Scythians provides new geographical and temporal elements about the prehistoric genetic influences from western and eastern Eurasia identified in Chapter One and contributes to the description of the mtDNA diversity in prehistoric nomadic populations of the central Eurasian Steppe.

Chapter Four: I investigate local genetic continuity between Bronze Age central Sardinians (3,200 – 3,400 yBP) and modern Sardinians. The geographical isolation of Sardinians was previously proposed as a cause for their status as European genetic outliers (Morelli et al., 2000). A similar approach is taken as for the Saami

people, using aDNA to date demographic events in the population history of Sardinians.

A general discussion of the results follows these four chapters. This section provides comments about how the difficulties associated with ancient human DNA have impacted the present work and how they have been dealt with. In the discussion chapter, I also attempt to place and discuss the ancient mtDNA data from this study within the wider context of Eurasian human mitochondrial diversity by comparing the new data with mtDNA data previously obtained from ancient and modern-day populations.

The conclusion section finally summarizes the study, emphasising its significance to the field, contribution to current knowledge and suggests future directions.

Table 1: List of samples examined in this study.

	Ancient North East Europeans				Scythians	Ancient Sardinians
Sites	Uznyi Oleni Ostrov	Popovo	Bolshoy Oleni Ostrov	Chalmny-Varre	various	Su Bittuleris/ Su Cannisoni
Location	Karelia, North West Russia	Archangelsk region, North West Russia	Kola Peninsula, North West Russia	Kola Peninsula, North West Russia	Rostov- on-Don, South West Russia	Central Sardinia
Age	~7,000 - 7,500 yBP	~7,000 - 7,500 yBP	3,500 yBP	200 – 300 yBP	2,300 – 2,800 yBP	3,200 – 3,400 yBP
Thesis Chapter	Chapters One and Two	Chapter One	Chapter One	Chapter One	Chapter Three	Chapter Four
Number of samples (individuals)	84 (42)	3 (3)	45 (23)	69 (42)	34 (17)	34 (34)
	201 (110)					
TOTAL	269 (161)					

yBP: years Before Present

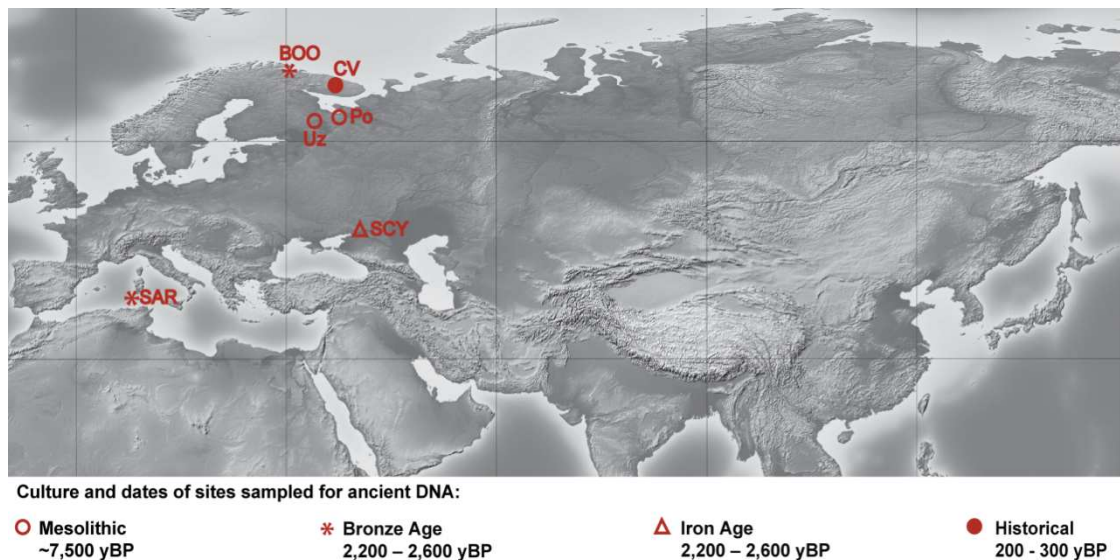


Figure 1: Map showing the geographical location of the six sites sampled for ancient DNA in this study.

Methodology

The mitochondrial diversity in the human remains examined here was characterized through the sequencing of 354 – 412 bp in HVR-I and the typing of 22 phylogenetically informative SNPs in the coding region, in experimental conditions limiting contamination by modern DNA. Several criteria for authentication of endogenous mtDNA sequences with regard to contamination or DNA damage were followed. Ancient mtDNA data was analysed using a range of statistical methods commonly applied to modern haplogroup and haplotype-based genetic data (e.g., principal component analysis, multi-dimensional scaling, haplotype sharing), as well as methods that account for diachronic sampling (i.e., at different points in time; coalescent simulations), some of which are described below.

HVR-I sequencing

The HVR-I of ancient individuals was sequenced between positions 15997 and 16409 (according to the rCRS) in three or four short overlapping fragments (Figure 2). The same system was used as described in Haak et al., 2005, Haak et al., 2008; and Haak et al., 2010. Amplicons range in size from 126 bp to 240 bp.

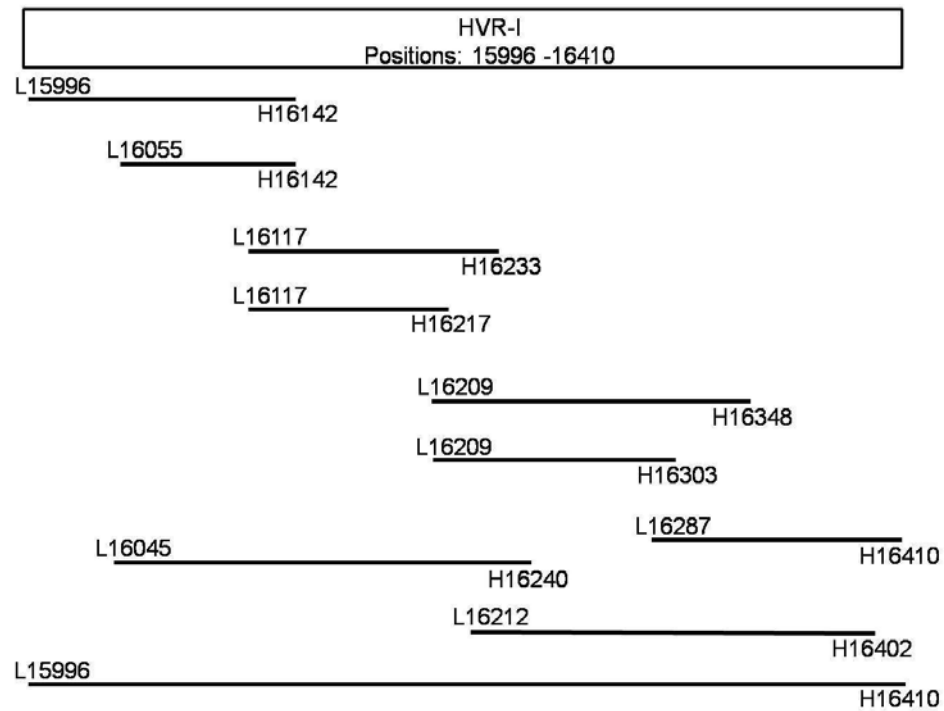


Figure 2: System of primers for HVR-I amplification and sequencing.

Typing of 22 coding region SNPs (GenoCore22 reaction)

The assessment of mtDNA haplogroups based on HVR-I sequences was refined and/or confirmed by typing 22 haplogroup diagnostic SNPs in the coding region of the mtDNA genome.

SNaPshot minisequencing has been previously applied to type SNPs in the mitochondrial genome (Brandstätter et al., 2003; Brandstätter et al., 2006; Grignani et al., 2006; Niederstätter et al., 2006; Haak et al., 2010), on autosomal chromosomes (Grimes et al., 2001; Dixon et al., 2005; Sanchez et al., 2006; Bouakaze et al., 2007) and on the Y-chromosome (Sanchez et al., 2003; Sanchez et al., 2005; Brion et al., 2005a; Brion et al., 2005b; Onofri et al., 2006; Bouakaze et al., 2009). These studies include work on modern, forensic and ancient DNA and have demonstrated the robustness, reliability and sensitivity of the SNaPshot minisequencing approach (Bouakaze et al., 2009).

In general, coding region SNPs were characterised using a multiplex PCR followed by a multiplexed Single Base Extension (SBE) reaction (GenoCore22, Haak et al., 2010) using the SNaPshot minisequencing technique (Applied Biosystems). This approach consists of a preliminary multiplex PCR amplification, in the present

case targeting 22 short fragments encompassing haplogroup-specific SNPs in the coding region (as defined in Behar et al., 2007; Figure 3). A SBE reaction follows, where oligonucleotides designed to hybridise one base pair upstream of each SNP of interest are used as primers for the addition of a single dideoxyribonucleoside triphosphate (ddNTP). Each ddNTP is labelled with a different fluorochrom to allow detection of the SNPs after fragment length-based electrophoresis separation of the 22 amplicons.

The GenoCoRe22 reaction is particularly relevant to the study of ancient mtDNA for several reasons. First, because the GenoCoRe22 targets 22 SNPs in a single reaction it is a time and cost-efficient method that uses very small amounts of valuable aDNA extracts. Because it limits the number of PCRs necessary to type 22 SNPs, GenoCoRe22 also reduces the risk of introducing modern contamination in the typing reactions (as opposed to 22 single reactions). As the multiplex is designed to target short DNA fragments spanning sizes from 60 bp to 80 bp, it is suitable for highly degraded DNA. Finally, the GenoCoRe22 reaction is used here as an internal control. The concordance between the haplogroup assessments independently obtained from the analysis of HVR-I sequences and SNaPshot profiles provides a key means to authenticate the aDNA results. In addition, because of the hierarchical nature of its design, i.e., typing of SNPs diagnostic of mtDNA macro-haplogroups, haplogroups and sub-haplogroups, the GenoCore22 reaction allows the detection of contamination and/or DNA damage through the observation of superimposed and/or phylogenetically inconsistent profiles.

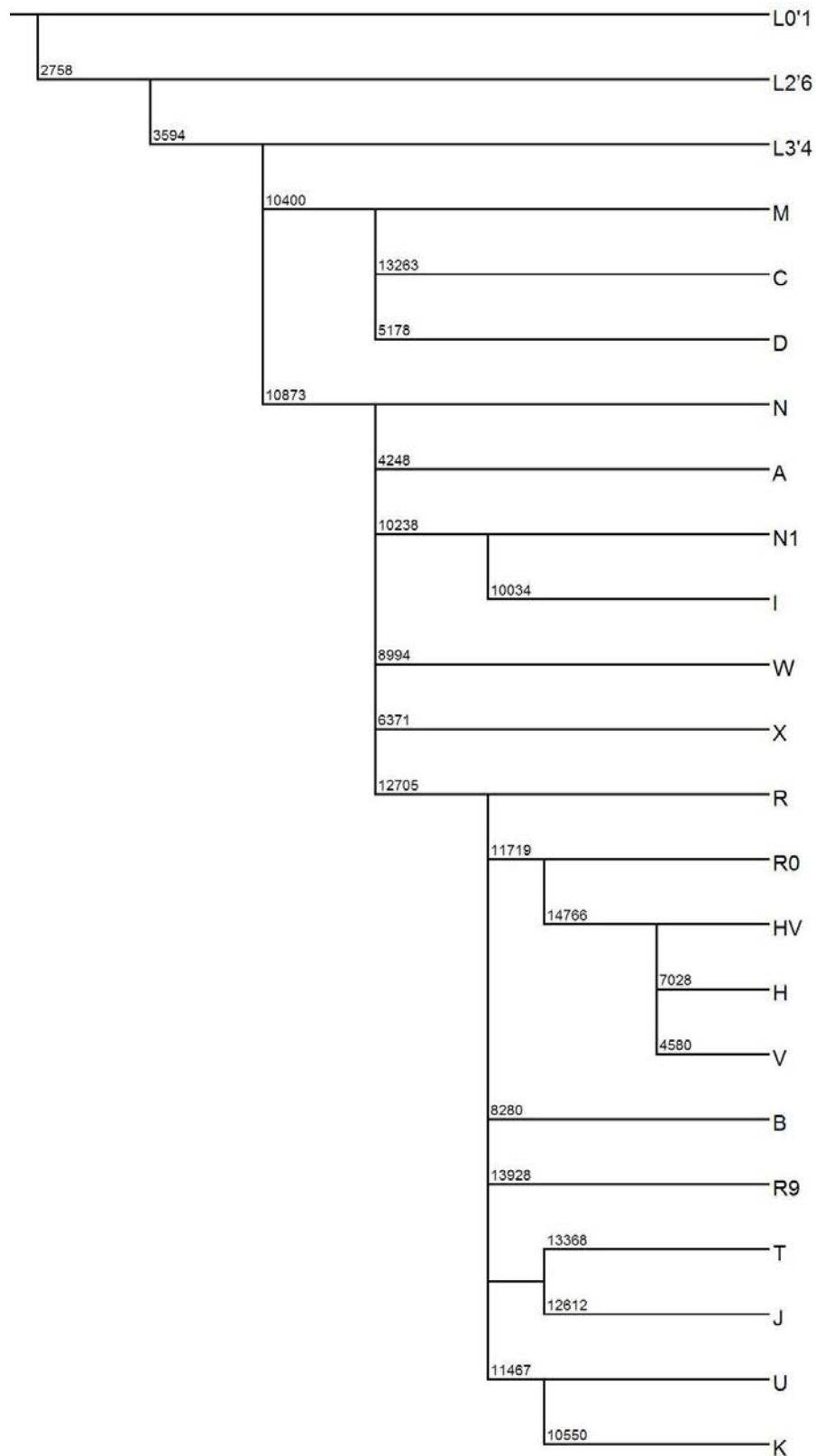


Figure 3: Schematic representation of the human mitochondrial phylogeny indicating the positions of the Single Nucleotide Polymorphisms typed in the GenoCore22 reaction.

Comparative genetic database

In the absence of large datasets of contemporaneous populations, the analyses of ancient genetic data were restricted to the published studies of equivalent markers in modern-day and rare ancient populations. However, ancient population analyses require access to extensive and reliable comparative databases. Ideally such a database should be regularly updated with new entries from the literature, but also contain the necessary information about samples/populations in terms of location, ethnicity, language, history. The reliable database should be free from common sequencing artefacts (Bandelt et al., 2006a; Bandelt et al., 2008) and also, for HVR-I sequence databases, contain information from the coding-region in order to determine/confirm the haplogroup assessment based on HVR-I sequences. In this regard, it is unfortunate that genetic data can sometimes be challenging to retrieve from publications due to the use of inconvenient formats and/or differing nomenclature.

The number of available human HVR-I sequences that have been published to date is quite significant; e.g., the comparative database used in the study of the ancient North East Europeans (Chapter One) and Scythians (Chapter Three) contained around 168,000 HVR-I haplotypes. The compilation of such databases represents a significant amount of work and unfortunately, despite several attempts (e.g., EMPOP), no universally agreed database is freely available. Instead, several databases are maintained privately. In the present study, I relied on collaborative agreements with laboratories specialised in population genetics of modern-day populations to access such databases (Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia and the Department of Genetics and Microbiology, University of Pavia). In contrast, to search information from complete mitochondrial genomes, the 'Phylotree' database (van Oven & Kayser, 2009), freely available online (<http://www.phylotree.org>, build 11, 7th February 2011, currently 8731 entries) and frequently updated, was found to be very reliable and useful and was extensively used throughout this work.

In addition to these databases I compiled a database of HVR-I sequences from ancient populations. The number of entries was significantly lower than for modern populations (516 entries) but the same quality checks were necessary.

To date, databases of HVR-I sequences for ancient populations, as well as databases of complete mitochondrial genomes are still geographically and temporally under-sampled, an issue that needs to be addressed in the coming years.

Principal Component Analysis

Haplogroup frequency data was visualised through Principal Component Analysis (PCA). Introduced by Pearson (1901), PCA is a form of multivariate analyses, a family of analyses that detect the internal structure among multiple variables, here the haplogroup frequencies. By identifying the relationships between the variables, PCA allows the visualisation of the data in a reduced number of dimensions, called 'components'. Here, components are constructed through a mathematical orthogonal transformation that converts the observed frequencies of potentially correlated haplogroups, into values of uncorrelated components. Components can be described, here, as the weighed combination of the haplogroup frequencies of the dataset. The first component is the component that accounts for most of the variability in the data and the following components are listed in descending order according to their variance. By reducing the dimensionality of the dataset, PCA allows the visualisation of patterns of -in this case - haplogroup frequencies in ancient and modern populations (Novembre et al., 2008).

Classical multidimensional scaling

I assessed similarities and dissimilarities among modern and ancient populations using classical multidimensional scaling (MDS). MDS was applied to visualise population differentiation as measured by fixation indices F_{ST} (Slatkin, 1995) calculated from haplogroup-frequency data. The MDS procedure allows the representation of the F_{ST} computed between populations in a two-dimensional graph (Borg & Groenen, 2005).

Haplotype sharing

Genetic links between comparative populations (ancient and modern) are detected here by elevated percentages of shared haplotypes, calculated by haplotype sharing analyses. In this study, haplotype sharing analyses were used to calculate the proportion of a particular haplotype detected in both an ancient population and a comparative modern-day population of equal size. These percentages of shared haplotypes were calculated, first, considering all the haplotypes sequenced in the population of interest, then, considering only non-basal haplotypes. In a phylogenetic tree, a basal haplotype represents the sequence at the root of a clade or of a sub-clade. Conversely, non-basal haplotypes are sequences derived from the basal haplotype, i.e.,

displaying additional mutations (Bandelt et al., 1995). Non-basal haplotypes can show geographical distributions that are more specific than those of basal haplotypes, and hence, are more informative.

Coalescence simulation: Bayesian Serial SimCoal and Approximate Bayesian Computation

Coalescence simulations (as described above in ‘Reconstructing past human population history using modern mitochondrial DNA’/‘The coalescent theory and coalescent simulations’) are performed using the program Bayesian Serial SimCoal (Anderson et al., 2005) and Approximate Bayesian Computation (Beaumont et al., 2002) in order to statistically test models of population history.

Expected outcomes

- (1) Obtaining reliable ancient human mtDNA datasets;
- (2) Detecting genetic affinities, at the haplogroup frequency or haplotypic levels, among ancient populations and/or between ancient and modern populations;
- (3) Providing answers to specific anthropological, historical, archaeological questions;
- (4) Discovering extinct past mtDNA diversity in the form of lineages that have not been detected in the currently available mtDNA database;
- (5) Identifying and timing past population processes, e.g., changes in population sizes, migrations, population replacements;
- (6) Providing elements to the reconstruction of the genetic history of European genetic outliers: the Saami and the Sardinians.

REFERENCES

1. Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V., Moral, P., Dugoujon, J.M., Roostalu, U., Loogväli, E.L., Kivisild, T., Bandelt, H.-J., Richards, M., Villems, R., Santachiara-Benerecetti, A.S., Semino, O., Torroni, A. (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75(5), 910-918.
2. Adachi, N., Shinoda, K., et al. (2009). Mitochondrial DNA analysis of Jomon skeletons from the Funadomari site, Hokkaido, and its implication for the origins of Native American. *Am J Phys Anthropol* 138(3), 255-265.
3. Adcock, G.J., Dennis, E.S., Easteal, S., Huttley, G.A., Jermini, L.S., Peacock, W.J., Thorne, A. (2001). Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins. *Proc Natl Acad Sci U S A* 16,98(2):537-42.
4. Adler, C. J., Haak, W., Donlon, D. & Cooper, A. (2010). Survival and recovery of DNA from ancient teeth and bones. *J Archaeol Sci* 38, 956-964.
5. Ammerman, A.J., Cavalli-Sforza, L.L. (1984). *The Neolithic Transition and the genetics of population in Europe*. Princeton Univ Press, Princeton.
6. Anderson, S., Bankier, A. T., et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290(5806), 457-465.
7. Anderson, C., Ramakrishnan, U., Chan, Y., Hadly, E. (2005). Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics* 21, 1733-1734.
8. Andrews, R., Kubacka, I., Chinnery, P., Lightowlers, R., Turnbull, D., Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23, 147.
9. Avise, J.C., J. Arnold, J, Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N.C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst* 18, 489-522.
10. Bailey, J.F., Richards, M.B., Macaulay, V.A., Colson, I.B., James, I.T., et al. (1996). Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. *Proc R Soc London Ser B* 263, 1467-73.
11. Bandelt, H.J., Forster, P., Sykes, B.C., Richards, M.B. (1995). Mitochondrial portraits of human populations using median networks. *Genetics* 141, 743-753.
12. Bandelt, H. J. (2005). Mosaics of ancient mitochondrial DNA: positive indicators of nonauthenticity. *Eur J Hum Genet* 13(10), 1106-1112.
13. Bandelt, H.-J., Macaulay, V., Richards, M. (2006a). *Mitochondrial DNA and the evolution of Homo sapiens*. Berlin: Springer-Verlag.
14. Bandelt, H.-J., Kong, Q.P., Yao, Y.G., Richards, M., Macaulay, V. (2006b). Estimation of mutation rates and coalescence times: Some caveats. In *Mitochondrial DNA and the evolution of Homo sapiens*, H.-J. Bandelt, V. Macaulay, and M. Richards, eds. (Berlin: Springer-Verlag), pp. 47-90.
15. Bandelt, H.J., Parson, W. (2008). Consistent treatment of length variants in the human mtDNA control region: a reappraisal. *Int J Legal Med.* 122(1):11-21.

16. Barbujani, G., Bertorelle, G., Chikhi, L. (1998). Evidence for Paleolithic and Neolithic Gene Flow in Europe. *Am J Hum Genet* 62, 488–491.
17. Barbujani, G., Goldstein, D.B. (2004). Africans and Asians abroad: Genetic Diversity in Europe. *Annu. Rev. Genomics Hum Genet* 5, 119–50.
18. Barnes, I., Young, J.P.W., Dobney, K.M. (2000). DNA-based identification of goose species from two archaeological sites in Lincolnshire. *J Archaeol Sci* 27, 91–100.
19. Barnes, I., Matheus, P., et al. (2002). Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* 295(5563), 2267–2270.
20. Behar, D.M., Rosset, S., Blue-Smith, J., Balanovsky, O., Tzur, S., Comas, D., Mitchell, R.J., Quintana-Murci, L., Tyler-Smith, C., Wells, R.S.; Genographic Consortium. (2007). The Genographic Project public participation mitochondrial DNA database. *PLoS Genet* 3(6):e104.
21. Beaumont, M.A., Zhang, W., Balding, D.J. (2002). Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035.
22. Belle, E. M., Ramakrishnan, U., et al. (2006). Serial coalescent simulations suggest a weak genealogical relationship between Etruscans and modern Tuscans. *Proc Natl Acad Sci U S A* 103(21), 8012–8017.
23. Belle, E. M., Benazzo, A., et al. (2009). Comparing models on the genealogical relationships among Neandertal, Cro-Magnon and modern Europeans by serial coalescent simulations. *Heredity* 102(3), 218–225.
24. Beja-Pereira, A., Luikart, G., England, P.R., Bradley, D.G., Jann, O.C., Bertorelle, G., Chamberlain, A.T., Nunes, T.P., Metodiev, S., Ferrand, N., Erhard, G. (2003). Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nat Genet* 35, 311–313.
25. Bensasson, D., Zhang, D., et al. (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol Evol* 16(6), 314–322.
26. Bertranpetit, J., Sala, J., Calafell, F., Underhill, P.A., Moral, P., Comas, D. (1995). Human mitochondrial DNA variation and the origin of Basques. *Ann Hum Genet* 59(Pt 1):63–81.
27. Bischoff, J. L. (2003). Neanderthals. *J Archaeol Sci* 30, 275.
28. Bocquet-Appel, J. P., Naji, S., Vander Linden, M., Kozłowski, J. K. (2009). Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of ¹⁴C dates. *J Archaeol Sci* 36: 807–820.
29. Borg, I., Groenen, P. (2005). *Modern Multidimensional Scaling: theory and applications* (2nd ed.). New York: Springer-Verlag. pp. 207–212.
30. Bouakaze, C., Keyser, C., Amory, S., Crubézy, E., Ludes, B. (2007). First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. *Int J Legal Med* 121(6):493–9.
31. Bouakaze, C., Keyser, C., Crubézy, E., Montagnon, D., Ludes, B. (2009). Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. *Int J Legal Med* 123(4):315–25.
32. Bramanti, B., Thomas, M., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M., Jankauskas, R., Kind, C., Lueth, F., Terberger, T., Hiller, J., Matsumara, S., Forster, P., Burger, J. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326, 137–140.

33. Brandstätter, A., Parsons, T.J., Parson, W. (2003). Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Legal Med* 117, 291–298.
34. Brandstätter, A., Salas, A., Niederstatter, H., Gassner, C., Carracedo, A., Parson, W., (2006). Dissection of mitochondrial superhaplogroup H using coding region SNPs. *Electrophoresis* 27, 2541–2550.
35. Briggs, A. W., Stenzel, U., et al. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* 104(37), 14616-14622.
36. Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Schmitz, R., Doronichev, V.B., Golovanova, L.V., de la Rasilla, M., Fortea, J., Rosas A., Pääbo, S. (2009). Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325(5938):318-21.
37. Brion, M., Sobrino, B., Blanco-Verea, A., Lareu, M.V., Carracedo, A. (2005a). Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *Int J Legal Med* 119, 10–15.
38. Brion M, Sanchez JJ, Balogh K et al (2005b) Introduction of a single nucleotide polymorphism-based “major Y-chromosome haplogroup typing kit” suitable for predicting the geographical origin of male lineages. *Electrophoresis* 26:4411–4420.
39. Brotherton, P., Endicott, P., Sanchez, J., Beaumont, M., Barnett, R., Austin, J., Cooper, A. (2007). Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35, 5717-5728.
40. Brown, W. M. (1980). Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc Natl Acad Sci U S A* 77(6), 3605-3609.
41. Brown, P., Sutikna, T., et al. (2004). A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* 431(7012), 1055-1062.
42. Burger, J., Hummel, S., et al. (1999). DNA preservation: a microsatellite-DNA study on ancient skeletal remains. *Electrophoresis* 20(8), 1722-1728.
43. Burger, J., Hummel, S., Pfeiffer, I., Herrmann, B. (2000). Palaeogenetic analysis of (pre)historic artifacts and its significance for anthropology. *Anthropol Anz* 58(1):69-76.
44. Burger, J., Kirchner, M., Bramanti, B., Haak, W., Thomas, M. G. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc Natl Acad Sci U S A* 104(10), 3736-42.
45. Campbell, K. L., Roberts, J. E., et al. (2010). Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. *Nat Genet* 42(6), 536-540.
46. Campos, P. F., Willerslev, E., et al. (2010). Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proc Natl Acad Sci U S A* 107(12), 5675-5680.
47. Cann, R.L., Stoneking, M., Wilson, A.C. (1987). Mitochondrial DNA and human evolution. *Nature* 325, 31-36.
48. Caramelli, D., Lalueza-Fox, C., Vernesi, C., Lari, M., Casoli, A., Mallegni, F., Chiarelli, B., Dupanloup, I., Bertranpetit, J., Barbujani, G., Bertorelle, G. (2003). Evidence for a genetic discontinuity between Neandertals and 24,000-

- year-old anatomically modern Europeans. *Proc Natl Acad Sci U S A* 100, 6593-6597.
49. Caramelli, D., Vernesi, C., Sanna, S., Sampietro, L., Lari, M., Castri, L., Vona, G., Floris, R., Francalacci, P., Tykot, R., Casoli, A., Bertranpetit, J., Lalueza-Fox, C., Bertorelle, G., Barbujani, G. (2007a). Genetic variation in prehistoric Sardinia. *Hum Genet* 122(3-4) 327-336.
 50. Caramelli, D., Lalueza-Fox, C., et al. (2007b). Genetic analysis of the skeletal remains attributed to Francesco Petrarca. *Forensic Sci Int* 173(1), 36-40.
 51. Caramelli, D., Milani, L., et al. (2008). A 28,000 years old Cro-Magnon mtDNA sequence differs from all potentially contaminating modern sequences. *PLoS One* 3(7), e2700.
 52. Charlesworth, B. (2009). Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 10(3), 195-205.
 53. Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC. (1995). Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* 57(1):133-49.
 54. Chikhi, L., Destro-Bisol, G., Pascali, V., Baravelli, V., Dobosz, M., Barbujani, G. (1998). Clinal variation in the nuclear DNA of Europeans. *Hum Biol* 70(4):643-57.
 55. Chikhi, L., Nichols, R., et al. (2002). Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci U S A* 99(17), 11008-11013.
 56. Collins, M.J., Penkman, K.E., Rohland, N., Shapiro, B., Dobberstein, R.C., Ritz-Timme, S., Hofreiter, M. (2009). Is amino acid racemization a useful tool for screening for ancient DNA in bone? *Proc Biol Sci* 276(1669):2971-7.
 57. Colson, I.B., Richards, M.B., Bailey, J.F., Sykes, B.C., Hedges, R.E.M. (1997). DNA analysis of seven human skeletons excavated from the terp of Wijnaldum. *J Archaeol Sci* 24, 911-17.
 58. Cooper, A. (1992). Removal of colourings, inhibitors of PCR, and the carrier effect of PCR contamination from ancient DNA samples. *Anc DNA Newslett* 1, 31-32.
 59. Cooper, A., Cooper, R.A. (1995). The Oligocene bottleneck and New Zealand biota: genetic record of a past environmental crisis. *Proc Biol Sci* 261(1362), 293-302.
 60. Cooper, A., Poinar, H.N. (2000). Ancient DNA: do it right or not at all. *Science* 289(5482), 1139.
 61. Cooper, A., Rambaut, A., Macaulay, V., Willerslev, E., Hansen, A.J., Stringer, C. (2001). Human origins and ancient human DNA. *Science* 292 (5522), 1655-1656.
 62. Cooper A., Willerslev, E. (2005). Ancient DNA. *Proc Biol Sci* 272(1558), 3-16.
 63. Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (1st ed.). London: John Murray.
 64. De Benedetto, G., Nasidze, I.S., Stenico, M., Nigro, L., Krings, M., Lanzinger, M., Vigilant, L., Stoneking, M., Pääbo, S., Barbujani, G. (2000). Mitochondrial DNA sequences in prehistoric human remains from the Alps. *Eur J Hum Genet* 8(9):669-77.

65. Debruyne, R., Barriel, V., et al. (2003). Mitochondrial cytochrome b of the Lyakhov mammoth (Proboscidea, Mammalia), new data and phylogenetic analyses of Elephantidae. *Mol Phylogenet Evol* 26(3), 421-434.
66. DeSalle, R., Gatesy, J., et al. (1992). DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science* 257(5078), 1933-1936.
67. DeSalle R. (1994). Implications of ancient DNA for phylogenetic studies. *Experientia* 50, 543-550.
68. Dixon, L.A., Murray, C.M., Archer, E.J., Dobbins, A.E., Koumi, P., Gill P. (2005). Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. *Forensic Sci Int* 154, 62–77.
69. Dolukhanov, P. (1993). Foraging and farming groups in north-eastern and north-western Europe: identity and interaction. In *Cultural Transformations and Interactions in Eastern Europe*, J. Chapman and P. Dolukhanov, eds. (Aldershot: Avebury).
70. Douzery, E. J., Delsuc, F., et al. (2003). Local molecular clocks in three nuclear genes: divergence times for rodents and other mammals and incompatibility among fossil calibrations. *J Mol Evol* 57 Suppl 1, S201-213.
71. Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22(5):1185-92.
72. Drummond, A., Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7, 214.
73. Edwards, C.J., Magee, D.A., Park, S.D., McGettigan, P.A., Lohan, A.J., Murphy, A., Finlay, E.K., Shapiro, B., Chamberlain, A.T., Richards, M.B., Bradley, D.G., Loftus, B.J., MacHugh, D.E. (2010). A complete mitochondrial genome sequence from a mesolithic wild aurochs (*Bos primigenius*). *PLoS One* 17;5(2):e9255.
74. Endicott, P., Metspalu, M., et al. (2006). Multiplexed SNP typing of ancient DNA clarifies the origin of Andaman mtDNA haplogroups amongst South Asian tribal populations. *PLoS One* 1, e85.
75. Endicott, P., Ho, S.Y. (2008). A Bayesian evaluation of human mitochondrial substitution rates. *Am J Hum Genet* 82(4), 895-902.
76. Endicott, P., Ho, S.Y., Metspalu, M., Stringer, C. (2009a). Evaluating the mitochondrial timescale of human evolution. *Trends Ecol Evol* 24(9), 515-9.
77. Endicott, P., Sanchez, J. J., et al. (2009b). 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.
78. Ermini, L., Olivieri, C., Rizzi, E., Corti, G., Bonnal, R., Soares, P., Luciani, S., Marota, I., De Bellis, G., Richards, M.B., Rollo, F. (2008). Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr Biol* 18, 1687-1693.
79. Excoffier, L., Novembre, J., Schneider, S. (2000). SimCoal: a general coalescent program for simulation of molecular data in interconnected populations with arbitrary demography. *J Hered* 91, 506–509.
80. Finlayson, C., Pacheco, F.G., Rodríguez-Vidal, J., Fa, D.A., Gutierrez López, J.M., Santiago Pérez, A., Finlayson, G., Allue, E., Baena Preysler, J., Cáceres, I., Carrión, J.S., Fernández Jalvo, Y., Glead-Owen, C.P., Jimenez Espejo, F.J., López, P., López Sáez, J.A., Riquelme Cantal, J.A., Sánchez Marco, A.,

- Guzman, F.G., Brown, K., Fuentes, N., Valarino, C.A., Villalpando, A., Stringer, C.B., Martinez Ruiz, F., Sakamoto, T. (2006). Late survival of Neanderthals at the southernmost extreme of Europe. *Nature* 443(7113), 850-3.
81. Fisher, R.A. (1930). *The Genetical Theory of Natural Selection*, Clarendon Press, Oxford.
 82. Forster, P., Harding, R., Torroni, A., Bandelt, H. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59, 935-945.
 83. Gamble, C., Davies, W., Pettitt, P., and Richards, M. (2004). Climate change and evolving human diversity in Europe during the last glacial. *Phil. Trans. R. Soc. B.* 359, 243–253.
 84. Gamble, C., Davies, W., Pettitt, P., Hazelwood, L., and Richards, M. (2005). The archaeological and genetic foundations of the European population during the Late Glacial: implications for ‘agricultural thinking’. *Camb Archaeol J* 15, 193–223.
 85. Ghirotto, S., Mona, S., Benazzo, A., Paparazzo, F., Caramelli, D., Barbujani, G. (2010). Inferring genealogical processes from patterns of Bronze-Age and modern DNA variation in Sardinia. *Mol Biol Evol* 27, 875-886.
 86. Gilbert, M. T., Hansen, A.J., Willerslev, E., Rudbeck, L., Barnes, I., Lynnerup, N., Cooper, A. (2003). Characterization of genetic miscoding lesions caused by postmortem damage. *Am J Hum Genet* 72(1), 48-61.
 87. Gilbert, M.T., Wilson, A.S., Bunce, M., Hansen, A.J., Willerslev, E., Shapiro, B., Higham, T.F., Richards, M.P., O'Connell, T.C., Tobin, D.J., Janaway, R.C., Cooper, A. (2004a). Ancient mitochondrial DNA from hair. *Curr Biol* 14, R463-464.
 88. Gilbert, M. T., Cuccui, J., et al. (2004b). Absence of *Yersinia pestis*-specific DNA in human teeth from five European excavations of putative plague victims. *Microbiology* 150 (Pt 2), 341-354.
 89. Gilbert, M.T., Bandelt, H.J., Hofreiter, M., Barnes, I. (2005a). Assessing ancient DNA studies. *Trends Ecol Evol* 20, 541-544.
 90. Gilbert, M.T.P., Rudbeck, L., Willerslev, E., Hansen, A.J., Smith, C., Penkman, E.H., Prangeberg, K., Nielsen-Marsh, C.M., Jans, M.E., Arthur, P., Lynnerup, N., Turner-Walker, G., Biddle, M., Kjølbye-Biddle, B., Collins, M. (2005b). Biochemical and physical correlates of DNA contamination in archaeological bones and teeth excavated at Matera, Italy. *J Archaeol Sci* 32, 785-793.
 91. Gilbert, M. T., Binladen, J., et al. (2007). Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Res* 35(1), 1-10.
 92. Gilbert, M.T., Kivisild, T., Grønnow, B., Andersen, P.K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Götherström, A., Campos, P.F., Rasmussen, M., Metspalu M., Higham, T.F., Schwenniger, J.L., Nathan, R., De Hoog, C.J., Koch, A., Møller, L.N., Andreasen, C., Medgaard, M., Villems, R., Bendixen, C., Willerslev, E. (2008). Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* 320, 1787-1789.
 93. Goldman, N. Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 11(5), 725-736.
 94. Green, R. E., Malaspinas, A. S., et al. (2008). A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134(3), 416-426.

95. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., Hansen, N.F., Durand, E.Y., Malaspinas, A.S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H.A., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V.B., Golovanova, L.V., Lalueva-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R.W., Johnson, P.L., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science* 328(5979), 710-22.
96. Grignani, P., Peloso, G., Achilli, A et al. (2006). Subtyping mtDNA haplogroup H by SNaPshot minisequencing and its application in forensic individual identification. *Int J Legal Med* 120, 151–156.
97. Grimes, E.A., Noake, P.J., Dixon, L., Urquhart, A. (2001) Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic Sci Int* 122, 124–129.
98. Grousset, R. (1970) *The Empire of the Steppes: History of Central Asia*. Ed. Rutgers University Press.
99. Guglielmino, C.R., Piazza, A., Menozzi, P., Cavalli-Sforza, L.L. (1990). Uralic genes in Europe. *Am J Phys Anthropol* 83, 57-68.
100. Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K.W., Burger, J. (2005). Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310, 1016-1018.
101. Haak, W., Brandt, G., et al. (2008). Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. *Proc Natl Acad Sci U S A* 105(47), 18226-18238.
102. Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., Der Sarkissian, C.S., Brandt, G., Schwarz, C., Nicklisch, N., Dresely, V., Fritsch, B., Balanovska, E., Villems, R., Meller, H., Alt, K.W., Cooper, A., Genographic consortium. (2010). Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. *PLoS Biol* 8, e1000536.
103. Hadly, E. A., Ramakrishnan, U., et al. (2004). Genetic response to climatic change: insights from ancient DNA and phylochronology. *PLoS Biol* 2(10), e290.
104. Haensch, S., Bianucci, R., et al. (2010). Distinct clones of *Yersinia pestis* caused the black death. *PLoS Pathog* 6(10).
105. Hagelberg, E., Sykes, B., Hedges, R. (1989). Ancient bone DNA amplified. *Nature*. 30;342(6249):485.
106. Hagelberg, E., Bell, L.S., Allen, T., Boyde, A., Jones, S.J., Clegg, J.B. (1991). Analysis of ancient bone DNA: techniques and applications. *Philos Trans R Soc Lond B Biol Sci* 333, 399-407.
107. Hall, B.G. (2001). *Phylogenetic trees made easy: a how-to for molecular biologists*. Third Edition. Sinauer Associates.
108. Handt, O., Höss, M., Krings, M., Pääbo, S. (1994). Ancient DNA: methodological challenges. *Experientia* 50 (6), 524-529.
109. Handt, O., Krings, M., Ward, R.H., Pääbo, S. (1996). The retrieval of ancient human DNA sequences. *Am J Hum Genet* 59(2), 368-376.

110. Hansen, A., Willerslev, E., Wiuf, C., Mourier, T., Arctander, P. (2001). Statistical evidence for miscoding lesions in ancient DNA templates. *Mol Biol Evol* 18(2), 262-5.
111. Harvati, K; (2007). *Handbook of Paleoanthropology*, W. Henke, I. Tattersall, Eds. (Springer, Berlin, 2007), vol. 3, pp.1717–1748.
112. Hasegawa, M., Kishino, H., et al. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22(2), 160-174.
113. Hawass, Z., Gad, Y. Z., et al. (2010). Ancestry and pathology in King Tutankhamun's family. *JAMA* 303(7), 638-647.
114. Helgason, A., Sigurethardóttir, S., Gulcher, J.R., Ward, R., Stefánsson, K. (2000). mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet* 66, 999-1016.
115. Helgason, A., Pálsson, S., Lalueza-Fox, C., Ghosh, S., Sigurethardottir, S., Baker, A., Hrafnkelsson B., Arnadottir L., Thornorsteinsdottir U., Stefansson K. (2007). A Statistical Approach to Identify Ancient Template DNA. *J Mol Evol* 65, 92-102.
116. Helgason, A., Lalueza-Fox, C., Ghosh, S., Sigurethardóttir, S., Sampietro, M.L., Gigli, E., Baker, A., Bertranpetit, J., Arnadóttir, L., Thornorsteinsdottir, U., Stefánsson, K. (2009). Sequences from first settlers reveal rapid evolution in Icelandic mtDNA pool. *PLoS Genet* 5(1):e1000343.
117. Herrnstadt, C., Elson, J., L., Fahy, E., Preston, G., Turnbull, D., M., Anderson, C., Ghosh, S., S., Olefsky, J., M., Beal, M., F., Davis, R., E., Howell, N. (2002). Reduced-median-network analysis of complete mitochondrial DNA coding region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* 70, 1152-1179.
118. Heyer, E., Zietkiewicz, E., et al. (2001). Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. *Am J Hum Genet* 69(5), 1113-1126.
119. Heyn, P., Stenzel, U., et al. (2010). Road blocks on paleogenomes--polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA. *Nucleic Acids Res* 38(16), e169.
120. Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A., Wilson, A.C. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312(5991):282-4.
121. Higuchi, R., Fockler, C., Dollinger, G., Watson, R. (1993). Kinetic PCR analysis: realtime monitoring of DNA amplification reactions. *Biotechnology (N Y)*.11, 1026-1030.
122. Ho, S.Y., Heupink, T.H., Rambaut, A., Shapiro, B. (2007). Bayesian estimation of sequence damage in ancient DNA. *Mol Biol Evol* 24(6), 1416-22.
123. Ho, S., Endicott, P. (2008). The crucial role of calibration in molecular date estimates for the peopling of the Americas. *Am J Hum Genet* 83, 142-146; author reply 146-147.
124. Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A.A., Pääbo, S. (2001a). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res* 29, 4793-4799.
125. Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., Pääbo, S. (2001b). Ancient DNA. *Nat Rev Genet* 2(5), 353-359.
126. Höss, M., Jaruga, P., Zastawny, T.H., Dizdaroglu, M., Pääbo, S. (1996). DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res* 24, 1304-1307.

127. Hudson, R.R. (1983). Testing the constant-rate neutral allele model with protein sequence data. *Evolution* 37, 203–207.
128. Hudson, R.R. (1990). Gene genealogies and the coalescent process. *Oxf Surv Evol Biol* 71–44.
129. Hummel, S., Schmidt, D., Kahle, M., Herrmann, B. (2002). AB0 blood group genotyping of ancient DNA by PCR-RFLP. *Int J Legal Med* 116(6):327-33.
130. Ingman, M., Kaessmann, H., Pääbo, S., Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* 408(6813), 708-13.
131. Izagirre, N., de la Rúa, C. (1999). An mtDNA analysis in ancient Basque populations: implications for haplogroup V as a marker for a major paleolithic expansion from southwestern Europe. *Am J Hum Genet* 65(1):199-207.
132. Jaenicke-Després, V., Buckler, E.S., Smith, B.D., Gilbert, M.T., Cooper, A., Doebley, J., Pääbo, S. (2003). Early allelic selection in maize as revealed by ancient DNA. *Science* 302(5648):1206-8.
133. Jans, M.M.E., Nielsen-Marsh, C.M., Smith, C.I., Collins, M.J., Kars, H. (2004). Characterisation of microbial attack on archaeological bone. *J Archaeolog Sci* 31, 87–95.
134. Jhaes, E., Toprak, K., Vanderheyden, N., Pfeiffer, H., Cassiman, J.J., Brinkmann, B., Decorte, R. (2001). Pitfalls in the analysis of mitochondrial DNA from ancient specimens and the consequences for forensic DNA analysis: the historical case of the putative heart of Louis XVII. *Int J Legal Med* 115(3):135-41.
135. Jobling, M.A., Hurles, M.E., Tyler-Smith, C. (2004) *Human Evolutionary Genetics: origins, peoples and disease*. London/New York: Garland Science Publishing, 523 pp.
136. Jukes, T.H., Cantor, C.R. (1969). Evolution of protein molecules. In Munro, H.N..*Mammalian protein metabolism*. New York: Academic Press. pp. 21–123.
137. Kemp, B. M., Tung, T. A., et al. (2009). Genetic continuity after the collapse of the Wari empire: mitochondrial DNA profiles from Wari and post-Wari populations in the ancient Andes. *Am J Phys Anthropol* 140(1), 80-91.
138. Keyser, C., Bouakaze, C., Crubézy, E., Nikolaev, V., 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.
139. 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.
140. Kingman, J.F.C. (1982). On the Genealogy of Large Populations. *J App. Prob* 19A, 27–43.
141. Kingman, J.F.C. (2000). Origins of the coalescent 1974–1982. *Genetics* 156, 1461–1463.
142. Kivisild, T., Shen, P., Wall, D.P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P.A., Scharfe, C., Torroni, A., Scozzari, R., Modiano, D., Coppa, A., de Knijff, P., Feldman, M., Cavalli-Sforza, L.L., Oefner, P.J. (2006). The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172(1), 373-87.
143. Knight, A., Zhivotovsky, L. A., et al. (2004). Molecular, forensic and haplotypic inconsistencies regarding the identity of the Ekaterinburg remains. *Ann Hum Biol* 31(2), 129-138.

144. Kolman, C. J., Tuross, N. (2000). Ancient DNA analysis of human populations. *Am J Phys Anthropol* 111(1), 5-23.
145. Koon, H.E.C., Nicholson, R.A., Collins, M.J. (2003). A practical approach to the identification of low temperature heated bone using TEM. *J Archaeolog Sci* 13, 1393-99.
146. Kozłowski J., Bandi H.G. (1984). The paleohistory of circumpolar arctic colonization. *Arctic* 37(4) 359-372.
147. Krause, J., Dear, P. H., et al. (2006). Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* 439(7077), 724-727.
148. Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R.E., Burbano, H.A., Hublin, J.J., Hänni, C., Fortea, J., de la Rasilla, M., Bertranpetit, J., Rosas, A., Pääbo, S. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. *Curr Biol* 17(21):1908-12.
149. Krause, J., Fu, Q., et al. (2010a). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464(7290), 894-897.
150. Krause, J., Briggs, A., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A., Pääbo, S. (2010b). A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol* 20, 231-236.
151. Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M., Pääbo, S. (1997). Neandertal DNA sequences and the origin of modern humans. *Cell* 90, 19-30.
152. Krings, M., Capelli, C., et al. (2000). A view of Neandertal genetic diversity. *Nat Genet* 26(2), 144-146.
153. Lacan, M., Keyser, C., Ricaut, F.X., Brucato, N., Duranthon, F., Guilaïne, J., Crubézy, E., Ludes, B. (2011). Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc Natl Acad Sci U S A* 108(24), 9788-9791.
154. Lahermo, P., Sajantila, A., Sistonen, P., Lukka, M., Aula, P., Peltonen, L., Savontaus, M.L. (1996). The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. *Am J Hum Genet* 58(6):1309-22.
155. Lalueza-Fox, C., Sampietro, M., Gilbert, M., Castri, L., Facchini, F., Pettener, D., Bertranpetit, J. (2004). Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians. *Proc Biol Sci* 271, 941-947.
156. Lalueza-Fox, C., Sampietro, M.L., Caramelli, D., Puder, Y., Lari, M., Calafell, F., Martinez-Maza, C., Bastir, M., Fortea, J., de la Rasilla, M., Bertranpetit, J., Rosas, A. (2005). Neandertal evolutionary genetics, mitochondrial DNA data from the Iberian peninsula. *Mol Biol Evol* 22, 1077-1081.
157. Lalueza-Fox, C., Römpler, H., Caramelli, D., Stäubert, C., Catalano, G., Hughes, D., Rohland, N., Pilli, E., Longo, L., Condemi, S., de la Rasilla, M., Fortea, J., Rosas, A., Stoneking, M., Schöneberg, T., Bertranpetit, J., Hofreiter, M. (2007). A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318, 1453-1455.
158. Lalueza-Fox, C., Gigli, E., et al. (2010). Genetic analysis of the presumptive blood from Louis XVI, king of France. *Forensic Sci Int Genet*.
159. Lalueza-Fox, C., Rosas, A., et al. (2011). Genetic evidence for patrilocal mating behavior among Neandertal groups. *Proc Natl Acad Sci U S A* 108(1), 250-253.
160. Lappalainen, T., Laitinen, V., Salmela, E., Andersen, P., Huoponen, K., Savontaus, M.L., Lahermo, P. (2008). Migration waves to the Baltic Sea

- region. *Ann Hum Genet* 72, 337-348.
161. Larson, G., Liu, R., et al. (2010). Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proc Natl Acad Sci U S A* 107(17), 7686-7691.
 162. Laval, G., Excoffier, L. (2004). SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*. 20(15), 2485-7.
 163. Leonard, J.A., Shanks, O., Hofreiter, M., Kreuz, E., Hodges, L., Ream, W., Wayne, R.K., and Fleischer, R.C. (2006). Animal DNA in PCR reagents plagues ancient DNA research. *J Archaeol Sci* 34, 1361-1366.
 164. Leonard, J.A., Shanks, O., Hofreiter, M., Kreuz, E., Hodges, L., Ream, W., Wayne, R.K., Fleischer, R.C. (2007). Animal DNA in PCR reagents plagues ancient DNA research. *J Arch Sci* 34(9),1361-1366.
 165. Lewontin, R. (1972). The Apportionment of Human Diversity. *Evol Biol* 6, 391–398.
 166. Lindahl, T. (1993). Recovery of antediluvian DNA. *Nature* 365(6448), 700.
 167. Liu, W., Jin, C. Z., et al. (2010). Human remains from Zhirendong, South China, and modern human emergence in East Asia. *Proc Natl Acad Sci U S A* 107(45), 19201-19206.
 168. Loreille, O., Orlando, L., et al. (2001). Ancient DNA analysis reveals divergence of the cave bear, *Ursus spelaeus*, and brown bear, *Ursus arctos*, lineages. *Curr Biol* 11(3), 200-203.
 169. Ludwig, A., Pruvost, M., Reissmann, M., Benecke, N., Brockmann, G.A., Castaños, P., Cieslak, M., Lippold, S., Llorente, L., Malaspinas, A.S., Slatkin, M., Hofreiter, M. (2009). Coat color variation at the beginning of horse domestication. *Science* 324(5926):485.
 170. Maca-Meyer, N., González, A.M., Larruga, J.M., Flores, C., Cabrera, V.M. (2001). Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2, 13.
 171. Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonn -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(1), 232-49.
 172. Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., Taha, A., Shaari, N.K., Raja, J.M., Ismail, P., Zainuddin, Z., Goodwin, W., Bulbeck, D., Bandelt, H.J., Oppenheimer, S., Torroni, A., Richards, M. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308(5724), 1034-6.
 173. Malmstr m, H., Svensson, E.M., Gilbert, M.T., Willerslev, E., G therstr m, A., Holmlund, G. (2007). More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA. *Mol Biol Evol* 24, 998-1004. behavior to identify authentic ancient human DNA. *Mol Biol Evol* 24, 998-1004.
 174. Malmstr m, H., Gilbert, M., Thomas, M., Brandstr m, M., Stor , J., Molnar, P., Andersen, P., Bendixen, C., Holmlund, G., G therstr m, A., Willerslev, E. (2009). Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. *Curr Biol* 19, 1758-1762.

175. Malmström, H., Linderholm, A., et al. (2010). High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. *BMC Evol Biol* 10, 89.
176. Malmström, H., Vretemark, M., Tillmar, A., Durling, M.B., Skoglund, P., Gilbert, M.T., Willerslev, E., Holmlund, G., Götherström, A. (2011). Finding the founder of Stockholm - A kinship study based on Y-chromosomal, autosomal and mitochondrial DNA. *Ann Anat*.
177. Malyarchuk, B., Grzybowski, T., Derenko, M., Perkova, M., Vanecek, T., Lazur, J., Gornolcak, P., Tsybovsky, I. (2008). Mitochondrial DNA phylogeny in Eastern and Western Slavs. *Mol. Biol. Evol.* 25, 1651–1658.
178. Marchant, J. (2011). Ancient DNA: Curse of the Pharaoh's DNA. *Nature* 472(7344), 404-406.
179. Margulis, L., (1981). *Symbiosis in cell evolution*. W. H. Freeman and Company.
180. Mellars, P. (2006). A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature* 439(7079), 931-935.
181. Meyer, S., Weiss, G., et al. (1999). Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152(3), 1103-1110.
182. Meyer, E., Wiese, M., et al. (2000). Extraction and amplification of authentic DNA from ancient human remains. *Forensic Sci Int* 113(1-3), 87-90.
183. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D. M., Gilbert, M. T., Endicott, P., Mastana, S., Papiha, S. S., 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.
184. Miller, W., Drautz, D. I., Ratan, A., Pusey, B., Qi, J., Lesk, A.M., Tomsho, L.P., Packard, M.D., Zhao, F., Sher, A., Tikhonov, A., Raney, B., Patterson, N., Lindblad-Toh, K., Lander, E.S., Knight, J.R., Irzyk, G.P., Fredrikson, K.M., Harkins, T.T., Sheridan, S., Pringle, T., Schuster, S.C. (2008). Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456(7220), 387-390.
185. Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., Sukernik, R.I., Olckers, A., Wallace, D.C. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A* 100, 171-176.
186. Mitchell, D., Willerslev, E., et al. (2005). Damage and repair of ancient DNA. *Mutat Res* 571(1-2), 265-276.
187. Morelli L., Grosso M.G., Vona G., Varesi L., Torroni A., Francalacci P. (2000). Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia. *Hum Biol* 72, 585-595.
188. Mourant, A.E., Kopec, A.C., Domaniewska-Sobczak K. (1978). *Blood Groups and Diseases*. Oxford, England: Oxford University Press.
189. Mullis, K. B., Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 155, 335-350.
190. Nei, M., Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3(5), 418-426.
191. Niederstatter, H., Coble, M.D., Grubwieser, P., Parsons, T.J., Parson, W. (2006). Characterization of mtDNA SNP typing and mixture ratio assessment with

- simultaneous real-time PCR quantification of both allelic states. *Int J Legal Med* 120, 18–23.
192. Nielsen-Marsh, C.M., Hedges, R.E.M., Mann, T., Collins, M.J. (2000). A preliminary investigation of the application of differential scanning calorimetry to the study of collagen degradation in archaeological bone. *Thermochim Acta* 365, 129–39.
 193. Nogués-Bravo, D., Rodríguez, J., et al. (2008). Climate change, humans, and the extinction of the woolly mammoth. *PLoS Biol* 6(4), e79.
 194. Noonan, J.P., Hofreiter, M., Smith, D., Priest, J.R., Rohland, N., Rabeder, G., Krause, J., Detter, J.C., Pääbo, S., Rubin, E.M. (2005). Genomic sequencing of Pleistocene cave bears. *Science* 309, 597–599.
 195. Nordborg, M. (2001). Coalescent theory. In D. J. Balding, M. J. Bishop, and C. Cannings, editors, *Handbook of Statistical Genetics*, John Wiley & Sons, Inc., Chichester, U.K., p179–212.
 196. Novembre, J., Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nat Genet* 40(5), 646–649.
 197. O’Connell, J.F., Allen, F.J. (2004). Dating the colonization of Sahul (Pleistocene Australia-New Guinea): A review of recent research. *J Arch Sci* 31, 835–853.
 198. Onofri, V., Alessandrini, F., Turchi, C., Pesaresi, M., Buscemi, L., Tagliabracchi, A. (2006). Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. *Forensic Sci Int* 157, 23–35.
 199. Orlando, L., Darlu, P., Toussaint, M., Bonjean, D., Otte, M., Hänni, C. (2006). Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. *Curr Biol* 16, R400–402.
 200. Oskam, C. L., Haile, J., et al. (2010). Fossil avian eggshell preserves ancient DNA. *Proc Biol Sci* 277(1690), 1991–2000.
 201. Ottoni, C., Ricaut, F. X., et al. (2011). Mitochondrial analysis of a Byzantine population reveals the differential impact of multiple historical events in South Anatolia. *Eur J Hum Genet* 19(5), 571–576.
 202. Ovchinnikov, I.V., Götherström, A., Romanova, G.P., Kharitonov, V.M., Lidén, K., Goodwin, W. (2000). Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404, 490–493.
 203. Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314, 644–645.
 204. Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci U S A* 86, 1939–1943.
 205. Pääbo, S. (2000). Of bears, conservation genetics, and the value of time travel. *Proc Natl Acad Sci U S A* 97(4), 1320–1321.
 206. Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annu Rev Genet* 38, 645–679.
 207. Parsons, T. J., Muniec, D. S., Sullivan K. et al. (1997). A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* 15, 363–368.
 208. Passarino, G., Semino, O., Bernini, L.F., Santachiara-Benerecetti, A.S. (1996). Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. *Am J Hum Genet* 59(4), 927–34.
 209. Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine* 2 (6), 559–572.

210. Pereira, L., Richards, M., Goios, A., Alonso, A., Albarrán, C., Garcia, O., Behar, D.M., Gölge, M., Hatina, J., Al-Gazali, L., Bradley, D.G., Macaulay, V., Amorim, A. (2005). High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res* 15(1), 19-24.
211. Pliss, L., Tambets, K., Loogväli, E.L., Pronina, N., Lazdins, M., Krumina, A., Baumanis, V., Villems, R. (2006). Mitochondrial DNA portrait of Latvians: towards the understanding of the genetic structure of Baltic-speaking populations. *Ann Hum Genet* 70, 439-458.
212. Poinar, G. O., Waggoner, B. M., et al. (1993). Terrestrial soft-bodied protists and other microorganisms in triassic amber. *Science* 259(5092), 222-224.
213. Poinar, H.N., Hoss, M., Bada, J.L., Pääbo, S. (1996). Amino acid racemization and the preservation of ancient DNA. *Science* 272, 864-866.
214. Poinar, H. N., Hofreiter, M., et al. (1998). Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* 281(5375), 402-406.
215. Poinar, H.N., Stankiewicz, B.A. (1999). Protein preservation and DNA retrieval from ancient tissues. *Proc Natl Acad Sci USA* 96, 8426-31.
216. Poinar, H., Kuch, M., McDonald, G., Martin, P., Pääbo, S. (2003). Nuclear gene sequences from a late pleistocene sloth coprolite. *Curr Biol* 13, 1150-1152.
217. Price, T. D., ed. (2000) *Europe's first farmers*. Cambridge: Cambridge University Press.
218. Pruvost, M., Schwarz, R., Correia, V.B., Champlot, S., Braguier, S., Morel, N., Fernandez-Jalvo, Y., Grange, T., Geigl, E.M. (2007). Freshly excavated fossil bones are best for amplification of ancient DNA. *Proc Natl Acad Sci U S A* 104, 739-744.
219. Pruvost, M., Schwarz, R., Bessa Correia, V., Champlot, S., Grange, T., Geigl, E-M. (2008). DNA diagenesis and palaeogenetic analysis: critical assessment and methodological progress. *Paleo Paleo Paleo* 266, 211-219.
220. Ramakrishnan, U., Hadly E. A. (2009). Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Mol Ecol* 18(7), 1310-1330.
221. Raoult, D., Aboudharam, G., et al. (2000). Molecular identification by suicide PCR of *Yersinia pestis* as the agent of medieval black death. *Proc Natl Acad Sci U S A* 97(23), 12800-12803.
222. Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J., S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., Bertalan, M., Nielsen, K., Gilbert, M., T., Wang, Y., Raghavan, M., Campos, P. F., Kamp, H., M., Wilson, A. S., Gledhill, A., Tridico, S., Bunce, M., Lorenzen, E. D., Binladen, J., Guo, X., Zhao, J., Zhang, X., Zhang, H., Li, Z., Chen, M., Orlando, L., Kristiansen, K., Bak, M., Tommerup, N., Bendixen, C., Pierre, T. L., Grønnow, B., Meldgaard, M., Andreasen, C., Fedorova, S. A., Osipova, L. P., Higham, T. F., Ramsey, C. B., Hansen, T. V., Nielsen, F. C., Crawford, M. H., Brunak, S., Sicheritz-Pontén, T., Villems, R., Nielsen, R., Krogh, A., Wang, J., Willerslev, E. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463(7282), 757-62.
223. Rawlence, N. J., Wood, J. R., et al. (2009). DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proc Biol Sci* 276(1672), 3395-3402.
224. Reich, D., Green, R. E., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468(7327), 1053-1060.

225. Richards, M., C  rte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H., Sykes, B. (1996). Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59, 185-203.
226. Richards, M., Macaulay, V., Bandelt, H., Sykes, B. (1998). Phylogeography of mitochondrial DNA in Western Europe. *Ann Hum Genet* 62, 241-260.
227. 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., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Renzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A., Bandelt, H.-J. (2000). Tracing European founder lineages in the near Eastern mtDNA pool. *Am J Hum Genet*, 1251-1276.
228. Richards, M., Macaulay, V. (2001). The mitochondrial gene tree comes of age. *Am J Hum Genet* 68(6), 1315-1320.
229. Richards, M. (2003). The Neolithic invasion. *Ann Rev Anthropol* 32, 135-162.
230. Richards, M., Bandelt, H.-J., Kivisild, T., Oppenheimer, S. (2006). A Model for the Dispersal of Modern Humans out of Africa. In *Mitochondrial DNA and the evolution of Homo sapiens*, H.-J. Bandelt, V. Macaulay, and M. Richards, eds. (Berlin: Springer-Verlag).
231. Rogaeve, E. I., Grigorenko, A. P., et al. (2009). Genotype analysis identifies the cause of the royal disease. *Science* 326(5954), 817.
232. Rollo, F., Ermini, L., et al. (2006). Fine characterization of the Iceman's mtDNA haplogroup. *Am J Phys Anthropol* 130(4), 557-564.
233. Rosset, S., Wells, R. S., et al. (2008). Maximum-likelihood estimation of site-specific mutation rates in human mitochondrial DNA from partial phylogenetic classification. *Genetics* 180(3), 1511-1524.
234. Rollo F., Ermini L., Luciani S., Marota I., Olivieri C., Luiselli, D. (2006). Fine characterization of the Iceman's mtDNA haplogroup. *Am J Phys Anthropol* 130, 557-564.
235. Rosenberg, N. A., Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat Rev Genet* 3(5), 380-390.
236. Sampietro, M. L., Gilbert, M.T., Lao, O., Caramelli, D., Lari, M., Bertranpetit, J., Lalueza-Fox, C. (2006). Tracking down human contamination in ancient human teeth. *Mol Biol Evol* 23(9), 1801-1807.
237. Sampietro, M., Lao, O., Caramelli, D., Lari, M., Pou, R., Mart  , M., Bertranpetit, J., Lalueza-Fox, C. (2007). Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc Biol Sci* 274, 2161-2167.
238. Sanchez, J.J., Borsting, C., Hallenberg, C., Buchard, A., Hernandez, A., Morling, N. (2003). Multiplex PCR and minisequencing of SNPs—a model with 35 Y chromosome SNPs. *Forensic Sci Int* 137, 74–84.
239. Sanchez, J.J., Borsting, C., Morling, N. (2005). Typing of Y chromosome SNPs with multiplex PCR methods. *Methods Mol Biol* 297, 209–228.
240. Sanchez, JJ, Phillips, C, Borsting, C et al. (2006). A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27, 1713–1724.
241. Savolainen, P., Zhang, Y.P., Luo, J., Lundeberg, J., Leitner, T. (2002). Genetic evidence for an East Asian origin of domestic dogs. *Science* 298, 1610-1613.

242. Schmitz, R. W., Serre, D., et al. (2002). The Neandertal type site revisited: interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. *Proc Natl Acad Sci U S A* 99(20), 13342-13347.
243. Schoetensack, O. (1908). *Der Unterkiefer des Homo heidelbergensis aus den Sanden von Mauer bei Heidelberg*. Leipzig: Wilhelm Engelmann.
244. Schwartz, M., Vissing, J. (2002). Paternal inheritance of mitochondrial DNA. *N Engl J Med* 347(8), 576-580.
245. Sharp, P. M., Shields, D. C., et al. (1989). Chromosomal location and evolutionary rate variation in enterobacterial genes. *Science* 246(4931), 808-810.
246. Shapiro, B., Drummond, A. J., et al. (2004). Rise and fall of the Beringian steppe bison. *Science* 306(5701), 1561-1565.
247. Shepherd, L. D. Lambert, D.M. (2008). Ancient DNA and conservation: lessons from the endangered kiwi of New Zealand. *Mol Ecol* 17(9), 2174-2184.
248. Sigurgardóttir, S., Helgason, A., et al. (2000). The mutation rate in the human mtDNA control region. *Am J Hum Genet* 66(5), 1599-1609.
249. Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139, 457-462.
250. Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., Richards, M.B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84, 740-759.
251. Soares P., Achilli A., Semino O., Davies W., Macaulay V., Bandelt H.J., Torroni A., Richards M.B. (2010). The archaeogenetics of Europe. *Curr Biol* 20, R174–R183.
252. Sokal, R. R., Oden, N. L., et al. (1991). Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* 351(6322), 143-145.
253. Stiller, M., Green, R.E., Ronan, M., Simons, J.F., Du, L., He, W., Egholm, M., Rothberg, J.M., Keates, S.G., Ovodov, N.D., Antipina, E.E., Baryshnikov, G.F., Kuzmin, Y.V., Vasilevski, A.A., Wuenschell, G.E., Termini, J., Hofreiter, M., Jaenicke-Despres, V., Pääbo, S. (2006). Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proc Natl Acad Sci U S A* 103, 13578-13584.
254. Stoneking, M., Hedgecock, D., Higuchi, R.G., Vigilant, L., Erlich, H.A. (1991). Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am J Hum Genet* 48(2), 370–82.
255. Stringer, C.B., Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. *Science* 239, 1263-1268.
256. Tajima, F. (1983). Evolutionary Relationship of DNA Sequences in finite populations. *Genetics* 105, 437–460.
257. Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogväli, E.L., Tolk, H.V., Reidla, M., Metspalu, E., Pliss, L., Balanovsky, O., Pshenichnov, A., Balanovska, E., Gubina, M., Zhadanov, S., Osipova, L., Damba, L., Voevoda, M., Kutuev, I., Bermisheva, M., Khusnutdinova, E., Gusar, V., Grechanina, E., Parik, J., Pennarun, E., Richard, C., Chaventre, A., Moisan, J.P., Barác, L., Perici, M., Rudan, P., Terzi, R., Mikerezi, I., Krumina, A., Baumanis, V., Koziel, S., Rickards, O., De Stefano, G.F., Anagnou, N., Pappa, K.I., Michalodimitrakakis, E., Ferák, V., Füredi, S., Komel, R., Beckman, L., Villems, R. (2004). The Western and Eastern roots of the Saami--the story of genetic

- "outliers" told by mitochondrial DNA and Y chromosomes. *Am J Hum Genet* 74, 661-682.
258. Tamura, K., Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10(3), 512-26.
 259. Tattersall, I., Schwartz, J.H. (1999). Hominids and hybrids: the place of Neanderthals in human evolution. *Proc Natl Acad Sci U S A* 96(13):7117-9.
 260. Tavaré, S. (1986). Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *Lectures on Mathematics in the Life Sciences* (American Mathematical Society) 17, 57–86.
 261. Templeton, A. R., Routman, E., et al. (1995). Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* 140(2), 767-782.
 262. Templeton, A.R. (2009). Statistical hypothesis testing in intraspecific phylogeography: NCPA versus ABC. *Mol. Ecol.* 18, 319–331.
 263. Templeton, A.R. (2010). Coalescent-based, maximum likelihood inference in phylogeography. *Mol Ecol* 19, 431–435.
 264. Thomsen, P. F., Elias, S., et al. (2009). Non-destructive sampling of ancient insect DNA. *PLoS One* 4(4), e5048.
 265. Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M, Smith, D.G., Vullo, C.M., Wallace, D.C. (1993). Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet.*53(3), 563-90.
 266. Torroni, A., Bandelt, H. J., et al. (1998). mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 62(5), 1137-1152.
 267. Torroni, A., Bandelt, H.-J., Macaulay, V., Richards, M., Cruciani, F., Rengo, C., Martinez-Cabrera, V., Villems, R., Kivisild, T., Metspalu, E., Parik, J., Tolk, H.V., Tambets, K., Forster, P., Karger, B., Francalacci, P., Rudan, P., Janicijevic, B., Rickards, O., Savontaus, M.L., Huoponen, K., Laitinen, V., Koivumäki, S., Sykes, B., Hickey, E., Novelletto, A., Moral, P., Sellitto, D., Coppa, A., Al-Zaheri, N., Santachiara-Benerecetti, A.S., Semino, O., Scozzari, R. (2001). A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 69(4): 844-852.
 268. Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykes, B.C., Bradley, D.G. (2001). Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410, 1088-1091.
 269. Valdiosera, C., Garcia, N., Dalén, L., Smith, C., Kahlke, R.D., Lidén, K., Angerbjörn, A., Arsuaga, J.L., Götherström, A. (2006). Typing single polymorphic nucleotides in mitochondrial DNA as a way to access Middle Pleistocene DNA. *Biol Lett* 2, 601-603.
 270. van Oven, M., Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30, E386-394. www.phylotree.org/
 271. Vernesi, C., Caramelli, D., Dupanloup, I., Bertorelle, G., Lari, M., Cappellini, E., Moggi-Cecchi, J., Chiarelli, B., Castri, L., Casoli, A., Mallegni, F., Lalueza-Fox, C., Barbujani, G. (2004). The Etruscans, a population-genetic study. *Am J Hum Genet* 74, 694-704.

272. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., Wilson, A.C. (1991). African populations and the evolution of human mitochondrial DNA. *Science* 253(5027), 1503-7.
273. Vilà, C., Leonard, J. A., et al. (2001). Widespread origins of domestic horse lineages. *Science* 291(5503), 474-477.
274. von Cramon-Taubadel, N., Pinhasi, R. (2011). Craniometric data support a mosaic model of demic and cultural Neolithic diffusion to outlying regions of Europe. *Proc Biol Sci*.
275. Vreeland, R.H., Rozenwieg, W.D, Powers, D.W. (2000). Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature* 407, 897-900.
276. Vuissoz, A., Worobey, M., Odegaard, N., Bunce, M., Machado, C.A., Lynnerup, N., Peacock, E.E., Gilbert, M.T.P. (2004). The survival of PCR-amplifiable DNA in cow leather. *J Archaeol Sci* 34:823-829.
277. Wakeley, J. (1993). Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *J Mol Evol* 37(6), 613-623.
278. Wallace, D.C., Brown, M.D., Lott, M.T. (1999). Mitochondrial DNA variation in human evolution and disease. *Gene* 238, 211-230.
279. Willerslev, E., Hansen, A. J., et al. (1999). Diversity of Holocene life forms in fossil glacier ice. *Proc Natl Acad Sci U S A* 96(14), 8017-8021.
280. Willerslev, E., Hansen, A.J., Binladen, J., Brand, T.B., Gilbert, M.T., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D.A., Cooper, A. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300, 791-795.
281. Willerslev, E., Hansen, A.J., Poinar, H.N. (2004). Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol Evol* 19, 141-147.
282. Willerslev, E., Cooper, A. (2005). Ancient DNA. *Proc Biol Sci* 272, 3-16.
283. Willerslev, E., Cappellini, E., Boomsma, W., Nielsen, R., Hebsgaard, M.B., Brand, T.B., Hofreiter, M., Bunce, M., Poinar, H.N., Dahl-Jensen, D., Johnsen, S., Steffensen, J.P., Bennike, O., Schwenninger, J.L., Nathan, R., Armitage, S., de Hoog, C.J., Alfimov, V., Christl, M., Beer, J., Muscheler, R., Barker, J., Sharp, M., Penkman, K.E., Haile, J., Taberlet, P., Gilbert, M.T., Casoli, A., Campani, E., Collins, M.J. (2007). Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* 317(5834):111-4.
284. Winters, M., Barta, J.L., Monroe, C., Kemp, B.M. (2011). To Clone or Not To Clone: Method Analysis for Retrieving Consensus Sequences In Ancient DNA Samples. *PLoS ONE* 6(6): e21247.
285. Wolpoff, M.H., Spuhler, J.N., Smith, F.H., Radovic, J., Pope G., Frayer, D.W., Eckhardt, R., Clark, G. (1988). Modern human origins. *Science* 241, 772-774.
286. Woodward, S.R., Weyand, N.J., Bunnell, M. (1994). DNA sequence from Cretaceous period bone fragments. *Science* 266, 1229– 32.
287. Wood, J.R., Wilmshurst, J.M., Worthy, T.H., Cooper, A. (2011). *Sporormiella* as a proxy for non-mammalian herbivores in island ecosystems. *Quat Sci Rev* 30, 915-920.
288. Wright, S. (1931). Evolution in Mendelian Populations. *Genetics* 16(2), 97-159.
289. Yang, Z. (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol Biol Evol* 10(6), 1396-1401.

290. Yang, Z., Kumar, S. (1996). Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites. *Mol Biol Evol* 13(5), 650-659.
291. Zeder, M.A., Emshwiller, E., Smith, B.D., Bradley, D.G. (2006). Documenting domestication: the intersection of genetics and archaeology. *Trends Genet* 22(3):139-55.
292. Zeder, M.A. (2008). Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc Natl Acad Sci U S A* 105(33):11597-604.
293. Zischler, H., Hoss, M., Handt, O., von Haeseler, A., van der Kuyl, A.C., Goudsmit, J. (1995). Detecting dinosaur DNA. *Science* 268, 1192–93.
294. Zuckerkandl, E. and L. Pauling (1965). Molecules as documents of evolutionary history. *J Theor Biol* 8(2), 357-366.
295. Zvelebil M., Dolukhanov P. (1991). The transition to farming in Eastern and Northern Europe. *J W Prehist* 5 (3), 233-278.

In this first chapter, I investigate the evolution of the mitochondrial gene pool of north east Europeans in a transect through time from the Mesolithic (~7,000 – 7,500 years Before Present) to the Bronze Age (~3,500 years Before Present), historical times (200 years Before Present) and the present. I also search for potential ancestors of the Saami, extreme European genetic outliers, in these ancient populations at the eastern border of Europe.

Chapter One:

Ancient Mitochondrial DNA Unravels Complex Population History in North East Europe

Abbreviations: ABC, approximate Bayesian computation; ACAD, Australian Centre for Ancient DNA; A.D., Anno Domini; aDNA, ancient DNA; AIC, Akaike information criterion; BayeSSC, Bayesian Serial SimCoal; CI, confidence interval; dNTP, deoxynucleoside triphosphate; Exo, exonuclease; HVR-I, hypervariable region I; m, arithmetic mean; min, minute; mtDNA, mitochondrial DNA; NE-E, north east Europe; np, nucleotide position; PCA, principal component analysis; qPCR, quantitative real-time Polymerase Chain Reaction; rCRS, revised Cambridge reference sequence; RPM, revolutions per minute; RSA, Rabbit Serum Albumin; SAP, shrimp alkaline phosphatase; SBE, single base extension; s.d., standard deviation; SNP, single nucleotide polymorphism; UV, ultra-violet; yBP, years Before Present.

ABSTRACT

Prehistoric human population processes are pivotal for understanding our genetic past and the present complexity of genetic marker distribution. Modern-day North East Europeans of the Baltic and Volga-Ural regions, Fennoscandia, and western Siberia are characterised by varying frequencies of genetic markers of both European and east Eurasian origins. The extent and timing of interactions between genetically differentiated populations of eastern and western Eurasia are critical aspects in understanding the dynamics that shaped the gene pool of North East Europeans. In order to reconstruct human population dynamics in north east Europe, we examined mitochondrial DNA (mtDNA) from four diachronically sampled archaeological sites: two Mesolithic sites at Uznyi Oleni Ostrov and Popovo ($n = 9$ and $n = 2$; 7,500 years before present, yBP), a Bronze Age site at Bolshoy Oleni Ostrov ($n = 23$; 3,550 yBP) and a historical Saami graveyard Chalmny-Varre ($n = 42$, 200 – 300 yBP). Important temporal information emerges from the new genetic data obtained from 76 human remains and suggests a complex genetic history in north east Europe. Prehistoric foragers - found at Uznyi Oleni Ostrov and Popovo - showed high frequencies and diversity in haplogroups U4 and U5a, a pattern previously observed in populations of ancient European hunter-gatherers. The Bronze Age population of Bolshoy Oleni Ostrov was clearly differentiated from Mesolithic hunter-gatherers due to the presence of mtDNA haplogroups C, D and Z of eastern Siberian origin. Historical Chalmny-Varre was found to be closely linked to modern-day Saami due to the presence of haplogroup V and the 'Saami motif' U5b1b1. This study identified genetic discontinuities between prehistoric and modern-day populations of north east Europe, indicating multiple migration waves into the area. Mitochondrial lineages detected in Mesolithic and Bronze Age populations were found to be rare or absent from modern-day Eurasian populations and from contemporaneous prehistoric groups previously sampled for ancient DNA. This result indicates a certain level of isolation of prehistoric north east European populations, as well as post-Bronze Age mtDNA lineage extinction and/or population replacement. This work provides unique genetic evidence for the understanding of human population movements at the scale of Eurasia.

INTRODUCTION

Prehistoric, historical and cultural background of north east Europe

North east Europe (NE-E) comprises the Baltic, and Volga-Ural regions, sub-Arctic Europe and western Siberia. The diversity of cultures, languages and lifestyles in this area today is indicative of a complex population history. Human population history in NE-E is thought to have involved multiple migrations from diverse origins driven by changes in climatic and environmental conditions. However, the precise origins, timing and genetic impact of the population events that have shaped the gene pool of North East Europeans remains largely unclear today.

During the Upper Palaeolithic (~40,000 – 30,000 years Before Present, yBP), anatomically modern-day humans first settled in NE-E. At this time, colonization of the northernmost latitudes of Europe was hindered by an ice sheet that covered NE-E until the end of the Ice Age around 11,500 yBP (Svendsen et al., 2004). During the Late Glacial (~15,000 yBP), it is thought that glacial recession and megafaunal extinction gradually drove small foraging groups out of their southern periglacial refuges towards newly freed territories in the North in order to find other forms of resources (Gamble et al., 2004; Kozłowski & Bandi et al., 1984; Dolukhanov et al., 1997; Hofreiter et al., 2009; Campos et al., 2010). Migrations from the west, the south and the east into NE-E are thought to have taken continued into the early Holocene (~10,000 – 8,000 yBP). These population movements were proposed to have led to the widespread establishment of complex Mesolithic societies of fishermen and hunter-gatherers that thrived in the steppe-forest zone of northern Europe (Price, 1991; Jacobs et al., 1995; Dolukhanov, 1997).

During the period of climatic conditions in the Holocene Climatic Optimum (8,000 – 4,200 yBP), foraging activities intensified in NE-E. At the same time, western Europe was undergoing the Neolithic transition, during which an agricultural lifestyle rapidly developed partly due to more favourable climatic and ecological conditions (Zvelebil & Dolukhanov, 1991). Agricultural communities were only observed in the southern fringe of NE-E. Archaeology established that these farming groups were culturally similar to central and eastern European Neolithic farmers. Around 3,000 yBP, metallurgy gradually appeared and spread into NE-E, supposedly from south Siberia and the Urals.

Historical records describe numerous subsequent population movements that originated in both the West (e.g., Scandinavia, western/central Europe) and the East (e.g., Volga-Ural region, Mongolia). Examples of such migrations are the Slavic expansion from the central European homeland (7th to 14th century Anno Domini; A.D.; 600 – 1,400 yBP) and the Mongol invasions (6th to 7th century A.D.; 1,300-1,500 yBP; Grousset et al., 1970), but expansions to the north were relatively limited by the cold and arid environment of the European sub-arctic region.

Influences from the west, the south and the east are thought to be at the origin of the diversity in cultures seen among extant North East Europeans. Two different linguistic families - Finno-Ugric and Indo European (Slavic, Baltic and Germanic), are spoken in NE-E. The cultural landscape of the area is constituted of the Slavic, Baltic, Scandinavian, Karelian, Volga-Ural and Saami traditions. Notably, local Saami are the only semi-nomadic indigenous people characterised by ‘traditional’ livelihoods (fishing, reindeer herding). They inhabit the northern area of Norway, Sweden, Finland and Russia (Fennoscandia) and their language is part of the Finno-Ugric branch of the Uralic language family (Sammallahti, 1998). Analysis of the physical features and DNA of the Saami defined them as marked outliers within the European population (Guglielmino et al., 1990; Cavalli-Sforza et al., 1994; Beckman et al., 1998; Tambets et al., 2004; Ingman & Gyllensten, 2007). In particular, discussions have focused on the relative proportions and contribution of gene flow from west and east Eurasia that contributed to the Saami population, and the distinctiveness of their population history. The question of the origin and population history of Saami is still mysterious and greatly debated today.

Reconstructing the population history in north east Europe using modern-day mitochondrial data

A variety of scenarios for the population genetic history of Europe have been proposed from the study of genetic markers in present-day populations. In particular, the extensive characterization of the maternally inherited mitochondrial DNA (mtDNA) in Europeans has helped identify the particular lineages thought to have participated in the first settlement of Europe and subsequent population movements (Richards et al., 1996; Richards et al., 1998; Richards et al., 2000). However, inhabitants of the Baltic area were found to be rather homogeneous at the mtDNA level, regardless of their spoken language. Slight differences could however be

observed due to unequal influences from subtly genetically differentiated neighbouring populations: central Europeans in the West, people of the Volga-Ural in the East. In the North, Baltic populations were affected by the particular genetic makeup of Saami, whose signature entails an enrichment of mtDNA lineages U5b1b1, V, Z1 and D5 (Pliss et al., 2006; Lappalainen et al., 2008; Tambets et al., 2004). However, population genetic processes, such as e.g., replacement and genetic drift, can significantly bias the reconstruction and timing of past migratory and demographic events inferred from the analysis of modern-day mtDNA distributions. This can lead to erroneous interpretations of ancient human population history, a problem that potentially can be circumvented by the direct assessment of the genetic diversity in ancient human remains.

Reconstructing the population history in north east Europe using ancient mitochondrial data

Ancient DNA (aDNA) studies have recently described mtDNA lineages in Palaeolithic/Mesolithic hunter-gatherers of central, eastern and Scandinavian Europe (Bramanti et al., 2009; Malmström et al., 2009; Krause et al., 2010), as well as in Neolithic farmers of southern and central Europe (Haak et al., 2005; Sampietro et al., 2007; Melchior et al., 2010; Haak et al., 2010). Ancient DNA uncovered a substantial heterogeneity in the distribution of mtDNA in western, central and northern Europe, both spatially and temporally, that could not be inferred from the analysis of modern genetic variation. Together, these aDNA studies revealed a significant mtDNA discontinuity among ancient populations and cultures, as well as between ancient and modern-day populations, emphasizing the complexity of the processes involved in the shaping of the European gene pool (Renfrew, 2010). As a result, it is probable that the analysis of genetic markers in modern-day populations has not reconstructed the full picture of human genetic history in NE-E.

In order to shed light on human origins and migrations in NE-E, we studied the mitochondrial diversity of ancient north east European populations that were temporally sampled from prehistoric to historical times. The samples were collected from archaeological sites in western Russia and the Kola Peninsula (Figure 1). They ranged in age from the Mesolithic (Uznyi Oleni Ostrov and Popovo; both ~7,000 - 7,500 yBP), to the Bronze Age (Bolshoy Oleni Ostrov; 3,500 yBP) and the 18th century A.D. (300 - 200 yBP; graveyard of Chalmny-Varre). Uznyi Oleni Ostrov, the

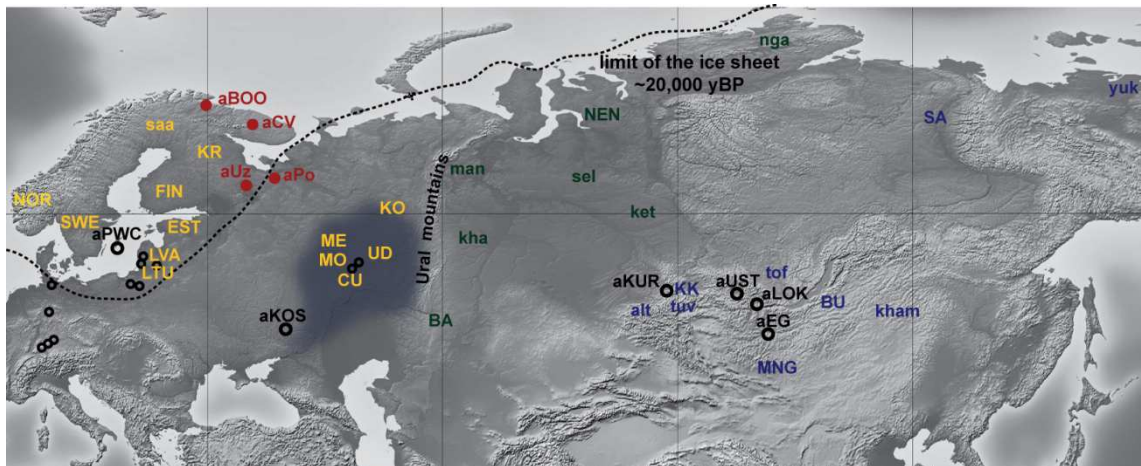


Figure 1. Map of Eurasia showing the approximate location of selected Eurasian populations. The dark grey area signifies the approximate location of the Volga-Ural basin. Red dots represent the archaeological sites sampled for ancient mitochondrial DNA in this study: aUZ, Uznyi Oleni Ostrov; aPo, Popovo; aBOO, Bolshoy Oleni Ostrov, aCV, Chalmny-Varre. White dots represent Palaeolithic/Mesolithic sites sampled for ancient mitochondrial DNA in Bramanti et al., 2009, unless specified otherwise (aPWC, aKOS). Yellow was used for modern populations of North East Europe, green for modern populations of West, South and Central Siberia, and blue for modern populations of East Siberia. See Material and Methods for population abbreviations.

‘Southern Reindeer Island’ in Russian, and Popovo are located in the Onega Lake region and the Archangelsk district respectively, along one of the proposed routes for the introduction of Saami-specific mtDNA lineages (Ingman & Gyllensten, 2007). The site of Bolshoy Oleni Ostrov, the ‘Great Reindeer Island’ in Russian, is located in the north Kola Peninsula, an area currently inhabited by Saami. Results from odontometric analyses suggested a direct genetic continuity between the Mesolithic population of Uznyi Oleni Ostrov and present-day Saami (Jacobs, 1992). Finally, the Chalmny-Varre site, located in the Kola Peninsula, is historically recognized as a Saami graveyard and provides an historical control.

The mitochondrial hypervariable region I (HVR-I) was sequenced and mtDNA haplogroups were assigned/confirmed by typing 22 haplogroup-diagnostic single nucleotide polymorphisms (SNPs) in the mtDNA coding region by Single Base Extension (SBE; Haak et al., 2010). Strict criteria for validating aDNA data were followed (Pääbo et al., 2004). The three ancient populations were tested as potential genetic ancestors to the present-day Saami and to other extant North East Europeans. Principal Component Analysis (PCA), genetic distance mapping, haplotype sharing

and coalescent simulations (using the program Bayesian Serial SimCoal, BayeSSC and Approximate Bayesian Computation, ABC) were performed to compare the ancient populations of NE-E with other ancient and modern-day populations.

MATERIALS AND METHODS

Sample description and archaeological context

A total of 215 human teeth - representing an initial 116 individuals - from the collection of the Kunstkamera Museum of St Petersburg, Russia (Valery Khartanovich, Alexandra Buzhilova and Sergey Koshel). The samples were collected from 4 archaeological sites in northwestern Russia: Uznyi Oleni Ostrov, Popovo, Bolshoy Oleni Ostrov and Chalmny-Varre.

Uznyi Oleni Ostrov: 96 teeth representing 48 individuals were obtained from the Uznyi Oleni Ostrov archaeological site, which is located on Uznyi Oleni Island, Onega Lake, Karelia (61°30'N 35°45'E). The site was first discovered in the 1920s through quarrying activities, which led to the destruction of the largest part of the graveyard. Scientific excavation of the site by Soviet archaeologists in the 1930s and the 1950s eventually unearthed a total of 177 burials in 141 different mortuary features (Gurina et al., 1956). The population size of the burial ground before its partial destruction was estimated around 500 individuals (O'Shea and Zvelebil, 1984), thus making it the largest Mesolithic graveyard yet discovered in Boreal Europe. First identified as a Neolithic graveyard, a reanalysis and radiocarbon dates centered around an age of 7,000 - 7,500 years Before Present (Price & Jacob, 1990; Wood, 2006), ultimately lead to its classification within the Baltic Mesolithic culture (Veretye culture). The abundance and diversity of mortuary artifacts uncovered make the Uznyi Oleni Ostrov graveyard remarkable among other Mesolithic sites.

Popovo: 6 teeth belonging to 3 individuals were obtained from the Popovo archaeological site located on the bank of the Kinema River, in the Archangelsk region (64°32'N 40°32'E). Associated artefacts were identified as representative of the Boreal Mesolithic Veretye culture. Long-term and successive usage of this graveyard is suggested by the wide range of dates obtained for this site (9,000 – 9,500 yBP and 7,500 – 8,000 yBP; Oshibikina, 1999). The small sample size, the temporal and

geographic proximity, and mainly, their cultural similarity, led to the samples of Uznyi Oleni Ostrov and Popovo sites being grouped together for statistical analysis.

Bolshoy Oleni Ostrov: 45 teeth representing 23 individuals were obtained from the Bolshoy Oleni Ostrov archaeological site, located in Murmansk region, Kola Peninsula (68°58'N 33°05'E). Radiocarbon dates for two graves of the Bolshoy Oleni Ostrov site were obtained from the Oxford Radiocarbon Accelerator Unit (ORAU), United Kingdom, at 3,525-3,440 yBP and 3,500-3,430 yBP for grave 12 and grave 13, respectively.

Chalmny-Varre: 68 teeth belonging to 42 individuals were obtained from the Chalmny-Varre archaeological site, located in the Murmansk region, middle part Ponoy River, Kola Peninsula. The graveyard was dated to the 18-19th centuries A.D. and was associated with the Saami culture.

Sample preparation and DNA extraction

At the ACAD, the outer surface of the each tooth was first decontaminated through exposure to ultraviolet (UV) light for 20 min on each side. Then, dirt was removed from the outer surface by gently wiping the teeth with a paper towel soaked in sodium hypochlorite (bleach).

Tooth powder was then prepared following two different protocols. The archaeological and anthropological value of the samples from Uznyi Oleni Ostrov, Popovo and Bolshoy Oleni Ostrov meant that the preservation of their morphological integrity was a requirement. A first protocol, involving minimal destruction of the tooth, consisted of cutting the root off the crown and powdering the inside of the root using a dental drill. A second protocol was used for the Chalmny-Varre samples. The outer surface of the teeth was first removed using a Dremel[®] drill, then the tooth root was cut off from the crown and finally, the root was ground into a fine powder in a Mikro-dismembrator ball mill (Sartorius). Tooth powder was stored at 5°C until further use.

Digestion was carried out by incubating the powdered teeth in 3.33 mL of buffer (0.5M EDTA, pH 8.0; 0.5% N-lauryl sarcosine; 20 mg/μL proteinase K) overnight on a rotary mixer at 37°C. DNA was isolated using a phenol:chloroform of one volume of phenol:chloroform:isoamyl-alcohol (25:24:1, pH 8.0) followed by a centrifugation at 4,600 revolutions per minute (RPM) for 10 min. The aqueous phase was recovered and this step was repeated. The aqueous phase was then gently mixed

to one volume of 99% chloroform and centrifuged at 4,600 RPM for 10 min before being recovered and desalted/concentrated using an Amicon Ultra-4 centrifugal dialysis (50 kDa cut-off; Millipore) for 10 min at 4,600 RPM. The extract was washed twice with distilled water (3.5 uL for 10 min, 2.5 uL for 5 min), taken up in 30 uL and stored at -18°C.

Hypervariable-Region I sequencing

The HVR-I was amplified from positions 15997 to 16409 (according to the revised Cambridge reference sequence, rCRS; Andrews et al., 1999) by PCR in four overlapping fragments. PCRs were set up in 25 µL containing 2 µL of DNA extract and the following reaction mix: 1x PCR Gold Buffer (Applied Biosystems), 2.5 mM MgCl₂ (Applied Biosystems), 0.5 mM deoxynucleoside triphosphate (dNTP) Mix (Invitrogen), 2 U AmpliTaq Gold DNA Polymerase (Applied Biosystems), 0.2 µM each primer (See Table S1), 1 mg/mL of Rabbit Serum Albumin (RSA, Sigma). PCRs were performed in a DNA engine Tetrad 2 thermocycler (BioRAD) under the following conditions: initial enzyme activation at 95°C for 6 min; 40 cycles of denaturation at 95°C for 30 s, annealing at 56°C for 30 s, elongation at 72°C for 30 s; followed by final elongation at 65°C for 10 min. The presence of PCR products of the expected size was visually checked by running 5 µL of PCR reaction in a 3.5% agarose electrophoresis gel at 100 V for 30 min, post-staining in ethidium bromide and UV-light exposure. Leftover primers and dNTPs were removed from PCR reactions showing successful amplification products of expected lengths by incubating 5 µL of PCR product with 1 U of Shrimp Alkaline Phosphatase (SAP) and 0.8 U of Exonuclease I (ExoI) at 37°C for 40 min, followed by heat inactivation at 80°C for 10 min. Purified PCR products were sequenced on both strands using the BigDye® Terminator technology (Applied Biosystems). 3 µL of PCR product were added to 7 µL of the following mix: 1x Sequencing buffer, 1x Terminator Ready reaction mix and 2.5 µM primer. The sequencing reaction was performed under the following conditions: 95°C for 1 min; 25 cycles of: denaturation at 95°C for 10 s, annealing at 55°C for 10 s, elongation at 60°C for 2 min 30s. Sequencing products were purified using a Multiscreen_{HTS} Vacuum Manifold (Millipore), according to the manufacturer's protocol. Sequences were read by an ABI PRISM 3130xl® Genetic Analyzer (Applied Biosystems) and aligned using the Sequencer v4.7 software.

Typing of mtDNA coding region single nucleotide polymorphisms (SNPs) by GenoCoRe22 multiplex PCR

Twenty-two haplogroup-diagnostic mtDNA coding region single nucleotide polymorphisms (SNPs) were typed using the Single-Base Extension (SBE) system described in Haak et al., 2010 (Table S1). Multiplex PCRs were set up in 25 μ L containing 2 μ L of DNA extract and the following reaction mix: 1x PCR Gold Buffer (Applied Biosystems), 6 mM $MgCl_2$ (Applied Biosystems), 0.5 mM dNTP Mix (Invitrogen), 2 U AmpliTaq Gold DNA Polymerase (Applied Biosystems), 0.2 μ M each primer, 1 mg/mL of RSA (Sigma). Multiplex PCRs were carried out in a DNA engine Tetrad 2 thermocycler (BioRAD) under the following conditions: initial enzyme activation at 95°C for 6 min; 40 cycles of: denaturation at 95°C for 30 s, annealing at 60°C for 30 s, elongation at 65°C for 30 s; followed by a final elongation step at 65°C for 10 min. Multiplex PCR reactions showing amplification products were cleaned up with ExoSAP following the protocol described above. The ABI Prism SNaPshot multiplex reaction kit (Applied Biosystems) was used to perform the SNP typing by Single Base Extension (SBE) reactions. The manufacturer's instructions were modified in order to minimize the occurrence of artefacts by adding 10% 3M ammonium sulphate to the extension primer mix. The SBE reaction was performed under the following thermocycling conditions: 35 cycles of denaturation at 96°C for 10 s, annealing at 55°C for 5 s, and extension at 60°C for 30 s. SBE reaction products were purified through incubation with 1 U SAP at 37°C for 40 min followed by heat inactivation at 80°C for 10 min. Samples were prepared for by capillary electrophoresis by adding 2 μ L of purified SNaPshot product to 11.5 μ L Hi-Di™ Formamide (Applied Biosystems) and 0.5 μ L of Gene-Scan-120 LIZ™ (Applied Biosystems) size standard. Samples were run on an ABI PRISM 3130xl® Genetic Analyzer (Applied Biosystems) after a denaturation step according to the manufacturer's instructions using a POP-6® polymer (Applied Biosystems). Evaluation and analyses of SNaPshot profiles were performed using custom settings within the Genemapper v3.2 Software (Applied Biosystems; Table S2).

Cloning

Cloning of PCR amplified ancient DNA was performed by Guido Brandt at the Institute of Anthropology, Johannes Gutenberg-University, Mainz, Germany. Cloning of PCR products into pUC18 vectors and downstream steps leading to the sequencing

of the clones were carried out according to the protocol described in Haak et al., 2005.

Quantitative Real-Time PCR

The size distribution of endogenous aDNA molecules has previously been shown to be skewed towards smaller fragment sizes due to post-mortem damage, i.e. DNA fragmentation (Pääbo et al., 2004; Noonan et al., 2005; Malmström et al., 2007; Adler et al., 2010). The absence of significant amount of larger DNA fragments consistent with modern DNA contaminants in the PCR products was verified by DNA quantification in two populations of molecules of different sizes (133 bp and 179 bp). The copy-number of two HVR-I fragments: L16209/H16303 (133 bp) and L16209/H16348 (179 bp) was estimated in selected aDNA extracts by quantitative real-time PCR (qPCR). A standard curve giving the minimal number of cycles for detection of fluorescence above baseline (threshold cycle, Ct) was constructed for known concentrations of target DNA. The standard curve was then used to estimate absolute copy-number in the aDNA extracts. One standard curve was created per primer pair using DNA extracted from a fresh buccal cheek swab with the DNeasy Blood and Tissue Kit (Qiagen) as follows. The two HVR-I fragments L16209/H16303 and L16209/H16348 were amplified in 25 µL containing 1xHotmaster Buffer (Eppendorf), 0.5 U Hotmaster Taq (5Prime), 10 µM forward and reverse primers, distilled water, and 2 µL of DNA extract. Thermocycling conditions were: initial enzyme activation at 95°C for 6 min; 35 cycles of: denaturation at 94°C for 30 s, annealing at 60°C for 30 s, elongation at 72°C for 30s; followed by a final elongation step at 65°C for 10 min. Agencourt Ampure (Beckman Coulter) was used to purify the PCR products according to the manufacturer's instructions. DNA concentration of the two amplification products was averaged from multiple measurements with a NanoDrop spectrometer (ThermoScientific). Ten-fold dilutions ranging from 1×10^6 to 10 copies/µL were prepared for both L16209/H16303 and L16209/H16348 PCR products and amplified by qPCR. Reactions were set up in 10 µL containing 1x Brilliant II SYBR Green QPCR Master mix (Agilent), 10 mg/ml RSA, 10 µM forward and reverse primers and 1 µL of DNA extract, and the thermocycling conditions above. qPCRs for each dilution of each fragment were performed in triplicate and repeated on a different day to provide an average. qPCRs were performed on the Rotor-Gene 6000 thermocycler and data analysed with the Rotor-Gene 6000 Series Software 1.7 (Corbett). The construction of the standard curve allowed the linearity

and the efficiency of the two assays to be assessed. An $R^2 > 0.95$ for both reactions indicated that the C_t evolved linearly with the \log_{10} of DNA concentration and efficiencies of 0.91 indicated that each cycle of the exponential phase doubled the amount of DNA. The specificity of both primer pairs was also verified, first, by observation of a single band of PCR products on a 2% agarose electrophoresis gel, and second, by examination of a single dissociation peak on the melt-curve of qPCR products. The same conditions as the ones used for the standard curve were applied to qPCRs using aDNA extracts, as well as for negative and positive controls (Table S3). To compare the copy-number of the 133 bp (L16209/H16303) and the 179 bp (L16209/H16348) fragments, the Shapiro-Wilk W test was first used to verify that the number of copies for each fragment followed a normal distribution ($p = 0.2215$ for the L16209/H16303 short fragment and $p = 0.5381$ for the long L16209/H16348 fragment). A significantly larger number of copies for the shorter compared to the larger fragment was statistically confirmed by a one-tailed paired t-test ($p = 0.04337$) in R version 2.12 (R Development Core Team, <http://www.R-project.org>). This result is in accordance with relatively low levels of modern DNA contaminants on the aDNA extracts produced in this study.

Authentication of the mtDNA data

Strict precautions were taken in order to minimise the risk of contamination by modern DNA and to detect artefactual mutations arising from aDNA degradation. Seven criteria support the authenticity of the mtDNA data presented here.

(1) Pre-PCR DNA work was carried out at the ACAD, University of Adelaide, a purpose-built a positive air pressure laboratory dedicated to aDNA studies, which is physically isolated from any molecular biology laboratory amplifying DNA. Routine decontamination of the laboratory surfaces and instruments involves exposure to UV radiation and thorough cleaning using DNA oxidants such as bleach, Decon and Ethanol. In order to protect the laboratory environment from human DNA, researchers are required to wear protective clothes consisting of a whole body suit, a facemask, a face shield, gum boots, and three pairs of surgical gloves that are changed on a regular basis.

(2) Obvious large-scale contamination within the laboratory or in the reagents were monitored and controlled by blank controls (one extraction blank for every five ancient samples and two PCR/GenoCoRe22 blank controls for every 6 reactions). In

addition, no haplotype similar to any of those possessed by laboratory members was consistently amplified from aDNA extracts.

(3) Cloning - performed at the Institute of Anthropology, Johannes Gutenberg-University, Mainz, Germany - was used to verify mutations in six ancient mtDNA haplotypes. The sequences of the clones highlighted nucleotide positions modified by post-mortem damage, shown as inconsistent cytosine to thymine or guanine to adenosine base changes (Figure S2).

(4) Artefactual and hybrid sequences potentially arising from various exogenous DNA molecules, DNA degradation or jumping PCR were also tested through multiple replications of each individual HVR-I fragment. Sequences were obtained from at least two independent PCRs from independent extracts from two samples for each individual (i.e., a minimum of four independent PCRs). This strategy was chosen over cloning of single PCR products for most of the individuals examined, since under low-template conditions, clone sequences from one single PCR can represent a biased distribution of sequences that were selectively amplified from a single highly degraded starting DNA template. We believe that a repetitive approach, based on multiple independent repetitions is a powerful alternative to cloning. For one individual - BOO57-1 - of the Bolshoy Oleni Ostrov population, independent replications and cloning did not allow the allelic resolution at position 16390. At this position, double picks were observed on the direct sequencing chromatograms and alleles showed an equal distribution among clones. It is possible that the position might be heteroplasmic in the BOO57-1 individual.

(5) Low level of laboratory contamination was verified by the replication of matching DNA sequences (eight individuals: one from Uznyi Oleni Ostrov, five from Bolshoy Oleni Ostrov and two from Chalmny-Varre) in an independent aDNA laboratory (Institute of Anthropology, Johannes Gutenberg-University, Mainz, Germany).

(6) Quantitative PCR for selected aDNA extracts showed that copy-numbers of short DNA fragments were significantly larger than for long DNA fragments. This reflects the DNA fragmentation typically observed in aDNA extracts. Quantitative PCR results suggest a low-level of contaminating DNA molecules, whose presence would have been detected by higher copy-number of less fragmented, long, DNA molecules.

(7) The phylogenetic consistency of the haplotypes and matching haplogroup assignments of both HVR-I data and coding region SNPs, were indicative of the robustness of the mtDNA typing approach presented here.

Populations used in comparative analyses

Mitochondrial DNA data generated from the ancient populations of Uznyi Oleni Ostrov/Popovo, Bolshoy Oleni Ostrov and Chalmny-Varre was compared to data obtained from other ancient as well as modern-day populations. Data for extant populations were compiled in the MURKA mitochondrial DNA database and integrated software, which currently contains 168,000 HVR-I records from published studies and is curated by Valery Zaporozhchenko, Oleg Balanovsky and Elena Balanovska of the Russian Academy of Medical Sciences. A sub-sample of 97 ancient and modern Eurasian populations (~60,350 individuals) was used for comparative analysis. Names of modern-day populations were abbreviated using ISO codes in capital letters, and in small letters when ISO codes were not available. Unless specified otherwise, the same population codes were used for all the maps and analyses in this study, i.e. PCA, F_{ST} calculation, haplotype sharing and BayeSSC analyses. Populations were separated into six color-coded groups according to their age and geographical location (Figure S4). Populations sampled for ancient mtDNA in this study are represented in red and abbreviated as follows (e.g., in Figure 2): aUZ, Uznyi Oleni Ostrov; aPo, Popovo; aUzPo, Uznyi Oleni Ostrov/Popovo; aBOO, Bolshoy Oleni Ostrov; aCV, Chalmny-Varre. Populations previously sampled for ancient mtDNA were indicated in black and referred to as: aEG, confederated nomads of the Xiongnu (Mongolia; Keyser-Tracqui et al., 2003); aHG, Palaeolithic/Mesolithic hunter-gatherers of central/eastern Europe, (Bramanti et al., 2009); aKAZ, (Lalueza-Fox et al., 2004); aKOS, Kostenski single individual (Krause et al., 2010); aKUR, (Keyser et al., 2009); aLBK, Linearbandkeramik culture farmers (Haak et al., 2010); aLOK, Lokomotiv Kitoi Neolithic individuals (Mooder et al., 2006); aPWC, Scandinavian Pitted-Ware Culture foragers (Malmström et al., 2009); aSP, Spanish Neolithic farmers (Sampietro et al., 2007); aUST, Ust'Ida Neolithic population (Mooder et al., 2006). Grey symbolised modern-day populations of the Near East and the Caucasus regions, which were indicated by: ARM, Armenia; AZE, Azerbaijan; IRN, Iran; IRQ, Iraq; JOR, Jordania; kab, Kabardians; KAZ, Kazakhstan; kur, Kurds; nog, Nogays; PSE, Palestine; SE, Ossets; SAU, Saudi Arabia; SYR, Syria; TUR,

Turkey. Yellow corresponded to modern-day populations of NE-E and were referenced as: ALB, Albania; AUT, Austria; aro, Arorums; bas, Basques; BEL, Belarus; BGR, Bulgaria; BIH, Bosnia; CHE, Switzerland; CU, Chuvash; CYP, Cyprus; CZE, Czech Republic; DEU, Germany; ESP, Spain; EST, Estonia; FIN, Finland; FRA, France; GBR, United-Kingdom; GEO, Georgia; GRC, Greece; HRV, Croatia; HUN, Hungary; ing, Ingrians; IRL, Ireland; ISL, Iceland; IT-88, Sardinia; ITA, Italy; KO, Komis; KR, Karelians; LTU, Lithuania; LVA, Latvia; ME, Maris; MO, Mordvinians; NOR, Norway; POL, Poland; PRT, Portugal; ROU, Romania; RUS, Russia; saa, Saami; SVK, Slovakia; SVN, Slovenia; SWE, Sweden; TA, Tatars; UD, Udmurts; UKR, Ukraine; vep, Vepses. Modern-day populations of western/central Siberia were shown in green and abbreviated as follows: BA, Bashkirs; ket, Kets; khan, Khants; man, Mansi; NEN_A, eastern Nenets; NEN_E, Western Nenets; nga, Nganasans; sel, Selkups. Present-day populations of east Siberia were represented in blue and abbreviated as follows: ale, Aleuts; alt, Altaians; BU, Buryats; CHU, Chukchi; esk, Eskimos; eve, Evenks; evn, Evens; KK, Khakhassians; kham, Khamnigans; kor, Koryaks; MNG, Mongolians; niv, Nivkhs; SA, Yakuts; sho, Shors; Tel, Telenghits; tof, Tofalars; tub, Tubalars; tuv, Tuvinians; ulc, Ulchi; yuk, Yukaghirs.

Principal Component Analysis

PCA was performed using the haplogroup frequency database for ancient and modern-day populations described in Table S4. The variables were frequencies in 17 haplogroups: C, D, H, HV, I, J, K, N1, T, U2, U4, U5a, U5b, V, W, X, Z, frequencies in the six 'east Eurasian' haplogroups pooled into the 'EAS' group: A, B, F, G, Y, and finally frequencies in eight haplogroups found at lower frequencies in Eurasia and pooled into the 'misc' group: L, M*, N*, U1, U6, U7, U8. Pooling and removal of rare haplogroups (with frequencies below 1%) allowed statistical noise to be reduced. PCA was carried out using a script written for R version 2.12.

Genetic distance mapping

The genetic distance between both the ancient populations of Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov and extant populations were calculated using the software DJ. The software GeneGeo was used to plot genetic distances onto a geographic map (as described in Haak et al., 2010).

Haplotype sharing analysis

A database of mtDNA haplotypes was constructed for modern-day populations, each containing 500 individuals. This was achieved either by pooling together populations on the basis of their geographical and/or linguistic similarities or, when the population size was greater than 500, by randomly sub-sampling the given population. Previously published ancient populations were added to the database, which eventually contains mtDNA haplotypes for 17,097 individuals grouped into 38 populations (Table S4). For each haplotype of each prehistoric population presented here (Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov), the number of haplotypic matches found in each of the populations of the comparative database was estimated. This number was divided by the sample size in order to obtain the percentage of shared haplotypes. The same analysis was repeated with the exception that only the number of informative (non-basal) haplotypes shared between the compared populations was counted.

Coalescent simulations

We statistically tested two hypotheses to explain the apparent genetic differentiation observed between modern-day populations (North East Europeans, NEE and Saami, SAA) and ancient populations (Uznyi Oleni Ostrov/Popovo, aUzPo; Bolshoy Oleni Ostrov, aBOO, both described in this study; Palaeolithic/Mesolithic European hunter-gatherers, aHG, 4,250 – 30,000 yBP, Krause et al., 2010, Bramanti et al., 2009; Mesolithic Scandinavian Pitted-Ware culture individuals, 4,500 – 5,300 yBP, aPWC, Malmström et al., 2009). The null hypothesis (H0) considered that genetic drift alone was at the origin of the differences between ancient and modern-day populations. The alternative hypothesis (H1) considered that genetic discontinuity was introduced by post-Bronze Age migrations from central Europe.

Demographic models corresponding to the hypothesis H0 and H1 (Figure S3) were simulated using the program Bayesian Serial SimCoal (Anderson et al., 2005). Distributions of population statistics were calculated for a large number of simulated genealogies (100,000) and compared to the statistics directly calculated from the haplotypic data obtained for ancient and present-day populations. Comparison between simulated and observed population statistics allowed the fit of the tested models to be assessed.

Demographic models were simulated using the program BayeSSC. Sequence evolution was modeled using the following parameters: 25 years for the generation time, $7.5 \cdot 10^{-6}$ substitutions per site per generation (Ho et al., 2008) for the mutation rate, 0.9841 for the transition/transversion ratio, 0.205 and 10 for the theta and kappa parameters of the gamma distribution of rates along the sequence.

Under H0, the demographic model allowed a single NE-E population to evolve in size exponentially. Population statistics could be estimated at various points in time, corresponding to the age of the ancient populations considered (aUzPo, aBOO, HG and PWC). The values of the growth rate were drawn from a uniform prior distribution, such that the population has evolved from a Palaeolithic population of effective size 5,000 that lived 1,500 generations ago. The values for the modern-day (NEE or SAA) effective population size were drawn from a uniform distribution. We explored present-day effective population sizes between 100,000 and 30,000,000 for NEE and 1,000 to 500,000 for SAA.

Under H1, the demographic model allowed a single NE-E population to evolve in size exponentially and to be the recipient (sink population) of a migration from central Europe (source population). Population sizes of each of the present day sink population (NEE) and the central European source population were both drawn from a uniform distribution of population sizes varying from 100,000 to 15,000,000 individuals. Migration and divergence times were estimated from uniform distributions (from 2 to 139 generations for migration and from 620 to 2,600 generations for divergence). Three different percentages of migrants were tested: 10%, 50% and 75%.

Population statistics were calculated for 100,000 simulated genealogies using BayeSSC (<http://www.stanford.edu/group/hadlylab/ssc/index.html>). Six of these population measures (haplotype diversity and fixation indexes F_{ST}) were selected and compared to the observed measures calculated in Arlequin version 3.11 (Excoffier, Laval & Schneider, 2005; Table S5). Comparison of simulated and observed measures was carried out in an Approximate Bayesian Computation (ABC) framework (Beaumont et al., 2002, Ghirotto et al., 2010). The 1% of the simulations for which simulated population parameters exhibited the smallest Euclidian distance with observed population measures was retained to construct posterior distributions of population parameters. From these distributions, values of population parameters that optimize the likelihood of a given model were estimated. They were then implemented

in models in place of priors and 10,000 genealogies were generated in BayeSSC for these models. Goodness of fit of the different models tested was compared using Akaike's Information Criterion (AIC; Akaike et al., 1974) and Akaike's weights ω (Burnham & Anderson, 2002; Posada & Buckley, 2004). Treatment of BayeSSC output was carried out in R version 2.12.

RESULTS

Amplification success and authentication of the ancient DNA data

76 ancient individuals yielded HVR-I haplotypes (Table 1) that were considered unambiguous on the basis of seven criteria for aDNA authenticity (see Material and Methods). Amplification success was distributed among sites as follows: 21.5% for Uznyi Oleni Ostrov (nine haplotypes for 42 individuals processed), 66.7% for Popovo (two haplotypes for three individuals), 100% for Bolshoy Oleni Ostrov (23 haplotypes for 23 individuals) and 100% for Chalmny-Varre (42 haplotypes for 42 individuals).

The fact that samples from Bolshoy Oleni Ostrov and Chalmny-Varre yielded higher success rates (100%) is consistent with their younger age and the excellent general preservation state of the samples assessed by subjective observation (Figure S1). The cold climate of the Kola Peninsula where these two sites are located is advantageous for DNA preservation. Conversely, the less well-preserved samples from Uznyi Oleni Ostrov and Popovo were more problematic in terms of amplification success (amplification success rate of 24%). This can partly be explained by their apparent poorer preservation.

Table 1. Result overview for ancient mitochondrial DNA typing.

Sites and dates	Samples	HVRI sequence (np 16,056-16,409) ^a 16,000+	Hg ^b (HVRI)	Hg ^c (CR ^d)	Analyses ^e
Uznyi Oleni Ostrov (aUz) 7,500 yBP Mesolithic hunter-gatherers, Karelia, Russia (61°30'N 35°45'E)	UZOO-43	129c-189C-362C	U2e	U	E(2), Q
	UZOO-46	129c-189C-362C	U2e	U	E(2)
	UZOO-16	093C-356C	U4	U	E(2)
	UZOO-40	093C-356C	U4	U	E(2)
	UZOO-70	192T-256T-270T-318G	U5a	U	E(2)
	UZOO-77	311C-362C	H	H	E(2), I, C(22)
	UZOO-7	189C-223T-298C-325C-327T	C1	C	E(2)
	UZOO-8	189C-223T-298C-325C-327T	C1	C	E(2)
Popovo (aPo) 7,000 yBP (64°32'N 40°32'E)	Po4	356C	U4	U	E(2)
	Po2	093C-356C	U4	U	E(2)
Bolshoy Oleni Ostrov (aBOO) 3,500 yBP Bronze Age foragers, Kola Peninsula, Russia (68°58'N 35°05'E)	BOO49-3	093C-129A-134T-311C-356C	U4a1	U	E(2)
	BOO57-1	093C-129A-134T-311C-356C-390A/G	U4a1	U	E(2), I, C(8)
	BOO49-1	192T-256T-270T	U5a	U	E(2)
	BOO72-11	192T-256T-270T	U5a	U	E(2)
	BOO72-9	192T-256T-270T-399G	U5a1	U	E(1), Q
	BOO72-10	192T-256T-270T-399G	U5a1	U	E(2)
	BOO72-14	192T-256T-270T-399G	U5a1	U	E(2)
	BOO78-8	192T-256T-270T-399G	U5a1	U	E(2)
	BOO72-4	093C-126C-294T	T*	T	E(2), I, C()
	BOO49-2	223T-298C-327C	C*	C	E(2)
	BOO49-4	223T-298C-327C	C*	C	E(2)
	BOO57-3	223T-298C-327C	C*	C	E(2)
	BOO72-2	223T-298C-327C	C*	C	E(2)
	BOO72-7	223T-298C-327C	C*	C	E(2) I, C(4)
	BOO72-12	223T-298C-327C	C*	C	E(2)
	BOO72-5	148T-223T-288C-298C-311C-327C	C5	C	E(2)
	BOO72-6	148T-223T-288C-298C-311C-327C	C5	C	E(2)
	BOO49-6	223T-362C	D*	D	E(2)
	BOO72-13	223T-362C	D*	D	E(2)
	BOO72-15	223T-362C	D*	D	E(2) I, C(5)
	BOO49-5	129A-185T-223T-224C-260T-298C	Z1a	M	E(2)
	BOO72-3	129A-185T-223T-224C-260T-298C	Z1a	M	E(2)
	BOO72-1	129A-155G-185T-223T-224C-260T-298C	Z1a	M	E(2), I, C(6), Q

^a Variable nucleotide positions (np) when compared to the revised Cambridge Reference Sequence (rCRS, Andrews et al., 1999). Transitions are reported with lower-case letters, transversions with upper-case letters. ^b Haplogroup (Hg). ^c As determined by the GenoCoRe22 reaction. ^d CR, coding region, ^e E(), number of samples from which DNA was independently extracted; I, results replicated in an independent laboratory; C() number of HVRI clones; Q, HVRI DNA quantification performed.

Table 1 (continued). Result overview for ancient mitochondrial DNA typing.

Sites and dates	Samples	HVRI sequence (np 16,056-16,409) ^a 16,000+	Hg ^b (HVRI)	Hg ^c (CR ^d)	Analyses ^e
Chalmny-Varre (aCV) Saami graveyard, 18 th century AD, Kola Peninsula, Russia. (67°09'N 37°34'E)	ChV31	192T-256T-270T-304C-399G	U5a1	U	E(1)
	ChV43	192T-256T-270T-304C-399G	U5a1	U	E(1)
	ChV45	192T-256T-270T-304C-399G	U5a1	U	E(1)
	ChV6	144C-189C-270T	U5b1b1	U	E(2)
	ChV24	144C-189C-270T	U5b1b1	U	E(1)
	ChV25	144C-189C-270T	U5b1b1	U	E(2)
	ChV26	144C-189C-270T	U5b1b1	U	E(2)
	ChV35	144C-189C-270T	U5b1b1	U	E(1)
	ChV47	144c-144C-189C-270T-270t	U5b1b1	U	E(2)
	ChV15	144C-148T-189C-270T-335G	U5b1b1	U	E(1)
	ChV18	144C-148T-189C-270T-335G	U5b1b1	U	E(1)
	ChV30	144C-148T-189C-270T-335G	U5b1b1	U	E(2), I
	ChV40	144C-148T-189C-249C-270T	U5b1b1	U	E(2)
	ChV22	144C-148T-189C-249C-270T-335G	U5b1b1	U	E(1), Q
	ChV44	144C-148T-189C-249C-270T-335G	U5b1b1	U	E(2), Q
	ChV1	153A-298C	V7a	V	E(1)
	ChV2	153A-298C	V7a	V	E(2)
	ChV3	153A-298C	V7a	V	E(2)
	ChV4	153A-298C	V7a	V	E(2)
	ChV5	153A-298C	V7a	V	E(2)
	ChV8	153A-298C	V7a	V	E(2)
	ChV9	153A-298C	V7a	V	E(2)
	ChV10	153A-298C	V7a	V	E(2)
	ChV11	153A-298C	V7a	V	E(2)
	ChV13	153A-298C	V7a	V	E(2)
	ChV14	153A-298C	V7a	V	E(1)
	ChV16	153A-298C	V7a	V	E(2)
	ChV17	153A-298C	V7a	V	E(1)
	ChV21	153A-298C	V7a	V	E(2)
	ChV27	153A-298C	V7a	V	E(2)
	ChV28	153A-298C	V7a	V	E(2)
	ChV33	153A-298C	V7a	V	E(1)
	ChV39	153A-298C	V7a	V	E(2)
	ChV46	153A-298C	V7a	V	E(1)
	ChV7	114G-153A-298C	V7a	V	E(1)
	ChV36	114G-153A-298C	V7a	V	E(2)
	ChV37	114G-153A-298C	V7a	V	E(2)
	ChV38	114G-153A-298C	V7a	V	E(2), I
	ChV42	114G-153A-298C	V7a	V	E(2)
	ChV49	114G-153A-298C	V7a	V	E(2)
	ChV12	114G-153A-218T-298C	V7a	V	E(2)
	ChV23	114G-153A-218T-298C	V7a	V	E(1)

^a Variable nucleotide positions (np) when compared to the revised Cambridge Reference Sequence (rCRS, Andrews et al., 1999). Transitions are reported with lower-case letters, transversions with upper-case letters. ^b Haplogroup (Hg). ^c As determined by the GenoCoRe22 reaction. ^d CR, coding region, ^e E(), number of samples from which DNA was independently extracted; I, results replicated in an independent laboratory; C() number of HVRI clones; Q, HVRI DNA quantification performed.

Haplogroup distribution in modern-day populations of north Eurasia

The PCA biplot of the first two components (41.5% of the total variance, Figure 2) showed that the modern-day north Eurasian populations used for comparison cluster into three main groups: East Siberians, Europeans and Middle Easterners. Populations of Europe and Siberia appeared to spread along the first component axis (28.5% of the variance) according to their longitudinal position. As previously shown, populations of the east Siberian cluster are predominantly composed of haplogroups A, B, C, D, F, G, Y and Z whereas west Eurasian populations are characterized by high frequencies of mitochondrial haplogroups H, U, K, J, T, HV, V, W, X, and I (e.g., Wallace et al., 1999; Ingman et al., 2000; Maca-Meyer et al., 2001; Herrnstadt et al., 2002; Mishmar et al., 2003). The occurrence of ‘east Siberian’ and ‘west Eurasian’ haplogroups appears mutually exclusive in modern-day populations of the corresponding clusters. The European populations group tightly together confirming the strong and well documented mitochondrial homogeneity within the European metapopulation (Pult et al., 1994).

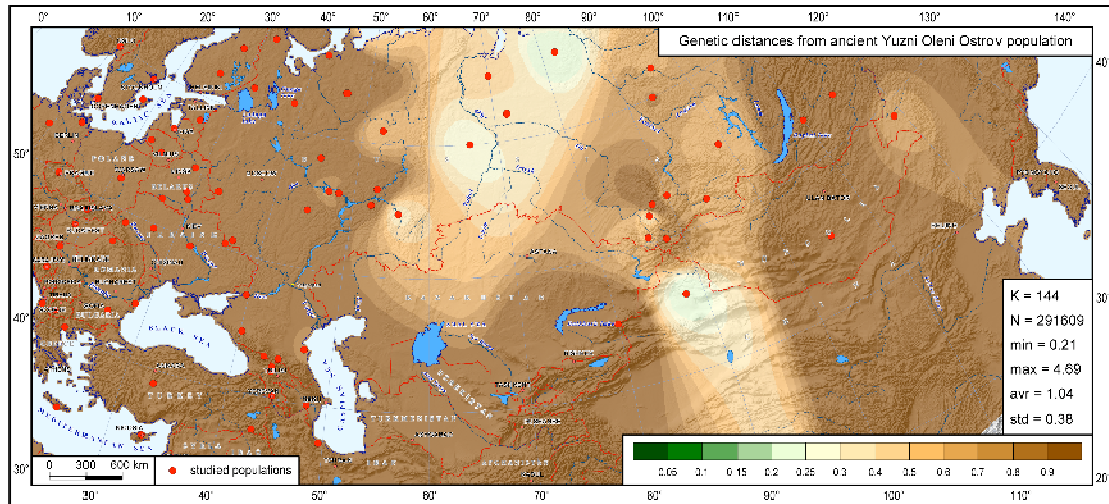
Mesolithic Uznyi Oleni Ostrov/Popovo compared to modern-day Eurasian populations

Mesolithic Uznyi Oleni Ostrov/Popovo clearly stood out from the three main groups of populations on the haplogroup frequency PCA graph (Figure 2), falling between the ‘European’ and ‘east Siberian’ clusters due to its mixed composition of haplogroups defined as ‘European’ (73%): U4 (37%), U2e (18%), U5a (9%), H (9%) and ‘Siberian’ C (27%). Haplogroup U4 is often regarded as a ‘western’ haplogroup, however, because of its prevalence in the Ural region and in western Siberia it should perhaps rather be considered as an ‘intermediate’ haplogroup. The presence of haplogroup U4, in addition to the mixed pattern of ‘western Eurasian’ (haplogroups U5a and U4) and ‘Siberian’ (haplogroup C) influences in Uznyi Oleni Ostrov are genetic characteristics that were also found in populations of the Urals (Bashkirs) and western Siberia (Khants, Mansi, Nganasan, Nenets, Selkups, Kets). Accordingly, these modern-day populations occupied an intermediate position on the biplot. The elevated component loadings (high frequencies) of haplogroups U4 and C in both Uznyi Oleni Ostrov/Popovo and Western Siberians contributed to the close grouping of these populations. The particular genetic link, i.e. low genetic distance, between the populations of Uznyi Oleni Ostrov/Popovo and modern-day populations of the Urals,

western and to a lesser extent southern Siberia is shown by a locally lighter colouring on the genetic distance map of north Eurasia (Figure 3A).

The genetic affinity of the ancient Uznyi Oleni Ostrov/Popovo samples with modern populations of Siberia was also detectable at the haplotypic level. Haplotype-sharing analysis showed that matches for Uznyi Oleni Ostrov were widely distributed across Eurasia with maximal percentages of shared haplotypes observed in the east and central Eurasian pools grouping Western Siberians (Khants/Mansi/Nenets/Selkups, 2.8%), southern Siberians (Altaians/Khakhassians/Shors/Tofalars, 2.2%) and population living south of the Urals (Bashkirs/Nogays/ Kazakhs, 2.0%) (Figure 4A). When only informative (non-basal) haplotypes were considered, very few matches could be observed in both west Eurasian and Siberian populations, with the Urals Bashkirs showing the highest percentage of shared haplotypes (0.4%) (Figure 4B). At the population level, genetic data from Uznyi Oleni Ostrov/Popovo suggests a strong genetic differentiation with modern-day North East Europeans.

A



B

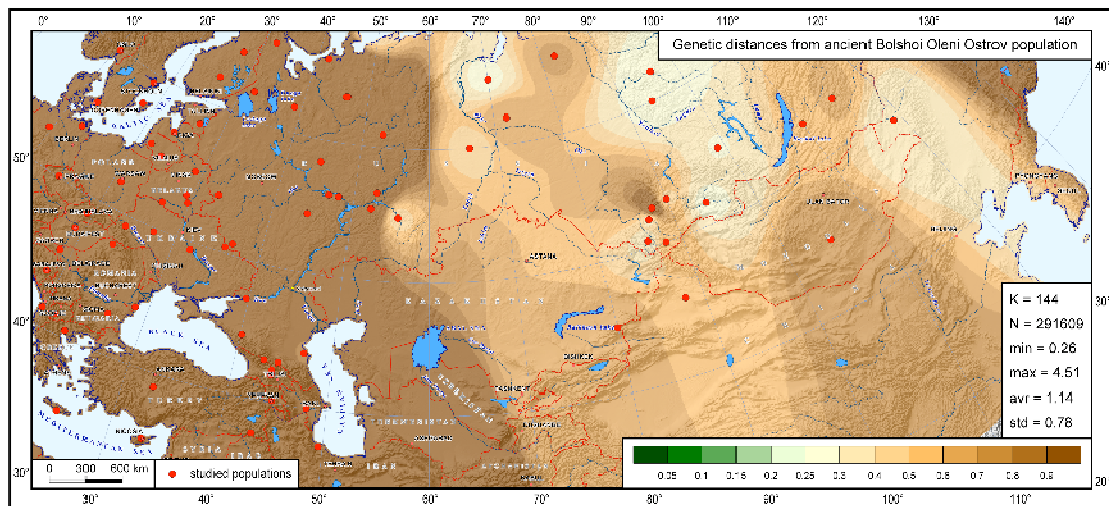


Figure 3. Map of genetic distances between the populations of Uznyi Oleni Ostrov/Popovo (A), Bolshoy Oleni Ostrov (B) and modern populations of Eurasia. Mitochondrial haplotypic data was used to compute genetic distances between 144 modern populations geographically delineated across Eurasia (red dots) and the eleven Mesolithic individuals from Uznyi Oleni Ostrov/Popovo (A) and the 23 Bronze Age individuals from Bolshoy Oleni Ostrov (B). The colour gradient represents the degree of similarity between the modern populations and the ancient population of interest, interpolated between sampling points: from ‘green’ for high similarity or small genetic distance to ‘brown’ for low similarity. ‘K’ designates the number of populations used for distance computation and mapping; ‘N’ represents the number of points in the grid used for extrapolation; ‘min’, ‘max’, ‘avr’ and ‘stdev’ correspond to the minimal, maximal, average and standard deviation values respectively of the computed distances between ancient and modern populations.

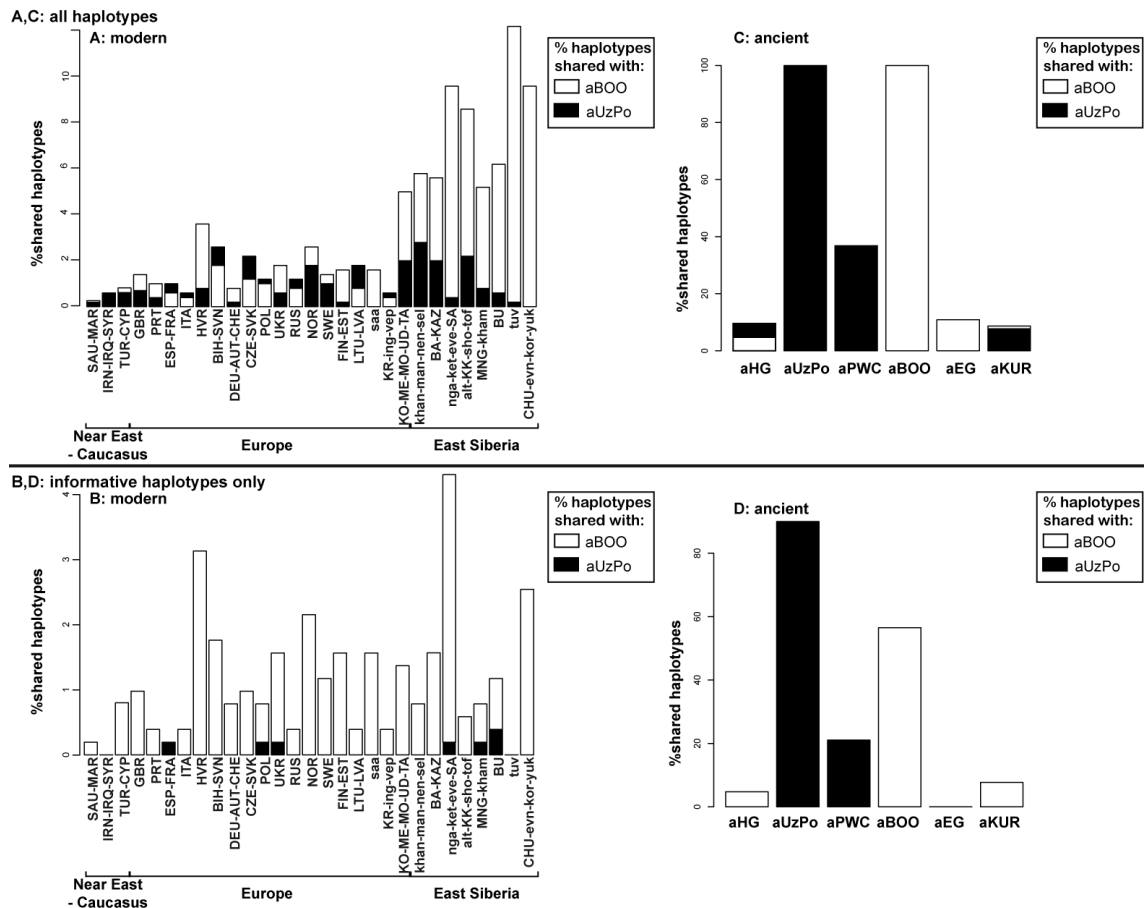


Figure 4. Percentages of Uznyi Oleni Ostrov/Popovo (black) and Bolshoy Oleni Ostrov (white) haplotypes matched in selected modern (A, B) and ancient (C, D) Eurasian population pools. The haplotype sharing analysis graph allows visualising the current distribution of the ancient haplotypes from Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov in modern populations from three main geographical regions: the Near-East/Caucasus, Europe, East Siberia (A, B). For each modern population, the number of occurrences of each haplotype from a given ancient population was divided by the modern population sample size ($N = 500$). The corresponding percentages of shared haplotypes are represented by a black bar for Uznyi Oleni Ostrov/Popovo haplotypes and by a white bar for Bolshoy Oleni Ostrov haplotypes. The same analyses were performed to compare the Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov haplotypes with haplotypes previously described in ancient populations (C, D). Figures A and C were constructed considering all haplotypes sequenced in the populations of Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov, whereas uninformative basal haplotypes were discarded in figures B and D. See Material and Methods for population abbreviations.

Bronze Age Bolshoy Oleni Ostrov compared to modern-day Eurasian populations

Around 3,500 years after Uznyi Oleni Ostrov/Popovo, the Bronze Age population of Bolshoy Oleni Ostrov showed a comparable heterogeneous composition of haplogroups from eastern and western Eurasia. Bolshoy Oleni Ostrov is composed of 39% ‘European’ haplogroups (including the ‘intermediate’ haplogroup U4): U5a (26%), U4 (9%), T (4%), and 61% ‘Siberian’ haplogroups: C (35%), Z (13%), D (13%), which resulted in a position close to Siberian populations on the PCA biplot between the east Siberian and western Eurasian clusters (Figure 2). The close genetic relationship between Siberians and Bolshoy Oleni Ostrov was also evident on the genetic distance map: the area representing low genetic distance from the population of Bolshoy Oleni Ostrov, in light brown, covered a broader area of Siberia than for Uznyi Oleni Ostrov, from the Urals to north east Mongolia (Figure 3B). The extant populations most genetically similar to the Bolshoy Oleni Ostrov population were found in Central and east Siberia, whereas the area of maximum similarity for Uznyi Oleni Ostrov lay in western Siberia (Figure 3A). However, it cannot be excluded that this observation could be an artefact of under-sampling of the Uznyi Oleni Ostrov/Popovo population.

At the haplotypic level, Bolshoy Oleni Ostrov showed a clear affinity with modern-day Eastern Eurasians, but also with populations of the Volga-Ural region (Figure 4A). Even though matches were found all over Siberia, the distribution was markedly skewed towards populations of east Siberia. This was caused by the presence of the basal C* and D* haplotypes in Bolshoy Oleni Ostrov, which displayed an east-centred distribution encompassing a broad range of Eurasian populations from the Volga–Ural region, west Siberia, central Siberia to east Siberia, Mongolia, and the Altai in the South. The maximum percentage of shared haplotypes was observed with the south Siberia Tuvinians (12.2%), and interestingly the haplotypes were found to be basal (Figure 4B). The maximal shared percentage of derived haplotypes was reached for the pool of north and east Siberian Nganasans/Kets/Evenks/Yakuts (4.2%), principally due to the presence of the particular U5a1 16192T-16256T-16270T-16399G and Z1a 16129A-16185T-16223T-16224C-16260T-16298C haplotypes. All ‘Siberian’ haplotypes as well as the ‘European’ U5a haplotype 16192T-16256T-16270T could be found in Buryats from the peri-Baikal region (Pakendorf et al., 2006; Derenko et al., 2007). This distinct genetic link between Bolshoy Oleni Ostrov and

present-day Buryats was supported by the occurrence of the rare haplotype C5 16148T-16223T-16288C-16298C-16311C-16327C in a single individual of the comparative dataset, who belongs to the Buryat population (Derenko et al., 2007).

Haplotype Z1a 16129A-16185T-16223T-16224C-16260T-16298C showed a different distribution from the other ‘Siberian’ haplotypes found in Bolshoy Oleni Ostrov. It was absent in Khants, Mansi, Nenets and Selkups of western Siberia but was found in northern Europe: in Swedes, Finns and Norwegians as well as in the Saami. Importantly, this Z1a haplotype was the only haplotype shared between an ancient population of north east European foragers (Bolshoy Oleni Ostrov), present-day Saami, and extant populations of NE-E (Scandinavia and the Volga-Ural region). Local persistence/continuity of mtDNA lineages since the Bronze Age was also shown by the U4 haplotypes: one exact match was found in a present-day Norwegian individual (Helgason et al., 2000), and closely related haplotypes were observed in modern-day Karelian and Komi individuals (Malyarchuk et al., 2004). On the whole, these results support a limited genetic legacy of the Bolshoy Oleni Ostrov gene pool in present-day North East Europeans and genetic similarities with populations of eastern Siberia and the Volga-Ural region.

18th century A.D. Chalmny-Varre compared to modern-day Eurasian populations

More than three thousand years after Bolshoy Oleni Ostrov, but in close geographical proximity, the population of 18th century A.D. (200 yBP) Chalmny-Varre demonstrated a markedly different genetic makeup. The population was well differentiated from both modern-day Europeans and Siberians, in accordance with the outlying position of Chalmny-Varre on the PCA biplot (Figure 2). The graph also showed that Chalmny-Varre is the closest population to the Saami. These results confirmed the tight relationship with modern-day Saami determined from the anthropological and archaeological analysis of the Chalmny-Varre graveyard. At the haplogroup level, a genetic signature of the modern-day Saami is their high frequencies of haplogroups V and U5b (41.5% and 48% respectively; Ingman & Gyllensten, 2007) this was also observed in Chalmny-Varre: V (64%), U5b (29%), and U5a (7%). A haplotypic signature of the Saami, the U5b1b1 haplotype 16144C-16189C-16270T, could also be detected in the population of Chalmny-Varre. This characteristic mutation motif has been coined the ‘Saami motif’ as it is very rare

outside Saami groups but highly frequent among them all (Sajantila et al., 1995). Moreover, the particular V haplotypes sequenced in the population of Chalmny-Varre also show affinities with the present-day Saami, as the basal 16153A-16298C haplotype was reported in all the modern-day Saami populations sampled so far. Haplogroup U5a, which is present but less frequent in modern-day Saami populations (2.3%), was also detected in 3 individuals from Chalmny-Varre (7.1%). Closely related U5a haplotypes were found in populations of east Europe (Russians and Byelorussians) as well as the Volga-Ural region (Tatars) and southern Siberia (Buryats; Malyarchuk et al., 2010a). Chalmny-Varre displays a close genetic relationship with modern-day Saami and to a lesser extent modern-day populations of east Europe.

Comparison among ancient Eurasian populations

Separated in time by 3,500 years, the Mesolithic Uznyi Oleni Ostrov/Popovo and Bronze Age Bolshoy Oleni Ostrov populations displayed a common affinity with modern-day populations of Siberia. Bolshoy Oleni Ostrov appeared genetically closer to Eastern Siberians than Uznyi Oleni Ostrov/Popovo on the basis of its position on the PCA biplot, which was attributed to its higher proportion (75% versus 39% in Uznyi Oleni Ostrov/Popovo) and variability of ‘Siberian’ haplogroups: haplogroups C, D and Z in Bolshoy versus haplogroup C in Uznyi Oleni Ostrov/Popovo. Haplotypic dissimilarity between Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov is shown by a total lack of overlap in haplotypes (Table 1, Figure 4C and D). These results support a significant differentiation between the population of Mesolithic Uznyi Oleni Ostrov/Popovo and Bronze-Age Bolshoy Oleni Ostrov.

The mtDNA structure of ancient European foraging populations was previously described for hunter-gatherers of Palaeolithic/Mesolithic central/eastern Europe (Bramanti et al., 2009; Krause et al., 2010) and Pitted-Ware culture individuals of Gotland Island, Sweden (Malmström et al., 2009). In agreement with these findings, the new data identified haplogroup U5a as a representative of central and NE-E’s Mesolithic mtDNA diversity. When included in the PCA, the ancient European hunter-gatherer and Pitted-Ware culture populations showed however a closer genetic proximity to European populations of NE-E than Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov (Figure 2).

The search for deep genetic ancestry of the Saami, i.e. haplotypic matches between 18th century A.D. (200 yBP) Chalmny-Varre and the two prehistoric sites Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov, revealed no matches for the U5b1b1 ‘Saami motif’ nor any particular V haplotype. This was interpreted as a lack of continuity between the prehistoric populations and Saami groups (both historical and modern-day). The significant difference in haplogroup composition between ancient populations and modern-day Saami was consistent with discrete positions on the PCA biplot. Haplogroup U5a is present in all three populations; although the U5a haplotypes were not shared among population, the closely related U5a haplotypes 16192T-16256T-16270T-16399G and 16192T-16256T-16270T-16304C-16399G were found in Bolshoy Oleni Ostrov and Chalmny-Varre. Of note, the Saami motif was absent from the previously described populations of European foragers (hunter-gatherers and Pitted-Ware culture individuals) and genetic continuity between Saami and Pitted-Ware culture individuals has been statistically rejected in a previous study (Malmström et al., 2009).

When compared to ancient populations of Siberia (Kurgans from south central Siberia; 3,800-1,600 yBP; Keyser et al., 2009 and Xiongnu individuals of the Egyin Gol Valley, Mongolia; 2,000 yBP; Keyser-Tracqui et al., 2003), only basal haplotypes were shared between Uznyi Oleni Ostrov/Popovo and Kurgans (U4, 16356C) and between Bolshoy Oleni Ostrov and Egyin Gol individuals (D, 16223T-16362C). These results illustrated the antiquity of the wide distribution of basal U4 and D haplotypes. The link shared between Bolshoy Oleni Ostrov and Siberian populations was strengthened further by the sharing of haplotype Z1a 16129A-16185T-16223T-16224C-16260T-16298C with a Bronze Age Kurgan individual of south central Siberia.

Testing population history hypothesis using Bayesian Serial SimCoal

The analysis of mtDNA at the haplogroup and haplotypic levels in the ancient populations of NE-E Uznyi Oleni Ostrov/Popovo, Bolshoy Oleni Ostrov and Chalmny-Varre led to the observation of: (1) a clear genetic continuity between the ancient population of Chalmny-Varre and present-day Saami; (2) a strong differentiation between Uznyi Oleni Ostrov/Popovo, Bolshoy Oleni Ostrov and present-day North East Europeans and Saami; (3) a greater differentiation between present-day populations (North East Europeans and Saami) and Bolshoy Oleni Ostrov

than the ancient foraging populations of European hunter-gatherers, Scandinavian Pitted-Ware Culture, and Uznyi Oleni Ostrov/Popovo. In contrast to (1), the relationships between other ancient and modern-day populations observed in (2) and (3) could have been the result of genetic drift or alternatively, pre- and post-Bronze Age migrations. Modern-day populations of NE-E lie between populations of western/central Europe and Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov on the PCA plot. This suggests a genetic input from the central Europe since the Bronze Age. We used the programme Bayesian Serial SimCoal (BayeSSC; Anderson et al., 2005) to test whether of genetic drift or migration best explains the population differentiation observed in (2) and (3) (Figure 5).

Among the models investigating genetic continuity between ancient and modern-day populations (hypothesis H0), the best model fits were obtained when testing modern-day Saami as the descendents of ancient European populations. However, previous analysis of shared haplotypes found little evidence for genetic continuity between ancient populations and present-day Saami. As a consequence, the results of the coalescent analysis could reflect the fact that lower haplotype diversities are observed in both ancient populations and extant Saami, whereas the modern-day north east European population was characterised by a higher haplotype diversity. The alleged low likelihood of genetic continuity between the prehistoric populations and modern-day populations of NE-E was supported statistically by a better fit for the model involving a post-Bronze Age migration from central Europe into NE-E (hypothesis H1) than for genetic continuity with either modern-day North East Europeans or Saami. When considering post-Bronze Age migration models in which ancient populations of European foragers (European hunter-gatherers and Scandinavian Pitted-Ware culture individuals) were in genetic continuity with either Uznyi Oleni Ostrov/Popovo (H1a) or Bolshoy Oleni Ostrov populations (H1b), a better fit was obtained for Uznyi Oleni Ostrov/Popovo (H1a). Overall, hypothesis testing using BayeSSC supports demographic models of genetic discontinuity: modern-day populations of NE-E and Saami are not the direct descendents of prehistoric populations of European hunter-gatherers. Instead BayeSSC analyses favour an important role of post-Bronze Age migrations in the composition of the north east European mtDNA gene pool. These analyses also confirmed that Uznyi Oleni Ostrov/Popovo better fits within the genetic continuum of European populations

of foragers than Bolshoy Oleni Ostrov, thus suggesting distinct population histories (Table 2).

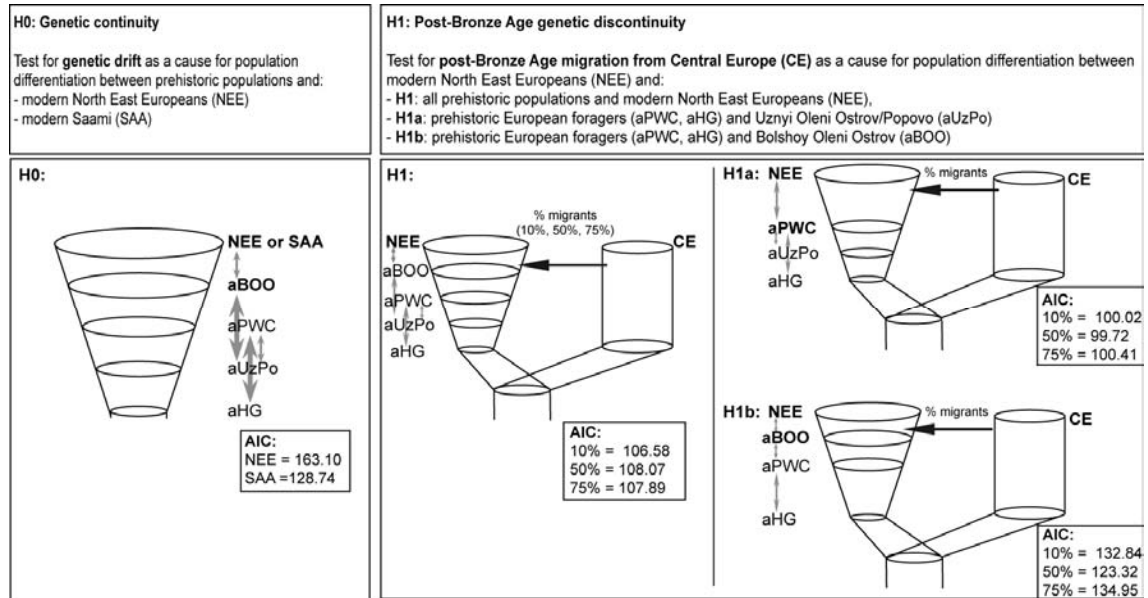


Figure 5. Graphical representation of the demographic models compared by coalescent simulation analysis and their associated Akaike information Criterion, AIC. Bold population labels indicate that the haplotype diversity of the corresponding populations was used in the analyses for parameter estimation and model selection. Fixation indexes of populations linked by grey double arrows were used in the analyses for parameter estimation and model selection. H0 tests genetic continuity between prehistoric populations (aBOO, aPWC, aUzPo and aHG) and a modern population (NEE or SAA). AICs are calculated for both the models testing genetic continuity with NEE and SAA. H1 tests models considering a post-Bronze Age migration from central Europe to North East Europe allowing for the percentage of migrants to equal 10%, 50% and 75%. H1 tests genetic continuity between all prehistoric populations (aBOO, aPWC, aUzPo and aHG), aBOO and aUzPo are excluded in models H1a and H1b, respectively. AICs are calculated for all models and all percentages of migrants.

Table 2. Comparison of the demographic models simulated with BayeSSC by relative model likelihood Akaike weights ω for all tested models ordered from the most to the least likely.

Demographic model and hypothesis tested	Akaike weights ω
H1a_50	$3.80e^{-1}$
H1a_10	$3.27e^{-1}$
H1a_75	$2.69e^{-1}$
H1_10	$1.23e^{-2}$
H1_75	$6.39e^{-3}$
H1_50	$5.84e^{-3}$
H1b_50	$2.85e^{-6}$
H0_SAA	$1.90e^{-7}$
H1b_10	$2.44e^{-8}$
H1b_75	$8.50e^{-9}$
H0_NEE	$6.56e^{-15}$

The tested demographic models are ordered from the most likely (as indicated by higher values of the Akaike weight) to the least likely (as indicated by lower values of the Akaike weight). The demographic models are referred to as follows: H0 models consider genetic continuity between ancient and modern populations of north east Europeans (_NEE) or Saami (_SAA); H1 models take into account migrations from central Europe that occurred after the Mesolithic (model H1a) and after the Bronze Age (model H1b). The tested proportions of migrants are 10% (_10), 50% (_50) and 75% (_75).

DISCUSSION

Genetic discontinuity between prehistoric populations and Saami

Genetic continuity with modern-day Saami was evident for the 18th century A.D. (200 yBP) Chalmny-Varre individuals. The widespread modern-day distribution of U5b1 and V lineages makes it difficult to identify the places of origin for the founders of the Saami (Tambets et al., 2004). Despite its clear association with Saami ancestry, the ‘Saami motif’ has also been found at low frequency (below 1%) in a wide range of non-Saami populations of Europe. This justified sampling ancient populations of NE-E such as Uznyi Oleni Ostrov/Popovo, and Bolshoy Oleni Ostrov in the modern-day homeland of the Saami. The absence of these lineages in ancient populations of European foragers suggests that the migration(s) that brought U5b1 and V north occurred later than the Bronze Age or that their genetic impact on the surrounding populations was weak. Another possibility is that these lineages reached Fennoscandia from western Europe along the western coast of Norway, hence in isolation from the post-glacial populations of east and north east Europe (Tambets et

al., 2004). The only way to test this hypothesis and track Saami-specific lineages is by sampling ancient populations along the proposed alternative western migration route into sub-arctic Europe. This would also provide insight into the demographic history of the Saami that is difficult to reconstruct on the basis of modern genetic data alone. Saami mtDNA diversity has supposedly been strongly influenced by founder events, multiple bottlenecks and reproductive isolation, probably in response to the challenging conditions of life in the subarctic taiga/tundra. Unless changes in their genetic makeup are followed in temporally sampled populations, the origins and population history of the Saami will certainly remain veiled with mystery.

Siberian influence in the mtDNA gene pool of North East Europeans

Samples from Mesolithic Uznyi Oleni Ostrov/Popovo and Bronze Age Bolshoy Oleni Ostrov were found to demonstrate limited genetic continuity with modern-day populations of NE-E. A cause of this discontinuity is a large proportion, in ancient populations, of haplogroups of east Eurasian origin, namely C, D and Z, which indicate a genetic link with diverse populations of Siberia and the Ural region. Haplogroups C and D are the most common haplogroups of northern, central and eastern Asia. They are thought to have originated in eastern Asia, from which they expanded throughout Asia in multiple migration events after the Late Glacial Maximum (~20,000 yBP; Derenko et al., 2010). Interestingly, genetic links were identified between Bolshoy Oleni Ostrov and modern-day Buryats of the peri-Baikal region, east Siberia, which is thought to have been the place of origin of most of these migrations. These migrations certainly led to the establishment of Siberian populations, which display relatively high frequencies in haplogroups C and D.

Genetic link with extant populations of the Volga-Ural

The sharp western boundary for the distribution of the ‘eastern Siberian’ lineages found in Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov lies in the Volga-Ural region. In this area, ‘eastern Siberian’ haplogroups are distributed as follows (Malyarchuk et al., 2010a): C, 0.3 to 11.8%; Z, 0.2 to 0.9%; and D, 0.6 to 12%. The ‘Siberian’ lineages Z1 and D5 are also found in modern-day Saami, with highest frequencies being reached in Finnish Saami (15.9%; Tambets et al., 2004). No ‘east Siberian’ lineages could be found in Chalmny-Varre either because of a bias in sampling or because of low frequencies of these lineages in Kola Peninsula Saami.

The presence of 'east Siberian' lineages in the populations of Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov indicates that the eastern genetic influence seen in populations of NE-E pre-dates the relatively recent westward expansions from the East of, for example, the Huns and the Mongols (4th to 15th century A.D.; Grousset, 1970). It can be then hypothesised that the east Siberian lineages observed today in east Europe represent a genetic legacy of prehistoric movements from Siberia rather than that of more recent 'warlike' movements, less likely to have impacted the mitochondrial gene pool.

The earliest migration from eastern Eurasia detected in the archaeological record occurred during the Tardiglacial (18,000 – 11,650 yBP) and is thought to explain the presence of final Upper Palaeolithic settlements exhibiting a clear Siberian tradition at the fringe of the ice-sheet (Talitskiy, upper level of Krutaya Gora, Medvezhja cave, in western Urals and Karacharovo in western Russia). A second migration was suggested to have moved from Siberia via the southern Urals to the Pechora and Vychegda basins (north west Urals) where it was associated with the appearance of the Kama culture at the beginning of the Holocene Atlantic period (~8,000 yBP) (Kozłowski & Bandi, 1984). This latter migration, which might have been responsible for the spread of Siberian lineages to Bolshoy Oleni Ostrov, was proposed earlier to have introduced Z1 to the populations of the Volga-Ural region and the Saami on the basis of coalescent ages calculated for Z1 lineages in modern-day eastern Europe (Tambets et al., 2004).

The presence of Z1 in samples from Bolshoy Oleni Ostrov establishes a direct genetic link between a Bronze Age population in NE-E and modern-day populations of the Volga-Ural region and Saami. The fact that Z1 was found in Bolshoy Oleni Ostrov but not in association with any Saami-specific haplotype further suggests that it must have had an independent origin and could have been integrated into the Saami gene pool via a separate migratory event. Z1 was not found in the population of 8,000 yBP Uznyi Oleni Ostrov/Popovo. This can be explained either by the bias introduced by low sample-size or by the fact that this population was genetically isolated from the migration involving Z1 carriers. The Z1 lineage could also have been brought into NE-E by a later migration that occurred between 8,000 yBP and 3,500 yBP, time at which it contributed to the Bolshoy Oleni Ostrov mtDNA gene pool. This scenario would be in agreement with the establishment of the Kama culture being associated with the spread of 'east Siberian' mtDNA lineages, as Bolshoy Oleni

Ostrov postdates the formation of the Kama culture in NE-E. It would be interesting to directly characterise the genetic structure of populations representative of the prehistoric migration from south Urals in order to assess their genetic link with the other sampled prehistoric populations of Europe and their contribution to the recolonisation of subarctic Europe.

Singularity of Uznyi Oleni Ostrov within the Mesolithic diversity

The genetic affinity OF Uznyi Oleni Ostrov/Popovo with Western Siberians is based on the presence of haplogroup C1, which is virtually absent from modern-day European populations and is generally considered to have an Asian origin. Five geographically restricted monophyletic clades (C1a, C1b, C1c, C1d, C1e) compose haplogroup C1, which is widely distributed from Iceland (C1e, Ebenesersdottir et al., 2011), east Asia (C1a, Derenko et al, 2007) to the Americas, where it is found at its highest frequencies (C1b, C1c and C1d). Very few haplotypes closely related to the Uznyi Oleni Ostrov C1 haplotype 16189C-16223T-16298C-16325C-16327T could be detected in Eurasian individuals (Canary Island, Germany, Iceland, Bashkirs, India; Rando et al., 1999; Pfeiffer et al., 1999; Helgason et al., 2000; Helgason et al., 2003; Ebenesersdottir et al., 2011; Bermisheva et al., 2002; Metspalu et al., 2004). In the absence of further information from the coding region of both the modern-day relatives and the Uznyi Oleni Ostrov C1 haplotype, its closest phylogenetic relatives and their geographical distribution cannot be identified.

All three individuals from Uznyi Oleni Ostrov showed identical C1 haplotypes, which means that a close maternal kinship between these individuals cannot be rejected. By overestimating the frequency of haplogroup C1 in the populations of Uznyi Oleni Ostrov/Popovo, this bias possibly led to an overestimation of the genetic affinity with modern-day Western Siberians detected in the haplogroup frequency-based analyses. In addition to the lack of diversity in 'east Siberian' lineages, the small sample size of the Uznyi Oleni Ostrov population casts doubts on its genetic affinity with Western Siberians. The poor preservation of the Uznyi Oleni Ostrov remains allowed only 21.5% of the samples to be reliably typed. Under-sampling certainly causes bias in the analysis of the haplogroup data of this population. Moreover, coalescent simulation analyses showed that the population of Uznyi Oleni Ostrov/Popovo fits better with the previously described Mesolithic European diversity than Bolshoy Oleni Ostrov. Given these circumstances, the genetic link between the

Mesolithic hunter-gatherers and Western Siberians seems less strongly supported and the Uznyi Oleni Ostrov C1 haplotype may in fact represent a distinct European-specific lineage not described before, along the same line as the newly identified C1e sub-clade, so far restricted to Iceland (Ebenesersdottir et al., 2011).

The absence of haplogroup C1 in other ancient and modern-day European populations, suggests that the spread of haplogroup C did not reach further west into central Europe. The C1 lineage detected here may rather represent a genetic outlier at the periphery of its proposed origin. The C1 haplotype may have been conserved by a relative isolation of the population of Uznyi Oleni Ostrov and/or the effects of drastic post-Mesolithic demographic events, such as reduction of population size or population replacement. Odontometric (Jacobs, 1992) and craniometric (von Cramon-Taubadel & Pinhasi, 2011) analyses of samples from Uznyi Oleni Ostrov support the idea that Uznyi Oleni Ostrov was part of a closed mating system in isolation with other Mesolithic populations of Scandinavia. This does not exclude a common origin for European foragers but highlights the differentiating role of post-glacial founder effects followed by reproductive isolation in the Palaeolithic and Mesolithic.

Importance of haplogroup U in the Palaeolithic/Mesolithic substratum

An outstanding difference between present-day Europeans and all ancient foraging populations of Europe described at the mtDNA level is the large proportion of haplogroup U in ancient populations: from 35% in Bolshoy Oleni Ostrov to 73% in European hunter-gatherers (Bramanti et al., 2009; Krause et al., 2010). In comparison, haplogroup U represents 20 to 23% of the present-day European mtDNA diversity (Richards et al., 2000). Haplogroup U is widely distributed in Europe, west Siberia, south west Asia, the Near East and north Africa, and encompasses the locations of all ancient populations of European foragers mentioned above. The widespread distribution and high variability of haplogroup U in modern-day as well as in prehistoric populations are consistent with the description of haplogroup U as one of the oldest European haplogroups. On the basis of modern genetic data, haplogroup U was proposed to have originated in the Near East and to have been prevalent in the populations of anatomically modern humans who first colonised Europe during the Palaeolithic (Richards et al., 2000). Estimation of the coalescent age of haplogroup U5, for example, lie within the Upper Palaeolithic and range from 30,400 yBP (95% confidence Interval, CI: 21,800 – 39,300; Malyarchuk et al., 2010b), 30,700 (95% CI:

21,400 – 40,500; based on the ρ statistics; Soares et al., 2009) to 36,000 (95% CI: 25,300 – 47,200; based on maximum likelihood tree; Soares et al., 2009). This study contributes to the assessment of the post-glacial geographical and temporal distribution of haplogroup U in Europe as well as its past diversity, by adding temporal and geographical coordinates for particular lineages as well as reporting U lineages that were not previously described in modern-day or ancient populations. In addition to two previous studies, this work manifests haplogroup U lineages as characteristic representatives of the Palaeolithic/Mesolithic genetic substratum of central and northern Europeans. In particular, the most characteristic feature of both Mesolithic populations of the peri-Baltic area (Uznyi Oleni Ostrov/Popovo and Scandinavian Pitted-Ware Culture individuals) was the high frequencies observed for haplogroup U4 combined with a smaller proportion of haplogroup U5a and quasi-absence of haplogroup U5b. It is possible that this feature was shared among contemporaneous groups of NE-E as a whole. These haplogroups might represent the Mesolithic genetic substratum of modern-day population of western Siberia as U4 and U5a are still observed at high frequencies in this area.

Limited genetic impact of ancient north east European foragers on modern-day populations

The sites of Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov show considerable differences in haplogroup frequencies compared to present-day North East Europeans. These differences are most strikingly exemplified by the current diversity and predominance of haplogroup H, which includes 44.5 to 48.2% in European haplogroups (Richards et al., 2000). This supports the idea of an important role of Bronze-Age or post-Bronze Age demographic events in the shaping of today's north east European genetic makeup. The migrations that brought the mtDNA lineages currently prevalent in NE-E, e.g. Germanic, Slavic or Scandinavian incursions, are believed to have mostly originated in the West and in the South.

It also seems that an important part of the past mtDNA variability has been lost since the Bronze Age as some derived haplotypes detected in Uznyi Oleni Ostrov/Popovo and Bolshoy Oleni Ostrov found no identical match in the comparative database of modern-day Eurasian populations. In order to confirm that these rare derived haplotypes did not arise from artefactual mutations caused by post-mortem

DNA damage, increase in sequencing depth and cloning were carried out at critical variable positions. The probable extinctions of particular lineages could be the result of lineage dilution by migration waves introducing new mtDNA lineages into NE-E. Alternatively, the random loss of genetic diversity by drift is more likely to be accelerated in small and isolated populations. Any of these dynamic population processes in the past cannot be inferred from the extant genetic makeup of North East Europeans. While studies on modern-day population are able to detect low diversity in isolated populations, ancient DNA is able to reveal these lost mtDNA lineages and unravel cryptic demographic processes that participated in the making of the maternal gene pool of North East Europeans.

LIST OF SUPPLEMENTARY MATERIALS

Figure S1: Pictures of selected samples from Uznyi Oleni Ostrov, Popovo, Bolshoy Oleni Ostrov and Chalmny-Varre. All samples yielded consistent haplotypes except ACAD4719.

Figure S2A-F: Clone sequences.

Figure S3: Additional models tested through coalescent simulation (BayeSSC) with associated Akaike Information Criterion (AIC) and relative model likelihood Akaike weights (ω).

Table S1: Sequences of primers and probes used for sequencing of HVRI and typing of mtDNA coding region SNPs.

Table S2: Results of SNP typing in the mtDNA coding region using the GenoCoRe22 SNaPshot assay. SNPs typed on the L-strand are reported in capital letters in the reference rCRS profile, whereas SNPs typed on the H-strand are reported in small letters. Missing data signifies allelic dropout or fluorescence signal below the background threshold (100 relative fluorescent units, rfu). 'g/a' indicates the presence of a mixed signal for the position interrogated. A mixed signal was repeatedly obtained at position 8994 (haplogroup W) with the detection of an additional G base. However, the rest of the profile never could phylogenetically support the presence of the G base at this particular position. For each individual, profiles were obtained from two independent extracts, except for individuals BOO72-9, ChV30, ChV38, for which a second samples was not available and for UZOO-77, BOO57-1, BOO72-10, BOO72-4, BOO72-7, BOO72-15, BOO72-1, ChV31, ChV43, ChV45, ChV24, ChV35, ChV15, ChV18, ChV22, ChV1, ChV14, ChV17, ChV33, ChV46, ChV7, for which the second individual was extracted in an independent laboratory.

Table S3: Results of quantitative PCR

Table S4: Description and references for the comparative population datasets and pools.

Table S5: Population parameters used in BayeSSC coalescent simulation analyses. Sample sizes and ages of the sequences were implemented in the demographic models used for genealogy simulations. Haplotype diversity and fixation indexes were used in the analyses of the simulation for parameter estimation and model selection.

ACKNOWLEDGEMENTS

We warmly acknowledge Oleg Balanovsky, Valery Zaporozhchenko and Elena Balanovska of the Russian Academy of Medical Sciences, Moscow, Russia, for compiling the comparative database of modern-day populations, performing the genetic distance maps and having actively participated in the haplotype sharing analysis. We thank Valery Zaporozhchenko for searching the comparative modern database. We are particularly grateful to Oleg Balanovsky for his valuable comments and suggestions. We acknowledge Guido Brandt of the Institute of Anthropology, Johannes Gutenberg-University, Mainz, Germany, for the independent replication of the analysis and cloning. We thank Valery Khartanovich (Kunstkamera Museum, St Petersburg, Russia), Alexandra Buzhilova (Institute for Archaeology, Russian Academy of Sciences, Moscow, Russia), Sergey Koshel (Faculty of Geography, Moscow State University, Moscow, Russia) for providing the samples. We thank Marta Kasper for help with project logistics; Kurt Alt, Jeremy Austin, Jessica Metcalf, David Soria, and Andrew Clarke for helpful comments.

REFERENCES

1. Adler, C. J., Haak, W., Donlon, D. & Cooper, A. (2010). Survival and recovery of DNA from ancient teeth and bones. *J Archaeol Sci* 38, 956-964.
2. Akaike H. (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19(6), 716-723.
3. Anderson, C., Ramakrishnan, U., Chan, Y., Hadly, E. (2005). Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics* 21, 1733-1734.
4. Andrews, R., Kubacka, I., Chinnery, P., Lightowlers, R., Turnbull, D., Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23, 147.
5. Beaumont, M.A., Zhang, W., Balding, D.J. (2002). Approximate Bayesian computation in population genetics. *Genetics* 162, 2025-2035.

6. Beckman, L., Sikström, C., Mikelsaar, A.V., Krumina, A., Ambrasiene, D., Kucinskas, V., Beckman, G. (1998). Transferrin variants as markers of migrations and admixture between populations in the Baltic Sea region. *Hum Hered* 48, 185-191.
7. Bermisheva, M., Tambets, K., Villems, R., Khusnutdinova, E. (2002). [Diversity of mitochondrial DNA haplotypes in ethnic populations of the Volga-Ural region of Russia]. *Mol Biol (Mosk)* 36, 990-1001.
8. Bramanti, B., Thomas, M., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M., Jankauskas, R., Kind, C., Lueth, F., Terberger, T., Hiller, J., Matsumara, S., Forster, P., Burger, J. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326, 137-140.
9. Burnham K. P., Anderson D. R (2002) Model selection and multimodel inference: A practical information-theoretic approach, 2nd edition. New York: Springer.
10. Campos, P. F., Kristensen, T., Orlando, L., Sher, A., Kholodova, M. V., Götherström, A., Hofreiter, M., Drucker, D. G., Kosintsev, P., Tikhonov, A., Baryshnikov, G. F., Willerslev, E., Gilbert, M. T. (2010). Ancient DNA sequences point to a large loss of mitochondrial genetic diversity in the saiga antelope (*Saiga tatarica*) since the Pleistocene. *Mol Ecol* 19, 4863-4875.
11. Cavalli-Sforza, L. L. , Menozzi, P. & Piazza, A. The History and Geography of Human Genes (Princeton Univ. Press, Princeton NJ, 1994).
12. Chairkina and, Kosinskaia (2009). in *Ceramics Before Farming. The Dispersal of Pottery Among Prehistoric Eurasian Hunter-Gatherers*. Jordan, P., Zvelebil, M. (eds.). Walnut Creek, CA: Left Coast Press.
13. Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Dambueva, I., Perkova, M., Dorzhu, C., Luzina, F., Lee, H.K., Vanacek, T., Villems, R., Zakharov, I. (2007). Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet* 81 (5), 1025-41.
14. 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.
15. Dolukhanov, P. (1997). The Pleistocene-Holocene transition in Northern Eurasia: Environmental changes and human adaptations. *Quat Internat*, 181-191.
16. Ebenesersdóttir, S.S., Sigurðsson, A., Sánchez-Quinto, F., Lalueza-Fox, C., Stefánsson, K., 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.
17. Ermini, L., Olivieri, C., Rizzi, E., Corti, G., Bonnal, R., Soares, P., Luciani, S., Marota, I., De Bellis, G., Richards, M.B., Rollo, F. (2008). Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr Biol* 18, 1687-1693.
18. Excoffier, L. G. Laval, S. Schneider. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50.
19. Gamble, C., Davies, W., Pettitt, P., Richards, M. (2004). Climate change and evolving human diversity in Europe during the last glacial. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 243-254.

20. Ghirotto, S., Mona, S., Benazzo, A., Paparazzo, F., Caramelli, D., Barbujani, G. (2010). Inferring genealogical processes from patterns of Bronze-Age and modern DNA variation in Sardinia. *Mol Biol Evol* 27, 875-886.
21. Grousset, R. (1970) *The Empire of the Steppes: History of Central Asia*. Ed. Rutgers University Press.
22. Guglielmino, C.R., Piazza, A., Menozzi, P., Cavalli-Sforza, L.L. (1990). Uralic genes in Europe. *Am J Phys Anthropol* 83, 57-68.
23. Gurina N.N. (1956). Oleneostrovski Mogil'nik. In *Materialy i Issledovaniya po Arkheologii SSSR*. Moscow: Nauka, Akademia Nauk SSSR.
24. Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K.W., Burger, J. (2005). Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310, 1016-1018.
25. Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., Der Sarkissian, C.S., Brandt, G., Schwarz, C., Nicklisch, N., Dresely, V., Fritsch, B., Balanovska, E., Villems, R., Meller, H., Alt, K.W., Cooper, A., Genographic consortium. (2010). Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. *PLoS Biol* 8, e1000536.
26. Helgason, A., Sigurethardóttir, S., Gulcher, J.R., Ward, R., Stefánsson, K. (2000). mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet* 66, 999-1016.
27. 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.
28. Herrnstadt, C., Elson, J., L., Fahy, E., Preston, G., Turnbull, D., M., Anderson, C., Ghosh, S., S., Olefsky, J., M., Beal, M., F., Davis, R., E., Howell, N. (2002). Reduced-median-network analysis of complete mitochondrial DNA coding region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* 70, 1152-1171.
29. Ho, S., Endicott, P. (2008). The crucial role of calibration in molecular date estimates for the peopling of the Americas. *Am J Hum Genet* 83, 142-146; author reply 146-147.
30. Hofreiter, M., Stewart, J. (2009). Ecological change, range fluctuations and population dynamics during the Pleistocene. *Curr Biol* 19, R584-594.
31. Ingman, M., Kaessmann, H., Pääbo, S., Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* 408, 708-713.
32. Ingman, M., Gyllensten, U. (2007). A recent genetic link between Sami and the Volga-Ural region of Russia. *Eur J Hum Genet* 15, 115-120.
33. Jacobs K. (1992). Human population differentiation in the peri-Baltic Mesolithic: the odontometrics of Oleneostrovskii mogilnik (Karelia). *Human Evolution* Vol. 7:N.4 (1992) (33-48).
34. Jacobs K. (1995). Returning to Oleni' ostrov: Social, Economic, and Skeletal Dimensions of a Boreal Forest Mesolithic Cemetery. *Jour Anthropol Archaeol* 14, 4, 359-403.
35. Keyser, C., Bouakaze, C., Crubézy, E., Nikolaev, V., 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.
36. 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.

37. Kozłowski J., Bandi H.G. (1984). The paleohistory of circumpolar arctic colonization. *Arctic* 37(4) 359-372.
38. Krause, J., Briggs, A., 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.
39. Lalueza-Fox, C., Sampietro, M., Gilbert, M., Castri, L., Facchini, F., Pettener, D., Bertranpetit, J. (2004). Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians. *Proc Biol Sci* 271, 941-947.
40. Lappalainen, T., Laitinen, V., Salmela, E., Andersen, P., Huoponen, K., Savontaus, M.L., Lahermo, P. (2008). Migration waves to the Baltic Sea region. *Ann Hum Genet* 72, 337-348.
41. Maca-Meyer, N., González, A.M., Larruga, J.M., Flores, C., Cabrera, V.M. (2001). Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2, 13.
42. Malmström, H., Svensson, E.M., Gilbert, M.T., Willerslev, E., Götherström, A., Holmlund, G. (2007). More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA. *Mol Biol Evol* 24, 998-1004.
43. Malmström, H., Gilbert, M., Thomas, M., Brandström, M., Storå, J., Molnar, P., Andersen, P., 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.
44. Malyarchuk, B., Derenko, M., Grzybowski, T., Lunkina, A., Czarny, J., Rychkov, S., Morozova, I., Denisova, G., Miścicka-Sliwka, D. (2004). Differentiation of mitochondrial DNA and Y chromosomes in Russian populations. *Hum Biol* 76, 877-900.
45. Malyarchuk, B., Derenko, M., Denisova, G., Kravtsova, O. (2010a). Mitogenomic diversity in Tatars from the Volga-Ural region of Russia. *Mol Biol Evol* 27, 2220-2226.
46. Malyarchuk, B., Derenko, M., Grzybowski, T., Perkova, M., Rogalla, U., Vanecek, T., Tsybovsky, I. (2010b). The peopling of Europe from the mitochondrial haplogroup U5 perspective. *PLoS One* 5, e10285.
47. Melchior, L., Lynnerup, N., Siegismund, H., Kivisild, T., Dissing, J. (2010). Genetic Diversity among Ancient Nordic Populations. *PLoS One* 5, e11898.
48. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D. M., Gilbert, M. T., Endicott, P., Mastana, S., Papiha, S. S., 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.
49. Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., Sukernik, R.I., Olckers, A., Wallace, D.C. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A* 100, 171-176.
50. Mooder, K., Schurr, T., Bamforth, F., Bazaliiski, V., Savel'ev, N. (2006). Population affinities of Neolithic Siberians: a snapshot from prehistoric Lake Baikal. *Am J Phys Anthropol* 129, 349-361.

51. Noonan, J.P., Hofreiter, M., Smith, D., Priest, J.R., Rohland, N., Rabeder, G., Krause, J., Detter, J.C., Pääbo, S., Rubin, E.M. (2005). Genomic sequencing of Pleistocene cave bears. *Science* 309, 597-599.
52. O'Shea J., Zvelebil M., Oleneostrovskii Mogilnik: Reconstructing Social and Economic Organisation of Prehistoric Hunter-Fishers in Northern Russia. *Journal of Anthropological Archaeology* 3, mo. 1 (1984): 1-40.
53. Oshibkina, S.V. (1999). Tanged Points Culture in Europe. Ed. Kozłowski S.K., Gurba J., Zaliznyak L.L.
54. Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annu Rev Genet* 38, 645-679.
55. Pakendorf, B., Novgorodov, I.N., Osakovskij, V.L., Danilova, A.P., Protod'jakonov, A.P., Stoneking, M. (2006). Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum Genet* 120, 33.
56. Pfeiffer, H., Brinkmann, B., Hühne, J., Rolf, B., Morris, A. A., Steighner, R., Holland, M. M., 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.
57. Pliss, L., Tambets, K., Loogväli, E.L., Pronina, N., Lazdins, M., Krumina, A., Baumanis, V., Villems, R. (2006). Mitochondrial DNA portrait of Latvians: towards the understanding of the genetic structure of Baltic-speaking populations. *Ann Hum Genet* 70, 439-458.
56. Posada, D., Buckley, T.R. (2004). Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst Biol* 53, 793-808.
57. Price T.D. (1991). The Mesolithic of Northern Europe. *Annu Rev Anthropol* 20, 211-233.
58. Rando, J.C., Cabrera, V.M., Larruga, J.M., Hernandez, M., Gonzalez, A.M., Pinto, F., Bandelt, H.J. (1999). Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* 63, 413-428.
59. Renfrew, C. (2010). Archaeogenetics--towards a 'new synthesis'? *Curr Biol* 20, R162-165.
60. Richards, M., Côrte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H., Sykes, B. (1996). Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59, 185-203.
61. Richards, M., Macaulay, V., Bandelt, H., Sykes, B. (1998). Phylogeography of mitochondrial DNA in Western Europe. *Ann Hum Genet* 62, 241-260.
62. 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., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Renzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A., Bandelt, H-J. (2000). Tracing European founder lineages in the near Eastern mtDNA pool. *Am J Hum Genet*, 1251-1276.

63. Sajantila, A., Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M.L., Aula, P., Beckman, L., Tranebjaerg, L., Gedde-Dahl, T., Issel-Tarver, L., Di Rienzo, A., Pääbo, S. (1995). Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res* 5, 42-5.
64. Sammallahti P (1998) The Saami languages: an introduction. Davvi Girji, Kárášjohka/Karasjoki, Vaasa.
65. Sampietro, M., Lao, O., Caramelli, D., Lari, M., Pou, R., Martí, M., Bertranpetit, J., Lalueza-Fox, C. (2007). Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc Biol Sci* 274, 2161-2167.
66. Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., Richards, M.B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84, 740-759.
67. Svendsen, J.I., Alexanderson, H., Astakhov, V.I., Demidov, I., Dowdeswell, J.A., Funder, S., Gataullin, V., Henriksen, M., Hjort, C., Houmark-Nielsen, M., Hubberten, H.W., Ingolfsson, O., Jakobsson, M., Kjaer, K.H., Larsen, E., Lokrantz, H., Lunkka, J.P., Lysa, A., Mangerud, J., Matiouchkov, A., Murray, A., Møller, P., Niessen, F., Nikolskaya, O., Polyak, L., Saarnisto, M., Siegert, C., Siegert, M.J., Spielhagen, R.F., Stein, R. (2004). Late quaternary ice sheet history of Northern Eurasia. *Quat Sc Rev*, 1229-1271.
68. Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogväli, E.L., Tolk, H.V., Reidla, M., Metspalu, E., Pliss, L., Balanovsky, O., Pshenichnov, A., Balanovska, E., Gubina, M., Zhadanov, S., Osipova, L., Damba, L., Voevoda, M., Kutuev, I., Bermisheva, M., Khusnutdinova, E., Gusar, V., Grechanina, E., Parik, J., Pennarun, E., Richard, C., Chaventre, A., Moisan, J.P., Barác, L., Perici, M., Rudan, P., Terzi, R., Mikerezi, I., Krumina, A., Baumanis, V., Koziel, S., Rickards, O., De Stefano, G.F., Anagnou, N., Pappa, K.I., Michalodimitrakakis, E., Ferák, V., Füredi, S., Komel, R., Beckman, L., Villems, R. (2004). The Western and Eastern roots of the Saami--the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *Am J Hum Genet* 74, 661-682.
69. von Cramon-Taubadel, N., Pinhasi, R. (2011). Craniometric data support a mosaic model of demic and cultural Neolithic diffusion to outlying regions of Europe. *Proc Biol Sci*.
70. Wallace, D.C., Brown, M.D., Lott, M.T. (1999). Mitochondrial DNA variation in human evolution and disease. *Gene* 238, 211-230.
71. Wood R. (2006). Chronometric and paleodietary studies at the Mesolithic and Neolithic burial ground of Minino, NW Russia. Dissertation for the MSc in archaeological Science. Oxford University.
72. Zvelebil M., Dolukhanov P. (1991). The transition to farming in Eastern and Northern Europe. *J W Prehist* 5 (3), 233-278.

SUPPLEMENTARY MATERIALS



Figure S1: Pictures of selected samples from Uznyi Oleni Ostrov, Popovo, Bolshoy Oleni Ostrov and Chalmny-Varre. All samples yielded consistent haplotypes except ACAD 4719.

A- Clone sequences for Uznyi Oleni Ostrov individual UZOO77

[illegible]

B- Clone sequences for Bolshoy Oleni Ostrov individual BO057-1

[illegible]

C-Clone sequences for Bolshoy Oleni Ostrov individual BO072-1

	16150	16160	16170	16180	16190	16200	16210	16220
crRS sequence	TTGGST	ACGACCCAAAGT	GTATTCATCCGACAT	CAACACGCGCTATGTAT	TCTGTCATTA	CTTGCAGCGCACAT	GAATATG	ATG
BO072-1: np 16045-16240								
BO072-1: clone 1								
BO072-1: clone 2								
BO072-1: clone 3								
BO072-1: clone 4								
BO072-1: clone 5								
BO072-1: clone 6								
BO072-1: clone 7								
BO072-1: clone 8								
BO072-1: clone 9								
BO072-1: clone 10								
BO072-1: clone 11								
BO072-1: clone 12								
BO072-1: clone 13								
BO072-1: clone 14								
BO072-1: clone 15								
BO072-1: clone 16								
BO072-1: clone 17								
BO072-1: clone 18								
BO072-1: clone 19								
BO072-1: clone 20								
BO072-1: clone 21								
BO072-1: clone 22								
BO072-1: clone 23								
BO072-1: clone 24								
BO072-1: clone 25								
BO072-1: clone 26								
BO072-1: clone 27								
BO072-1: clone 28								
BO072-1: clone 29								
BO072-1: clone 30								
BO072-1: clone 31								
BO072-1: clone 32								
BO072-1: clone 33								
BO072-1: clone 34								
BO072-1: clone 35								
BO072-1: clone 36								
BO072-1: clone 37								
BO072-1: clone 38								
BO072-1: clone 39								
BO072-1: clone 40								
BO072-1: clone 41								
BO072-1: clone 42								
BO072-1: clone 43								
BO072-1: clone 44								
BO072-1: clone 45								
BO072-1: clone 46								
BO072-1: clone 47								
BO072-1: clone 48								
BO072-1: clone 49								
BO072-1: clone 50								
BO072-1: clone 51								
BO072-1: clone 52								
BO072-1: clone 53								
BO072-1: clone 54								
BO072-1: clone 55								
BO072-1: clone 56								
BO072-1: clone 57								
BO072-1: clone 58								
BO072-1: clone 59								
BO072-1: clone 60								

	16130	16140	16150	16160	16170	16180	16190	16200	16210	16220
crRS sequence	TGACGTAT	CACACAT	CAACT	GGAACT	CGACACGCGCT	GACCGCAT	AGAT	ACCAACAACACCT	ATCCGACGCT	TACAGTACAT
BO072-1: np 16125-16340										
BO072-1: clone 1										
BO072-1: clone 2										
BO072-1: clone 3										
BO072-1: clone 4										
BO072-1: clone 5										
BO072-1: clone 6										
BO072-1: clone 7										
BO072-1: clone 8										
BO072-1: clone 9										
BO072-1: clone 10										
BO072-1: clone 11										
BO072-1: clone 12										
BO072-1: clone 13										
BO072-1: clone 14										
BO072-1: clone 15										
BO072-1: clone 16										
BO072-1: clone 17										
BO072-1: clone 18										
BO072-1: clone 19										
BO072-1: clone 20										
BO072-1: clone 21										
BO072-1: clone 22										
BO072-1: clone 23										
BO072-1: clone 24										
BO072-1: clone 25										
BO072-1: clone 26										
BO072-1: clone 27										
BO072-1: clone 28										
BO072-1: clone 29										
BO072-1: clone 30										
BO072-1: clone 31										
BO072-1: clone 32										
BO072-1: clone 33										
BO072-1: clone 34										
BO072-1: clone 35										
BO072-1: clone 36										
BO072-1: clone 37										
BO072-1: clone 38										
BO072-1: clone 39										
BO072-1: clone 40										
BO072-1: clone 41										
BO072-1: clone 42										
BO072-1: clone 43										
BO072-1: clone 44										
BO072-1: clone 45										
BO072-1: clone 46										
BO072-1: clone 47										
BO072-1: clone 48										
BO072-1: clone 49										
BO072-1: clone 50										
BO072-1: clone 51										
BO072-1: clone 52										
BO072-1: clone 53										
BO072-1: clone 54										
BO072-1: clone 55										
BO072-1: clone 56										
BO072-1: clone 57										
BO072-1: clone 58										
BO072-1: clone 59										
BO072-1: clone 60										

D Clone sequences for Bolshoy Oleni Ostrov individual BO072-7

	16150	16180	16210	16240	16270	16300	16330	16360	16390	16420	16450	16480	16510	16540	16570	16600	16630	16660	16690	16720	16750	16780	16810	16840	16870	16900	16930	16960	16990	17020	17050	17080	17110	17140	17170	17200	17230	17260	17290	17320	17350	17380	17410	17440	17470	17500	17530	17560	17590	17620	17650	17680	17710	17740	17770	17800	17830	17860	17890	17920	17950	17980	18010	18040	18070	18100	18130	18160	18190	18220	18250	18280	18310	18340	18370	18400	18430	18460	18490	18520	18550	18580	18610	18640	18670	18700	18730	18760	18790	18820	18850	18880	18910	18940	18970	19000	19030	19060	19090	19120	19150	19180	19210	19240	19270	19300	19330	19360	19390	19420	19450	19480	19510	19540	19570	19600	19630	19660	19690	19720	19750	19780	19810	19840	19870	19900	19930	19960	19990	20020	20050	20080	20110	20140	20170	20200	20230	20260	20290	20320	20350	20380	20410	20440	20470	20500	20530	20560	20590	20620	20650	20680	20710	20740	20770	20800	20830	20860	20890	20920	20950	20980	21010	21040	21070	21100	21130	21160	21190	21220	21250	21280	21310	21340	21370	21400	21430	21460	21490	21520	21550	21580	21610	21640	21670	21700	21730	21760	21790	21820	21850	21880	21910	21940	21970	22000	22030	22060	22090	22120	22150	22180	22210	22240	22270	22300	22330	22360	22390	22420	22450	22480	22510	22540	22570	22600	22630	22660	22690	22720	22750	22780	22810	22840	22870	22900	22930	22960	22990	23020	23050	23080	23110	23140	23170	23200	23230	23260	23290	23320	23350	23380	23410	23440	23470	23500	23530	23560	23590	23620	23650	23680	23710	23740	23770	23800	23830	23860	23890	23920	23950	23980	24010	24040	24070	24100	24130	24160	24190	24220	24250	24280	24310	24340	24370	24400	24430	24460	24490	24520	24550	24580	24610	24640	24670	24700	24730	24760	24790	24820	24850	24880	24910	24940	24970	25000	25030	25060	25090	25120	25150	25180	25210	25240	25270	25300	25330	25360	25390	25420	25450	25480	25510	25540	25570	25600	25630	25660	25690	25720	25750	25780	25810	25840	25870	25900	25930	25960	25990	26020	26050	26080	26110	26140	26170	26200	26230	26260	26290	26320	26350	26380	26410	26440	26470	26500	26530	26560	26590	26620	26650	26680	26710	26740	26770	26800	26830	26860	26890	26920	26950	26980	27010	27040	27070	27100	27130	27160	27190	27220	27250	27280	27310	27340	27370	27400	27430	27460	27490	27520	27550	27580	27610	27640	27670	27700	27730	27760	27790	27820	27850	27880	27910	27940	27970	28000	28030	28060	28090	28120	28150	28180	28210	28240	28270	28300	28330	28360
--	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

E Clone sequences for Bolshoy Oleni Ostrov individual BOO72-4

[illegible]

F Clone sequences for Bolshoy Oleni Ostrov individual BOO72-15

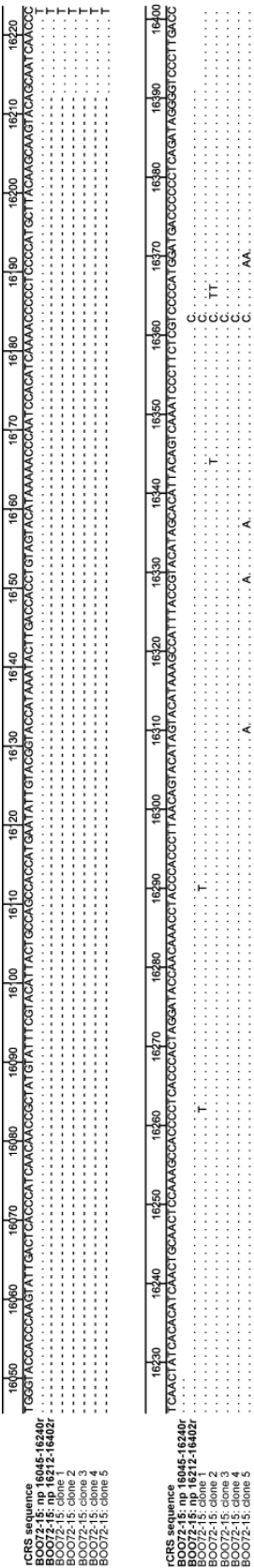


Figure S2A-F: Clone sequences.

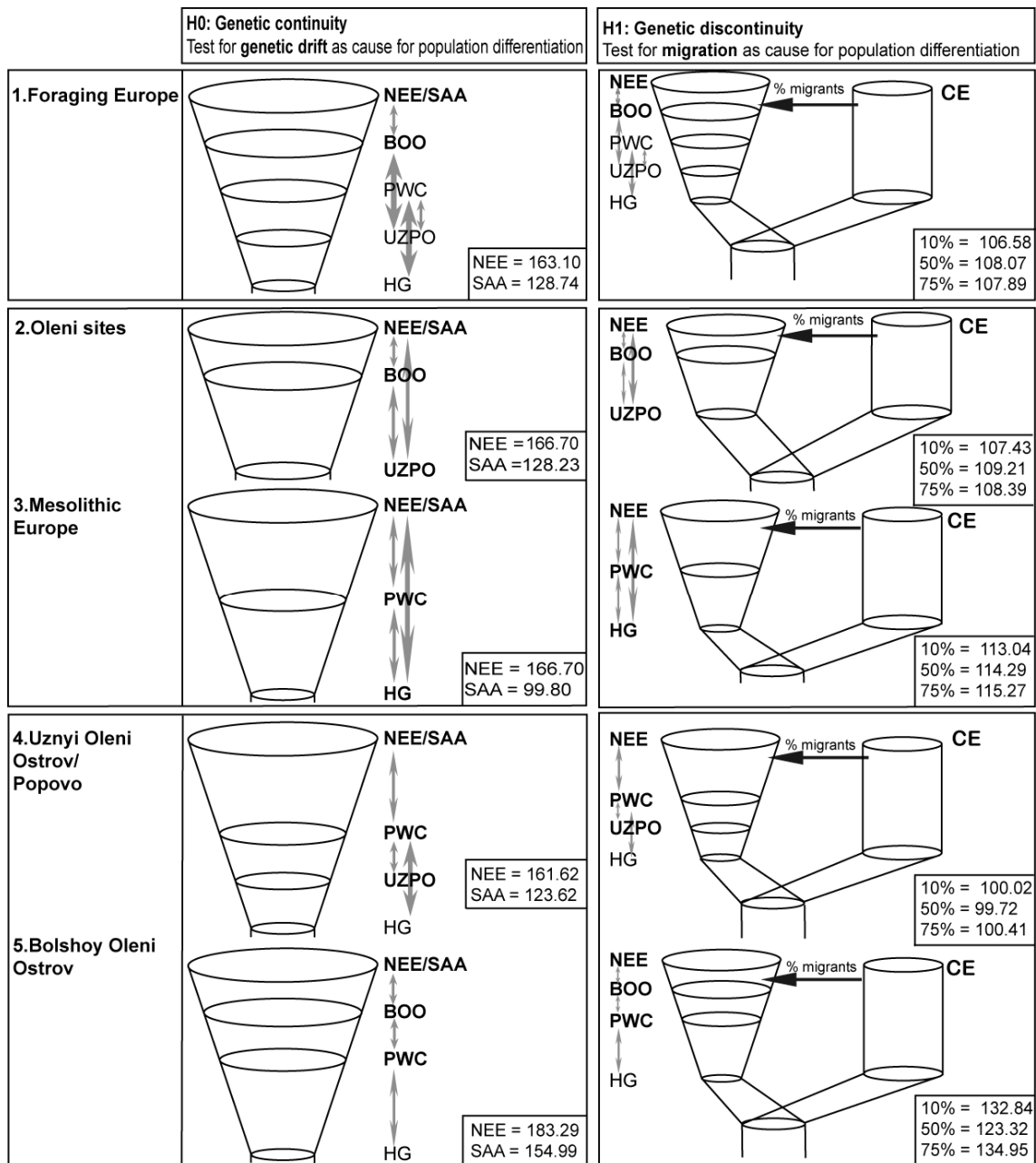


Figure S3: Additional models tested through coalescent simulation (BayeSSC) with associated Akaike Information Criterion (AIC).

[illegible]

REFERENCES:

- REFERENCES
1. Haak, W., Balanovsky, O., Sanchez, J., Koshel, S., Zaporozhchenko, V., Adler, C., Der Sarkissian, C., Brandt, G., Sotgiu, G., et al. (2008) Ancient DNA from Eastern Europe reveals a major prehistoric population shift. *Nature* 449: 91–95.
2. Handt, O., Krings, M., Ward, R., and Pääbo, S. (1996) The retrieval of ancient human DNA sequences: American Indians as a model system. *Proc. Natl. Acad. Sci. USA* 93: 2591–2596.
3. Stone, A., and Stoneking, M. (1998) mtDNA analysis of a prehistoric Onondaga population: Implications for the peopling of North America. *Am. J. Hum. Genet.* 63: 1561–1574.
4. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., and Wilson, A. (1991) African populations and the evolution of modern humans. *Science* 253: 1503–1507.

Table S2. Results of SNP typing in the mtDNA coding region using the GenoCoRe22 SNaPshot assay. SNPs typed on the L-strand are reported in capital letters in the reference rCRS profile, whereas SNPs typed on the H-strand are reported in lower-case. Missing data signifies allelic dropout or fluorescence signal below the background threshold (100 relative fluorescent units, rfu). 'g/a' indicates the presence of a mixed signal for the position interrogated. A mixed signal was repeatedly obtained at position 8994 (haplogroup W) with the detection of an additional G base. However, the rest of the profile never could phylogenetically support the presence of the G base at this particular position. For each individual, profiles were obtained from two independent extracts, except for individuals BOO72-9, ChV30, ChV38, for which a second samples was not available and for UZOO-77, BOO57-1, BOO72-10, BOO72-4, BOO72-7, BOO72-15, BOO72-1, ChV31, ChV43, ChV45, ChV24, ChV35, ChV15, ChV18, ChV22, ChV1, ChV14, ChV17, ChV33, ChV46, ChV7, for which the second individual was extracted in an independent laboratory.

		13928_R9	3594_L3'4	10550_K	11467_U	4248_A	8994_W	13263_C	13368_T	11719_preHV_R0	8280delB	4580_V	6371_X	10238_N1	7028_H	10034_I	5178_D	14766_HV	12705_R	10873_N	12612_J	2758_L2'6	10400_M	Hg ¹
rCRS		c	G	t	A	T	G	t	G	c	A	c	g	T	g	T	g	G	g	a	a	G	g	H
Sample/Extract																								
Uznyi Oleni Ostrov: UZ00-																								
43	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
46	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	g/a	T	G	A	G		A	G	G	U
16	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	g/a	T	G	A	G	A	A	G	G	U
40	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
70	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	g/a	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
77	1	C	G	T	A	T	G	T	G	C	A	C	G	T	G	T	G	G	G	A	A	G	G	H
7	1	C		T	A	T	G	C	G	T	A	C	G	T	A	T	G			G	A	G	A	C
	2	C	G	T	A	T	g/a	C	G	T	A	C	G	T	A	T	G		A	G	A	G	A	C
8	1	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G	A	A	G	A	G	A	C
	2	C	G	T		T	g/a	C	G	T	A		G	T	g/a	T	G			G	A	G		C
74	1	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G		A	G	A	G	A	C
	2	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G		A	G	A	G	A	C
Popovo																								
Po4	1	C	G	T	G	T	g/a		G	T	A		G	T	A	T	G		G		A	G		U
	2	C	G	T	G	T	g/a	T	G	T	A		G	T	A	T	G				A	G	G	U
Po2	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	g/a	T	G	T	A		G	T	A	T	G		G	A	A	G	G	U

Table S2 (Continued 1). Results of SNP typing in the mtDNA coding region using the GenoCoRe22 SNaPshot assay.

		13928_R9	3594_L3'4	10550_K	11467_U	4248_A	8994_W	13263_C	13368_T	11719_preHV, R0	8280delB	4580_V	5371_X	10238_N1	7028_H	10034_I	5178_D	14766_HV	12705_R	10873_N	12612_J	2758_L2'6	10400_M	Hg ¹
rCRS		c	G	t	A	T	G	t	G	c	A	c	g	T	g	T	g	G	g	a	a	G	g	H
Sample/Extract																								
Bolshoy Oleni Ostrov: BOO																								
49-3	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
57-1	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
49-1	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
72-11	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
72-9	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
72-10	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
72-14	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
72-8	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
72-4	1	C	G	T	A	T	G	T	A	T	A	C	G	T	A	T	G	A	G	A	A	G	G	T
49-2	1	C	G	T	A	T	G	C	G	T	A		G	T	A	T	G			G	A	G		C
	2	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G			G		G		C
49-4	1	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G	A	A	G	A	G	A	C
	2	C	G	T	A	T	G	C	G	T	A		G	T	A	T	G			G		G	A	C
57-3	1	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G		A	G	A	G	A	C
	2	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G	A	A	G	A	G	A	C
72-2	1	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G	A	A	G	A	G	A	C
	2	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G	A	A	G	A	G	A	C
72-7	1	C	G	T	A	T	G	C	G	T	A		G		A	T	G			G	A	G	A	C
72-12	1	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G		A	G	A	G	A	C
	2	C	G	T	A	T	G	C	G	T	A		G	T	A	T	G		A		A	G	A	C
72-5	1	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G		A	G	A	G	A	C
	2	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G	A	A	G	A	G	A	C
72-6	1	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G	A	A	G	A	G	A	C
	2	C	G	T	A	T	G	C	G	T	A	C	G	T	A	T	G		A	G	A	G	A	C
49-6	1	C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	T	A	A	G	A	G	A	D
	2		G	T	A	T	G	T	G	T	A	C	G	T	A	T	T	A		G	A	G		D
72-13	1	C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	T	A	A	G	A	G	A	D
	2	C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	T	A	A	G	A	G	A	D
72-15	1	C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	T	A	A	G	A	G	A	D

Table S2 (Continued 2). Results of SNP typing in the mtDNA coding region using the GenoCoRe22 SNaPshot assay.

		13928_R9	3594_L3'4	10550_K	11467_U	4248_A	8994_W	13263_C	13368_T	11719_preHV, R0	8280delB	4580_V	6371_X	10238_N1	7028_H	10034_I	5178_D	14766_HV	12705_R	10873_N	12612_J	2758_L2'6	10400_M	Hg ¹
rCRS		c	G	t	A	T	G	t	G	c	A	c	g	T	g	T	g	G	g	a	a	G	g	H
Sample/Extract																								
Bolshoy Oleni Ostrov: BOO																								
49-5	1		G	T	A	T	G	T	G	T	A	C	G	T	A	T	G		A	G	A	G	A	M
	2	C	G	T	A	T	G	T	G	T	A		G	T	A	T	G	A	A	G	A	G	A	M
72-3	1	C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	G			G	A	G	A	M
	2	C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	G	A	A	G	A	G	A	M
72-1	1	C	G	T	A	T	G	T	G	T	A		G	T	A	T	G			G	A	G	A	M

		13928_R9	3594_L3'4	10550_K	11467_U	4248_A	8994_W	13263_C	13368_T	11719_preHV, R0	8280delB	4580_V	6371_X	10238_N1	7028_H	10034_I	5178_D	14766_HV	12705_R	10873_N	12612_J	2758_L2'6	10400_M	Hg ¹
rCRS		c	G	t	A	T	G	t	G	c	A	c	g	T	g	T	g	G	g	a	a	G	g	H
Sample/Extract																								
Chalmny-Varre: ChV																								
31	1	C	G	T	G	T			G	T	A	C	G	T	A	T	G		G	A	A	G	G	U
43	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
45	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
6	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
24	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
25	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	g/a	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
26	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
35	1		G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
47	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
15	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
18	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
30	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
40	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
22	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U

Table S2 (Continued 3). Results of SNP typing in the mtDNA coding region using the GenoCoRe22 SNaPshot assay.

		13928_R9	3594_L3'4	10550_K	11467_U	4248_A	8994_W	13263_C	13368_T	11719_preHV, R0	8280delB	4580_V	6371_X	10238_N1	7028_H	10034_I	5178_D	14766_HV	12705_R	10873_N	12612_J	2758_L2'6	10400_M	Hg ¹
rCRS		c	G	t	A	T	G	t	G	c	A	c	g	T	g	T	g	G	g	a	a	G	g	H
Sample/Extract																								
Chalmny-Varre: ChV																								
44	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	G	T	G	T	A		G	T	A	T	G	A	G		A			U
1	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
2	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
	2	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
3	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
	2	C		T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G		V
4	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
	2	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
5	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
	2	C	G	T	A	T	G	T	G	C	A	T	G	T	A		G	G	G	A	A	G	G	V
8	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
	2	C	G	T	A	T	G	T	G	C		T	G	T	A	T					A			V
9	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
	2	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G		G	A	A	G	G	V
10	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
	2	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
11	1	C	G	T	A	T	g/a	T	G	C	A	T	G	T	A	T	G		G	A	A	G	G	V
	2	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
13	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
	2	C	G	T	A	T	g/a	T	G	C	A	T	G	T	A	T	G	g/a	G	A	A	G	G	V
14	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
16	1	C	G	T	A	T	G		G	C	A	T	G			T			G		A	G		V
	2	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
17	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
21	1	C	G	T	A	T	G	T	G	C	A		G	T	A	T	G	G	G	A	A	G		V
	2	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
27	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
	2	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
28	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
	2	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V
33	1	C	G	T	A	T	G	T	G	C	A	T	G	T	A	T	G	G	G	A	A	G	G	V

Table S2 (Continued 4). Results of SNP typing in the mtDNA coding region using the GenoCoRe22 SNaPshot assay.

|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|

¹ Hg: Haplogroup

Table S3: Results of quantitative PCR

PCR:L16209/H16303						
Site	Sample	Cycle Threshold (Ct)	Average Ct	Stdev¹	Stdev/AverageCt %	Copies/uL
Uznyi Oleni Ostrov	UZOO43	38.87	38.94	0.08	0.20	1,158
Uznyi Oleni Ostrov	UZOO43	39.02				
Uznyi Oleni Ostrov	UZOO43	38.92				
Uznyi Oleni Ostrov	UZOO74	42.32	42.03	0.51	1.22	160
Uznyi Oleni Ostrov	UZOO74	41.44				
Uznyi Oleni Ostrov	UZOO74	42.33				
Bolshoy Oleni Ostrov	BOO72-1	33.88	33.95	0.18	0.53	28,117
Bolshoy Oleni Ostrov	BOO72-1	34.16				
Bolshoy Oleni Ostrov	BOO72-1	33.82				
Bolshoy Oleni Ostrov	BOO79-9	37.38	36.65	2.12	5.79	5,004
Bolshoy Oleni Ostrov	BOO79-9	38.31				
Bolshoy Oleni Ostrov	BOO79-9	34.26				
Chalmny-Varre	ChV22	33.70	34.05	0.46	1.35	26,430
Chalmny-Varre	ChV22	33.88				
Chalmny-Varre	ChV22	34.57				
Chalmny-Varre	ChV23	33.79	33.51	0.25	0.74	37,424
Chalmny-Varre	ChV23	33.34				
Chalmny-Varre	ChV23	33.39				
Chalmny-Varre	ChV44	34.43	34.56	0.51	1.47	19,069
Chalmny-Varre	ChV44	35.12				
Chalmny-Varre	ChV44	34.13				

PCR:L16209/H16348						
Site	Sample	Cycle Threshold (Ct)	Average Ct	Stdev¹	Stdev/AverageCt %	Copies/uL
Uznyi Oleni Ostrov	UZOO43	34.54	34.69	0.15	0.43	6
Uznyi Oleni Ostrov	UZOO43	34.68				
Uznyi Oleni Ostrov	UZOO43	34.84				
Uznyi Oleni Ostrov	UZOO74	35.30	35.03	0.24	0.68	5
Uznyi Oleni Ostrov	UZOO74	34.84				
Uznyi Oleni Ostrov	UZOO74	34.96				
Bolshoy Oleni Ostrov	BOO72-1	31.11	30.89	0.64	2.06	70
Bolshoy Oleni Ostrov	BOO72-1	30.17				
Bolshoy Oleni Ostrov	BOO72-1	31.38				
Bolshoy Oleni Ostrov	BOO79-9	30.42	30.44	0.05	0.17	88
Bolshoy Oleni Ostrov	BOO79-9	30.50				
Bolshoy Oleni Ostrov	BOO79-9	30.40				
Chalmny-Varre	ChV22	31.76	32.11	0.62	1.92	31
Chalmny-Varre	ChV22	31.74				
Chalmny-Varre	ChV22	32.82				
Chalmny-Varre	ChV23	30.05	30.20	0.14	0.47	103
Chalmny-Varre	ChV23	30.33				
Chalmny-Varre	ChV23	30.21				
Chalmny-Varre	ChV44	31.86	31.76	0.57	1.78	39
Chalmny-Varre	ChV44	31.15				
Chalmny-Varre	ChV44	32.27				

¹Stdev: Standard deviation

Table S4: Description and references for the comparative population datasets and pools.

Population name	Population ID	Pool	Latitude	Longitude	N	Reference
Albanians	ALB	Europe	41.4	19.8	281	Belledi et al., 2000, Bosch et al., 2006, Balanovsky, personal communication
Aleuts	ale				199	Volodko et al., 2008
Altaians Telenghits	alt2		50.0	88.0	71	Derenko et al., 2007
Altaians Tubalars	alt3		52.0	87.0	72	Derenko et al., 2007
Altaians-Kizhi	alt1		50.5	86.0	90	Starikovskaya et al., 2005
Armenians	ARM		40.2	44.5	192	Metspalu et al., GenBank
Aromuns	aro		41.4	21.3	133	Bosch et al., 2006
Austrians	AUT	Europe	47.4	11.4	117	Parson et al., 1998, Handt et al., 1994
Azeris	AZE		40.0	48.0	88	Quintana-Murci et al., 2004, Richards et al., 2000
Bashkirs	RU-BA	Western/Central Siberia	54.8	56.0	207	Bermisheva et al., 2002
Bosnians	BIH	Europe	43.9	18.4	322	Malyarchuk et al., 2003, Harvey Unpublished
Bulgarians	BGR	Europe	42.9	26.1	141	Calafell et al., 1996, Richards et al., 2000
Buryats	RU-BU		53.0	110.0	411	Derenko et al., 2002, Derenko et al., 2007; Starikovskaya et al., 2005
Byelorussians Bobruisk	BEL	Europe	53.1	29.2	92	Belyaeva et al., 2000
Byelorussians Brest		Europe	52.0	26.8	89	Balanovsky, personal communication
Byelorussians Gomel		Europe	52.6	29.7	71	Balanovsky, personal communication
Byelorussians Vitebsk		Europe	55.0	29.0	100	Balanovsky, personal communication
Chukchi	RU-CHU		65.0	175.0	262	Volodko et al., 2008, Starikovskaya et al., 1998, Derenko et al., 2007
Chuvash	RU-CU	Europe	56.0	46.8	92	Bermisheva et al., 2002; Richards et al., 2000
Croatians coastal	HRV	Europe	43.3	17.0	96	Babalini et al., 2005
Croatians isles 1		Europe	43.3	16.6	311	Babalini et al., 2005
Croatians isles 2		Europe	45.2	14.4	133	Balanovsky, personal communication
Croatians northern		Europe	45.6	18.7	294	Balanovsky, personal communication
Croatians southern		Europe	42.8	18.0	146	Balanovsky, personal communication
Cyprus	CYP		35.2	33.3	91	Irwin et al., 2007
Czech	CZE	Europe	50.1	14.2	449	Balanovsky, personal communication, Richards et al., 2000, Vanecek et al., 2003, Malyarchuk et al., 2006
English Central	GBR	Europe	53.0	358.0	271	Sykes et al., 2006
English Cornwall		Europe	50.5	355.8	92	Richards et al., 2000
English eastern		Europe	53.0	1.0	339	Sykes et al., 2006
English northern		Europe	56.0	357.0	403	Sykes et al., 2006
Scottish		Europe	58.0	355.0	891	Helgason et al., 2001

Table S4 (Continued 1): Description and references for the comparative population datasets and pools.

Population name	Population ID	Pool	Latitude	Longitude	N	Reference
Wales		Europe	52.5	356.0	192	Piercy et al., 1993, Richards et al., 1996
Western isles and Skye		Europe	57.5	354.0	230	Helgason et al., 2001
Eskimo	esk				825	Volodko et al., 2008, Saillard et al., 2000, Starikovskaya et al., 1998, Helgason et al., 2005, Simonson et al., Genbank
Estonians	EST	Europe	58.4	26.7	662	Balanovsky, personal communication, Sajantila et al, et al., 1995, Sajantila et al., 1996; Pult et al., 1994; Lappalainen et al., 2008
Evenks East	eve		49.0	119.0	92	Derenko, et al., 2007
Evenks western 1			62.0	96.0	142	Starikovskaya et al., 2005, Pakendorf et al., 2006
Evenks western 2			60.0	94.0	73	Derenko et al., 2007
Evens	evn		60.0	140.0	100	Derenko et al., 2002, Tajima et al., 2004
Finns northern	FIN	Europe	64.4	27.5	403	Meinila et al., 2001
Finns southern		Europe	62.0	26.0	105	Kittles et al., 1999, Lahermo et al., 1996
France Finistere	FRA	Europe	48.0	355.9	142	Richard et al., 2007, Dubut et al., 2003
France Perigord-Limousin		Europe	46.1	1.0	72	Dubut et al., 2003
French Bearn, Pyrennees-Atlantiques		Europe	43.3	359.5	81	Richard et al., 2007
French Brittany, Loire-Atlantique		Europe	47.5	358.3	75	Richard et al., 2007
French central		Europe	47.3	0.1	135	Richard et al., 2007, Dubut et al., 2003
French Languedoc, Herault		Europe	43.6	3.9	85	Richard et al., 2007
French Morbihan		Europe	47.7	357.0	81	Richard et al., 2007, Dubut et al., 2003
French Normandy		Europe	49.3	0.2	85	Richard et al., 2007, Dubut et al., 2003
French Picardie, Somme		Europe	49.9	2.2	79	Richard et al., 2007
French Poitou, Vendee		Europe	46.9	358.5	80	Richard et al., 2007
French south-eastern		Europe	44.6	5.7	83	Richard et al., 2007, Dubut et al., 2003
Georgians	GEO	Europe	42.0	44.0	158	Quintana-Murci et al., 2004, Balanovsky, personal communication
Germans Lower Saxony	DEU	Europe	52.6	9.6	700	Pfeiffer et al., 1999
Germans south	GRC	Europe	48.3	11.5	247	Lutz et al., 1998, Richards et al., 1996
Germans west		Europe	51.7	7.3	159	Pfeiffer et al., 1999, Baasner et al., 1998
Germans Western Pomerania		Europe	54.0	13.0	300	Poetsch et al., 2003
Greek Crete		Europe	35.0	25.0	187	Balanovsky, personal communication

Table S4 (Continued 2): Description and references for the comparative population datasets and pools.

Population name	Population ID	Pool	Latitude	Longitude	N	Reference
Greek northern		Europe	40.5	22.9	469	Richards et al., 2000, Bosch et al., 2006, Irwin et al., 2007
Hungarians	HUN	Europe	47.5	19.0	190	Balanovsky, personal communication, Bogaszi-Szabo et al., 2006
Icelanders	ISL	Europe	64.5	337.4	448	Richards et al., 1996, Sajantila et al., 1995, Helgason et al., 2000
Iranians central	IRN		32.7	51.7	78	Metspalu et al., 2004, Quintana-Murci et al., 2004
Iranians northwest			37.3	49.6	284	Metspalu et al., 2004, Quintana-Murci et al., 2004
Iranians southwest			33.5	48.3	155	Metspalu et al., 2004, Quintana-Murci et al., 2004
Iraq	IRQ		33.3	44.4	168	Richards et al., 2000, Al-Zahery et al., 2003
Irish	IRL	Europe	53.5	350.9	300	McEvoy et al., 2004
Italians central	ITA	Europe	41.9	12.5	183	Babalini et al., 2005, Richards et al., 2000, Tagliabracci et al., 2001
Italians eastern		Europe	41.9	14.3	73	Babalini et al., 2005
Italians southern		Europe	41.1	15.5	74	Babalini et al., 2005
Italians Tuskany		Europe	43.8	11.2	432	Achilli et al., 2007, Francalacci et al., 1996, Varesi et al., GenBank
Italians Veneto		Europe	46.0	11.0	68	Mogentale-Profizi et al., 2001
Sicily		Europe	38.0	12.9	106	Cali et al., 2001
Jordanians	JOR		31.8	35.8	146	Cabrera et al., GenBank
Kabardians	kab		43.0	43.0	163	Richards et al., 2000, Balanovsky, personal communication
Karelians Aunus	RU-KR	Europe	62.0	32.0	218	Lappalainen et al., 2008
Karelians Viena		Europe	66.0	32.0	87	Lappalainen et al., 2008
Kazakhs	KAZ		45.0	80.0	125	Comas et al., 1998, Comas et al., 2004; Yao et al., 2000
Kets	ket		45.8	88.0	104	Derbeneva et al., 2002, Balanovsky, personal communication
Khakassians	RU-KK		53.0	90.0	110	Derenko et al., 2002, Derenko et al., 2007; Starikovskaya et al., 2005
Khamnigans	kham		53.0	115.0	99	Derenko et al., 2007
Khants	RU-KHM_khan	Western/Central Siberia	62.0	72.0	318	Pimenoff et al., 2008, Balanovsky, personal communication
Komi	RU-KO	Europe	61.0	53.0	127	Bermisheva et al., 2002
Koryaks	kor		55.0	160.0	147	Schurr et al., 1999
Kurds	kur		37.6	43.1	73	Richards et al., 2000, Quintana-Murci et al., 2004

Table S4 (Continued 3): Description and references for the comparative population datasets and pools.

Population name	Population ID	Pool	Latitude	Longitude	N	Reference
Latvians	LVA	Europe	57.0	24.0	413	Pliss et al., 2005, Lappalainen et al., 2008
Lithuanians Aukstaiciai	LTU	Europe	55.0	24.0	90	Balanovsky, personal communication
Lithuanians Zemaiciai		Europe	55.5	22.0	90	Balanovsky, personal communication
Mansi	RU-KHM_man	Western/Central Siberia	60.0	66.0	161	Derbeneva et al., 2002, Pimenoff et al., 2008
Mari	RU-ME	Europe	56.0	48.1	136	Bermisheva et al., 2002
Mongolians	MNG		45.0	105.0	262	Yao et al., 2004; Kolman et al., 1996; Derenko et al., 2007; Kong et al., 2003
Mordvinians	RU-MO	Europe	54.3	44.5	99	Bermisheva et al., 2002
Morocco	MAR		31.0	353.1	336	Rando et al., 1998, Balanovsky, personal communication
Nenets Asian	RU-NEN_A	Western/Central Siberia	65.0	70.0	79	Balanovsky, personal communication
Nenets European	RU-NEN_E	Western/Central Siberia	69.0	49.0	128	Saillard et al., 2000, Tonks et al., 2006
Nganasan	nga	Western/Central Siberia	69.5	86.2	118	Volodko et al., 2008, Derbeneva et al., 2002, Osipova et al., 2005
Nivkhs	niv		52.0	142.0	113	Starikovskaya et al., 2005, Tajima et al., 2004
Nogays	nog		44.0	47.0	206	Bermisheva et al., 2004
Norwegians	NOR	Europe	59.9	10.6	663	Helgason et al., 2001, Passarino et al., 2002, Richards et al., 2000, Dupuy et al., 1996, Opdal et al., 1998
Ossets northern	RU-SE		43.0	44.5	106	Richards et al., 2000
Ossets southern			42.3	44.0	183	Balanovsky, personal communication
Palestinians	PSE		31.8	35.1	117	Richards et al., 2000
Poles	POL	Europe	52.0	21.0	583	Balanovsky, personal communication, Richards et al., 2000, Malyarchuk et al., 2002
Portugal central	PRT	Europe	39.5	352.0	317	Gonzalez et al., 2003, Pereira et al., 2004
Portugal northern		Europe	41.3	351.5	271	Gonzalez et al., 2003, Pereira et al., 2004
Portugal southern		Europe	37.2	352.2	260	Gonzalez et al., 2003, Pereira et al., 2004
Romanians 1	ROU	Europe	47.6	23.6	92	Richards et al., 2000
Romanians south		Europe	44.1	28.6	105	Bosch et al., 2006
Russians Belgorod	RUS	Europe	50.8	36.5	148	Balanovsky, personal communication
Russians Cossacks		Europe	44.5	40.2	132	Balanovsky, personal communication
Russians Oshevensk		Europe	61.8	38.8	76	Belyaeva et al., 2000
Russians Pinega		Europe	63.4	46.5	144	Balanovsky, personal communication

Table S4 (Continued 4): Description and references for the comparative population datasets and pools.

Population name	Population ID	Pool	Latitude	Longitude	N	Reference
Russians Pomors		Europe	66.0	42.0	81	Tonks et al., 2006
Russians Rostov		Europe	47.2	39.7	111	Kornienko et al., 2004, Richards et al., 2000
Russians Smolensk		Europe	53.9	32.9	147	Balanovsky, personal communication
Russians Unja		Europe	58.3	44.8	79	Balanovsky, personal communication
Saami	saa		68.9	27.6	559	Sajantila et al., 1995, Dupuy et al., 1996, Kittles et al., 1999, Delghandi et al., 1998, Tambets et al., 2004, Tonks et al., 2006
Sardinia	IT-88	Europe	40.0	9.0	115	Richards et al., 2000
Saudi Arabia	SAU		24.6	46.5	325	Balanovsky, personal communication, Abu-Amero et al., 2007
Selkups	sel	Western/Central Siberia	65.8	82.5	120	Balanovsky, personal communication
Shors	sho		52.8	87.9	82	Derenko et al., 2007
Slovaks	SVK	Europe	48.8	19.3	510	Koledova et al., 2005 Unpublished
Slovenians	SVN	Europe	46.1	14.5	233	Malyarchuk et al., 2003, Zupanec et al., 2004
Spaniards Andalusia	ESP	Europe	38.1	355.2	65	Larruga et al., 2001, Corte-Real et al., 1996
Spaniards Cantabria		Europe	43.2	356.0	242	Maca-Meyer et al., 2003a
Spaniards central		Europe	40.6	356.2	129	Larruga et al., 2001, Corte-Real et al., 1996
Spaniards Galicia		Europe	43.0	352.0	135	Gonzalez et al., 2003, Salas et al., 1998
Spaniards north-eastern		Europe	41.6	1.9	133	Crespillo et al., 2000, Corte-Real et al., 1996
Swedes central	SWE	Europe	59.3	17.7	105	Balanovsky, personal communication
Swedes Gotland		Europe	57.7	18.1	267	Balanovsky, personal communication
Swedes northern		Europe	68.0	20.0	97	Balanovsky, personal communication
Swedes southern		Europe	55.9	12.7	177	Balanovsky, personal communication
Swiss	CHE	Europe	46.7	6.6	230	Dimo-Simonin et al., 2000, Pult et al., 1994
Syrians	SYR		33.6	36.2	169	Balanovsky, personal communication, Richards et al., 2000
Tatars	RU-TA	Europe	65.0	52.4	225	Bermisheva et al., 2002
Tofalars	tof		54.8	99.0	104	Derenko et al., 2002, Starikovskaya et al., 2005
Turkey	TUR		39.0	33.0	608	Balanovsky, personal communication, Quintana-Murci et al., 2004, Richards et al., 2000, Calafell et al., 1996

Table S4 (Continued 5): Description and references for the comparative population datasets and pools.

Population name	Population ID	Pool	Latitude	Longitude	N	Reference
Tuvinians	tuv		51.6	94.4	645	Derenko et al., 2002, Derenko et al., 2007, Balanovsky, personal communication, Tonks et al., 2006, Starikovskaya et al., 2005, Pakendorf et al., 2006
Udmurts	RU-UD	Europe	56.6	53.0	109	Bermisheva et al., 2002
Ukrainians Belgorod	UKR	Europe	50.4	35.8	95	Balanovsky, personal communication
Ukrainians Cherkasy		Europe	49.4	32.1	179	Balanovsky, personal communication
Ukrainians Hmel'nitskaya		Europe	49.7	27.3	179	Balanovsky, personal communication
Ukrainians western		Europe	49.3	24.0	157	Balanovsky, personal communication
Ulchi-Udegey-Negidal	ulc		50.0	135.0	166	Starikovskaya et al., 2005
Yakuts	RU-SA		65.0	125.0	770	Pakendorf et al., 2003, et al., 2006; Balanovsky, personal communication; Fedorova et al., 2003; Zlojutro et al., 2006; Derenko et al., 2002, Derenko et al., 2007
Yukagir	yuk		65.0	150.0	153	Pakendorf et al., 2006, Volodko et al., 2008

Table S5: Population parameters used in BayeSSC coalescent simulation analyses. Sample sizes and ages of the sequences were implemented in the demographic models used for genealogy simulations. Haplotype diversity and fixation indexes were used in the analyses of the simulation for parameter estimation and model selection.

	North East Europeans	Saami	Bolshoy Oleni Ostrov	Pitted Ware Culture hunter-gatherers	Uznyi Oleni Ostrov/ Popovo	European hunter-gatherers	Central Europeans
Abbreviation	NEE	SAA	aBOO	aPWC	aUzPo	aHG	CE
Sample size	621	118	23	19	11	20	1030
Age (generations)	0	0	140	116	300	168 - 614	0
Haplotype diversity	0.98	0.81	0.82	0.80	0.74	0.91	0.97

Fixation indeces Fst						
	NEE	SAA	aBOO	aPWC	aUzPo	aHG
NEE	-					
SAA	0.12	-				
aBOO	0.12	0.17	-			
aPWC	0.05	0.20	0.17	-		
aUzPo	0.05	0.14	0.06	0.06	-	
aHG	0.08	0.19	0.17	0.09	0.15	-

In Chapter One, I identified a mitochondrial haplotype whose origins were difficult to establish. This haplotype, found in a Mesolithic population of reindeer hunters of north east Europe, belongs to haplogroup C1, which is rare in Europe today.

In this chapter, I present the complete mitochondrial genome sequence of this C1 lineage, and from the improved phylogeny of haplogroup C1, I propose the most likely models for the presence of haplogroup C1 in Europe.

Chapter Two

Mitochondrial Genome Sequencing in Mesolithic North East Europe Unearths a New Sub-clade Within the Broadly Distributed Human Haplogroup C1

Abbreviations: ACAD, Australian Centre for Ancient DNA; A.D., Anno Domini; bp, base pair; aDNA, ancient DNA; CEL, Cell Intensity; dNTP, deoxynucleoside triphosphate; Exo, exonuclease; Hg, haplogroup; HVR-I and II, hypervariable region I and II; min, minute; mtDNA, mitochondrial DNA; mtgenome, mitochondrial genome; np, nucleotide position; PCR, Polymerase Chain Reaction; QST, quality score threshold; rCRS, revised Cambridge reference sequence; RPM, revolutions per minute; RSA, Rabbit Serum Albumin; SAP, shrimp alkaline phosphatase; SNP, single nucleotide polymorphism; UV, ultraviolet; yBP, years Before Present.

ABSTRACT

The human mitochondrial haplogroup C1 has a broad global distribution but is extremely rare in Europe today. Its presence in Europe in the Mesolithic has been demonstrated by ancient DNA (Chapter One). Three individuals from the 7,500 year old Mesolithic site of Uznyi Oleni Ostrov, Western Russia, could be assigned to haplogroup C1 based on mtDNA hypervariable region I (HVR-I) sequences. However, HVR-I data alone did not provide enough resolution to establish the phylogenetic relationship of these Mesolithic haplotypes with currently known haplogroup C1 mitochondrial DNA (mtDNA) sequences found today in populations of Europe, Asia and the Americas. In order to shed light on the origin of this European Mesolithic C1 haplotype, we sequenced the mitochondrial genome (mtgenome) of one Uznyi Oleni Ostrov C1 carrier. The mtgenome sequence was obtained using an Affymetrix[®] GeneChip[®] Human Mitochondrial resequencing Array 2.0 (MitoChip v.2.0) from a library enriched for the target ancient human mtDNA. The phylogenetic analysis of the C1 haplogroup indicated that the Uznyi Oleni Ostrov haplotype represents a new distinct clade, coined here 'C1f'. We show that all three C1 carriers of Uznyi Oleni Ostrov belong to this clade. No haplotype closely-related to the Uznyi Oleni Ostrov C1f sequence could be found in the current database of ancient and present-day mtgenomes. Hence, we have discovered past human mitochondrial diversity that has not been observed in modern populations so far. The lack of positive matches in modern populations may be explained by under-sampling of rare modern C1 carriers or, by severe demographic processes that may have acted on populations of Europe since prehistoric times.

INTRODUCTION

Current phylogeography of the human mitochondrial haplogroup C

Human mitochondrial haplogroup (Hg) C is part of the non-African macro-haplogroup M. Most of the diversity of Hg C is found today in indigenous populations of Asia and the Americas (Kong et al., 2003; Metspalu et al., 2006). In northern Asia, Hg C represents, with Hg D, more than half of the present-day mitochondrial (mtDNA) diversity (Derenko et al., 2010). Haplogroup Z, the sister-clade of Hg C, has a distribution ranging from northern Scandinavia (in Saami) to central Asia, Siberia, northern China and Korea. Analyses of complete mtDNA genomes reconstructed the phylogeny of Hg C into four sub-clades: C1, C4, C5 and C7 (e.g., Tamm et al., 2007; Achilli et al., 2008; Volodko et al., 2008; Perego et al., 2010; Derenko et al., 2010).

Hg C1 has one of the broadest distributions of all human mitochondrial haplogroups in the world: from Iceland in the West to East Asia and the Americas in the East. The C1 basal haplotype is defined by the hypervariable region I and II (HVR-I and HVR-II) motif: 16223T-16298C-16325C-16327T (HVR-I; numbering according to the revised Cambridge reference sequence rCRS; Andrews et al., 1999) and 73G-263G-489C (HVR-II).

Haplogroup C1 has been described as phylogenetically structured into five distinct monophyletic sub-clades - C1a, C1b, C1c, C1d, C1e - that exhibit a clear geographical distribution pattern (Starikovskaya et al., 2005; Tamm et al., 2007; Perego et al., 2010; Ebenesersdóttir et al., 2011; Figure 1). Three of the C1 sub-clades (C1b, C1c and C1d) are restricted to Native American populations, although spread widely across the American continent (Forster et al., 1996; Bandelt et al., 2003). It was hypothesized that these three Native American C1 sub-clades were the three ancestral founders brought to the Americas, along with Hg A2, B2 and D1, during the initial human colonisation of the continent (Starikovskaya et al., 2005; Tamm et al., 2007; Achilli et al., 2008; Perego et al., 2010). The source population of this migration was assumed to be in eastern Asia, where most of the diversity of Hg C is observed today, and where C1a, a sister clade of the American C1 clades, is found at low frequencies in diverse indigenous populations (Starikovskaya et al., 2005). This suggests an early differentiation from the ancestral C1 haplotype into the Asian (C1a) and American (C1b, C1c and C1d) sub-clades. Carriers of the latter subsequently migrated onto the Beringian ice-free land bridge that connected north-east Siberia and Alaska during the

last Ice Age and from there colonised the Americas (Shields et al., 1993; Starikovskaya et al., 2005). The place of origin of ancestral Hg C1 was approximated in the Amur River region just South of Beringia (eastern Asia) on the basis of the current distribution of Hg C1 in Asia (Starikovskaya et al., 2005).

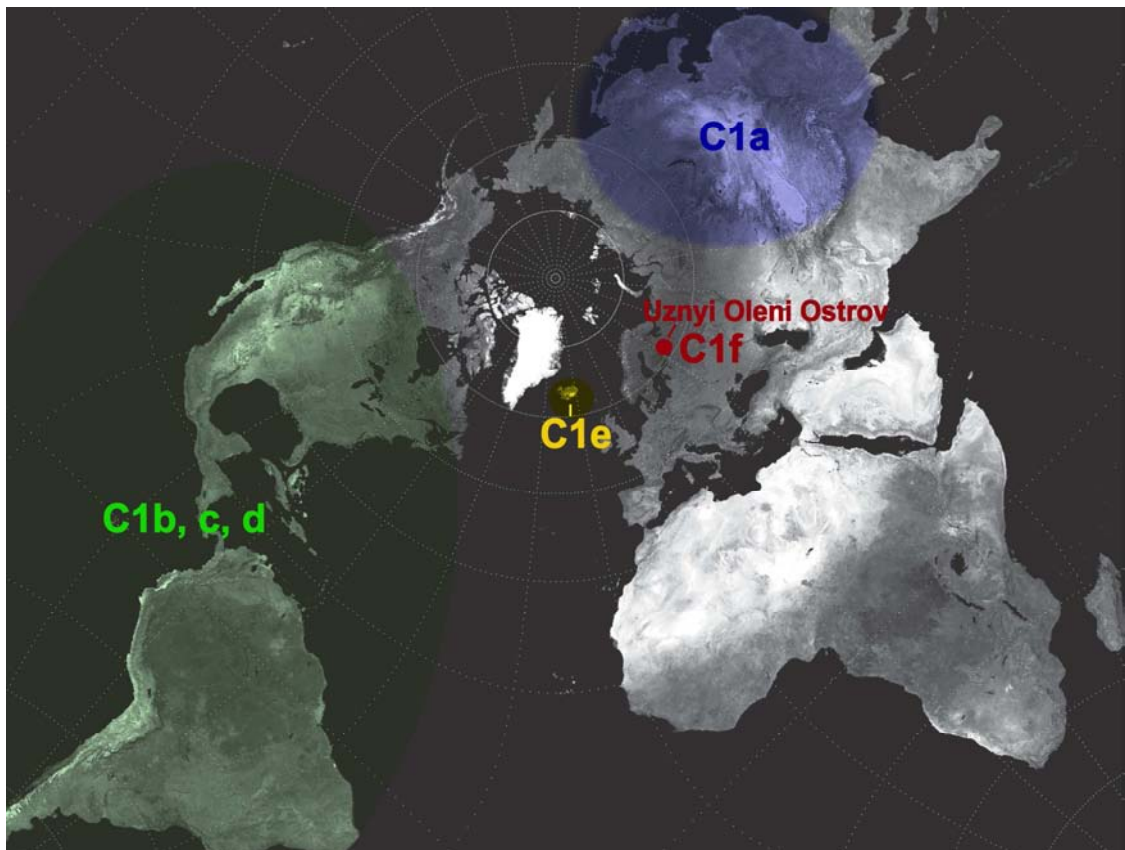


Figure 1. Approximate distribution of the C1 sub-clades in modern populations and location of the Uznai Oleni Ostrov archaeological site.

In Europe, the dense and extensive sampling of HVR-I diversity has revealed extremely low frequencies of Hg C1, with very few haplotypes (Figure 2) found in Germans (Pfeiffer et al., 1999), Canarians (Rando et al., 1999), Icelanders (Helgason et al., 2000; Helgason et al., 2003) and Bashkirs (Bermisheva et al., 2002). These sequences lack HVR-I Single Nucleotide Polymorphisms (SNPs) diagnostic of the sub-clades C1a (at nucleotide position, np, 16356) and C1d (at np 16051). A more detailed assignment of the European haplotypes into sub-haplogroups is limited by the

low resolution provided by HVR-I and the lack of information from the coding region. These limitations therefore impede the reconstruction of their phylogeny and origin.

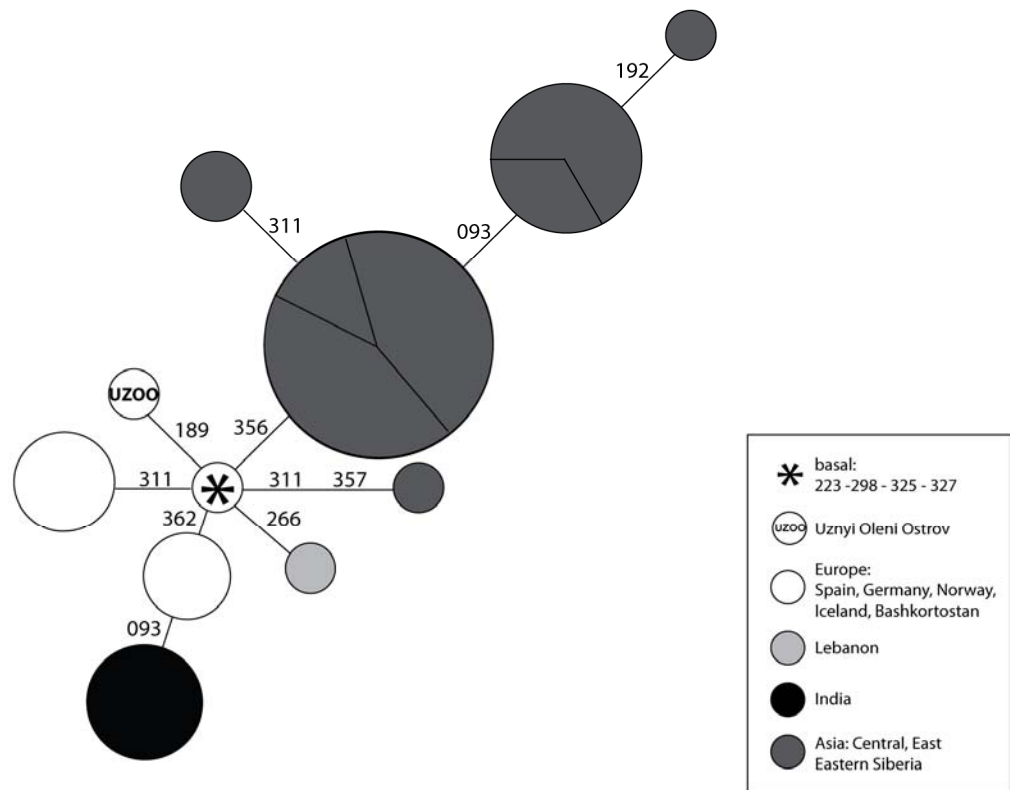


Figure 2. Network representation of haplogroup C1 HVR-I sequences in Mesolithic Uznyi Oleni Ostrov and modern Eurasian populations. Each haplotype is represented by a circle the area of which is proportional to the number of individuals that were found to carry this haplotype in the literature (See Supplementary Table S2 for description and references for the corresponding sequences). The haplotypes are colour-coded according to their geographical location: India (black), Asia (dark grey), Lebanon (light grey), and Europe (white). Each section of the circles represents individuals sampled from a same population. Mutations are all substitutions and are reported according to the revised Cambridge reference sequence minus 16,000. The star represents the HVR-I haplotype that characterises the root of the C1 clade (223-298-325-327). The haplotype labelled as 'UZOO' is the HVR-I haplotype sequenced from individuals of the archaeological site of Uznyi Oleni Ostrov. All the other haplotypes were found in modern populations.

Origins for the sub-clade C1 in Europe

Three hypotheses for the geographical origins of the European C1 lineages have been suggested (Helgason et al., 2000; Helgason et al., 2003; Ebenesersdóttir et al., 2011):

Model 1: Hg C1 in Europe has an American origin and has reached Europe through admixture between Native Americans and Europeans. This gene-flow may have occurred after the colonization of the New World by Europeans in the 15th century Anno Domini (A.D.), i.e., post-Columbian times (Helgason et al., 2000; Helgason et al., 2003). Alternatively, admixture may date to pre-Columbian times, as it is widely accepted that Icelandic Vikings have reached North America and built temporary pioneer settlements (Ebenesersdóttir et al., 2011). Consistent with this model, the European C1 HVR-I sequences could possibly belong to either the C1b or C1c American clades, as diagnostic SNPs for these two clades are located outside HVR-I (Helgason et al., 2000; Helgason et al., 2003).

Model 2: A recent genetic input from Asia into Europe during historical times is the source of Hg C1 in Europe. Historically, central and east Europe experienced repeated influxes from invasive groups from the neighbouring Asian steppes that could have introduced C1 into Europe. Well-documented examples include the Huns from Mongolia in the 4th - 5th centuries A.D. (1,500 – 1,700 years Before Present, yBP) and the Mongols in the 13th century A.D. (700 – 800 yBP; Grousset, 1970). However, the Asian C1 clade is characterised by the HVR-I transition at np 16356, which has not been found in any European C1 haplotype. In the case of a recent Asian ancestry, a reversal of the mutation at np 16356 in all the European sequences would be required. Alternatively, the Asian branch could represent a lineage that derived from the European-Asian C1 common ancestor.

Model 3: Haplogroup C1 could have been present in Europe since prehistoric times. The C1 HVR-I haplotypes found in Europe would then represent a distinct branch that diverged from a common ancestor somewhere in Eurasia and reached Europe during prehistoric migrations. This scenario is considered the most plausible in the light of C1 HVR-I haplotypes in three individuals (UZOO-7, UZOO-8, and UZOO-74) from the 7,500 year old Mesolithic site of Uznyi Oleni Ostrov, Karelia, in north west Russia (Chapter One).

Complete mitochondrial genome sequencing in ancient and present-day populations

To date, the precise phylogenetic relationship between the European C1 HVR-I lineages and the Asian and American clades remains unresolved. This gap in knowledge is caused by the lack of complete genome sequences for most European C1 mitochondrial mtDNA, which is probably a consequence of its extremely low frequency. More generally, the shortage in the number of sequenced mitochondrial genomes (mtgenomes) also impedes the resolution of the origins of Hg C1. So far, only one study addressed the phylogeny of European Hg C1 mtgenomes (Ebenesersdóttir et al., 2011), which were found in Iceland. This study showed that all present-day C1 lineages in Iceland belong to the same clade, named C1e, which is distinct from any of the Asian and American clades previously defined on the basis of mtgenome sequences (Ebenesersdóttir et al., 2011). Although an important step towards a more detailed reconstruction of Hg C1 diversity, this study does not resolve the origin of C1 in Europe due to the lack of comparative data from Europe, Asia and the Americas in modern, as well as ancient, populations.

In order to contribute to the description of the mitochondrial diversity within Hg C1, we sequenced the mtgenome of one of the three Mesolithic specimens from the Uznyi Oleni Ostrov archaeological site (individual UZOO-74). The classification of the corresponding mitochondrial haplotype within Hg C1 was previously determined by HVR-I sequencing (C1 characteristic mutation at np 16325) as well as by typing phylogenetically informative SNPs in the coding region (C characteristic mutations at np 13263; Chapter One).

Complete mtgenome sequences have successfully been determined for a 3,400 - 4,500 year-old Palaeo-Eskimo from Greenland (Gilbert et al., 2008), a 5,100 - 5,350 year-old Tyrolean mummy (Ermini et al., 2008) and a 30,000 year-old Western Russian individual from Kostenki (Krause et al., 2010). These mtgenomes were obtained by applying high-throughput ‘next-generation’ sequencing technologies, such as the 454 FLX sequence-by-synthesis platform (Roche; Margulies et al., 2005). In the present study, we used a different and novel method to sequence the ancient mtgenome of an Hg C1 carrier of Mesolithic Uznyi Oleni Ostrov. This alternative approach took advantage of the commercially available hybridisation array designed for the sequencing of mtgenomes, the Affymetrix® GeneChip® Human Mitochondrial resequencing Array 2.0 (MitoChip v.2.0; Maitra et al., 2004; Zhou et al., 2006;

Hartmann et al., 2009). A genomic library was first created from a DNA extract of the ancient specimen UZOO-74. This was achieved by enrichment for mtDNA in three iterative rounds of hybridisation to in-house designed biotinylated DNA probes, followed by re-sequencing via hybridisation to the commercial MitoChip v.2.0. This approach is currently under development at the Australian Centre for Ancient DNA (ACAD), University of Adelaide. This method resulted in the sequencing of ancient mtgenomes whose characteristic SNPs were concordant with the assignment of the sequences at unique locations within the human mitochondrial tree. DNA enrichment followed by MitoChip v.2.0 resequencing has then shown very powerful at recovering consistent phylogenetic signals and at discovering unknown SNPs from ancient DNA (aDNA) extracts (Brotherton, manuscript in preparation). A phylogeny of all currently available Hg C1 lineages was constructed and included the novel C1 mtgenome from the Uznyi Oleni Ostrov specimen. The objective of this study is to shed light on the evolutionary history and phylogeography of Hg C1, as well as on the population history of the Mesolithic Uznyi Oleni Ostrov hunter-gatherers.

MATERIAL and METHODS

DNA extraction

Among the three Mesolithic individuals from Uznyi Oleni Ostrov shown to carry Hg C1 (UZOO-7, UZOO-8, UZOO-74), individual UZOO-74 was selected for mtgenome sequencing on the basis of its subjectively good preservation and robust performance in previous PCR amplification experiments (Chapter One). DNA was re-extracted from individual UZOO-74 using an independent sample.

The outer surface of the each tooth was first decontaminated through exposure to ultraviolet (UV) light for 20 min on each side. Then, dirt was removed from the outer surface by gently wiping the teeth with a paper towel soaked in sodium hypochlorite (bleach). The outer surface was removed using a Dremel® drill, then the root was cut off from the crown. Finally, the root was ground into a fine powder in a Mikro-Dismembrator ball mill (Sartorius Stedim Biotech GmbH) and stored at 5°C until further use. Digestion was carried out by incubating the powdered teeth in 4.44 mL of buffer (0.5 M EDTA, pH 8.0; 0.5% N-lauroylsarcosine; 0.25 mg/μL proteinase K) overnight on a rotary mixer at 37°C. The digestion product was

centrifuged at 4,600 revolutions per minute (RPM) for 10 min and the supernatant was subjected to DNA isolation following the silica-based extraction protocol described below. A silica suspension was first prepared by adding 6 g of silica dioxide to 50 mL of distilled water. The suspension was left to settle for one hour before 40 mL of the supernatant containing the finest particles were removed and allowed to settle overnight. The final silica suspension characterized by silica particles in the medium size range was obtained by removing 30 mL of supernatant. 125 μ L of the resulting medium-sized silica suspension was added to the digestion supernatant and 16 mL of binding buffer (13.5 mL QG buffer (Qiagen), 1x Triton, 20 mM NaCl, 0.2 M acetic acid). The yellow coloration of the pH indicator included in the QG buffer was used to verify neutral pH conditions necessary to the binding of DNA to silica. After letting DNA bind to silica overnight at room temperature and under constant rotation, the sample was centrifuged for 3 min and the supernatant was poured off. The pellet was washed three times by resuspension in 80% ethanol (1 mL for the first wash and 0.9 mL for subsequent washes), centrifugation for 1 min and removal of the supernatant. Then, the pellet was air-dried for 30 min, resuspended and incubated 10 min in 200 μ L of TE buffer (10 mM Tris, 1 mM EDTA) pre-warmed to 50°C. Centrifugation of the pellet for 1 min allowed the collection of the final DNA extract in the supernatant, which was aliquoted and stored at -18°C until further use.

Enrichment of ancient human mitochondrial DNA

Ancient DNA extracts from specimens preserved in soil are expected to contain DNA molecules of various origins. In addition to the highly degraded DNA of the specimen under study, environmental, microbial DNA (bacteria and fungi), as well as from unidentified sources, has been shown to constitute a major proportion in the pool of DNA molecules present in aDNA extracts. For example, primate nuclear DNA in the extract obtained from a Neanderthal specimen was estimated to represent less than 6.2% of the total DNA extract (Green et al., 2006). The presence of a mixed population of DNA molecules from various organisms hampers the reliable sequencing of the DNA fragments of interest. Here, enrichment of ancient human mtDNA of the UZOO-74 individual was a crucial step prior to resequencing using the MitoChip v.2.0. The preliminary enrichment step had two aims: 1) bringing the DNA concentration above the concentration threshold required for obtaining unambiguous sequencing signals above the background noise of the MitoChip v.2.0; 2) preventing

sequencing noise arising from the co-extraction of environmental and microbial DNA when using the MitoChip v.2.0.

Ancient DNA libraries were enriched for human mtDNA using a hybridisation-based method described in Brotherton et al., manuscript in preparation. The first step of the library preparation consisted of enzymatic polishing and phosphorylation of aDNA molecules, which were subsequently ligated to library adaptors following the procedure described in Margulies et al., 2005. In the second step, tagged DNA molecules were amplified by PCR using primers specific to the adaptors. These first two steps allowed the non-specific amplification of DNA molecules present in the extract, i.e. mitochondrial/nuclear human and non-human DNA.

Specific enrichment of human mtDNA was achieved by ‘fishing out’ ancient human mtDNA using specific DNA probes in great excess. These probes were constructed by amplifying the complete mtgenome (Hg J) of a modern individual in presence of biotinylated deoxynucleoside triphosphate (dNTPs). This was achieved by long-range PCR amplification of two overlapping ~8,500 base pairs (bp)-long fragments followed by sonication to obtain oligonucleotides in the 200-600 bp size range. The probes bind to the ancient ‘target’ DNA and form heteroduplex constructs, thus ‘capturing’ human mtDNA present in the aDNA extract. This was achieved by following a protocol adapted from Wang & Brown, 1991; Patel & Sive, 1996; and Sagerström et al., 1997. Robust hybrids between biotinylated probes and target DNA were left to bind to Streptavidin beads and non-specific DNA molecules, i.e. not bound to Streptavidin, were removed by stringency washes (Brotherton et al., 2007). The remaining library templates were then re-amplified using library-specific PCR primers. The enrichment/re-amplification cycle was carried out twice more to produce a final DNA population of highly enriched human mtDNA fragments (Brotherton et al., manuscript in preparation).

Resequencing using the MitoChip v.2.0 (Affymetrix®)

The MitoChip v.2.0 is a hybridisation array that interrogates each of the 16 kb of the human mtgenome with eight unique in situ-synthesized 25-mer probes (four for each strand of the mtDNA molecule). For each strand, the central nucleotide of each of the four probes varies to allow hybridisation to each possible nucleotide (A, C, G or T). Common HVR-I and HVR-II variants are further interrogated by additional probes

on the chip (selected from the American Federal Bureau of Investigation database: www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm). Probes were designed on the basis of the sequence of the rCRS (Hg H). The MitoChip v2.0 allows for the detection of both known and unknown SNPs as well as heteroplasmic nucleotide positions, but not insertions or deletions (indels).

Amplification of the aDNA library obtained for individual UZOO-74 prior to analysis on the MitoChip v.2.0 was performed following the manufacturer's instructions. MitoChip v.2.0 resequencing was carried out at the Adelaide Microarray Centre, University of Adelaide. The resulting raw data of the MitoChip v.2.0 analysis, saved as Cell Intensity files (CEL), were imported into the Affymetrix® GeneChip® Sequence Analysis Software. Different algorithm parameters were used for the analysis of the CEL files in order to find a compromise between the call rate, i.e. the percentage of bases of the mtgenome successfully sequenced, and the accuracy, i.e., the concordance of the base calls. In particular, the analyses were performed using four different quality score thresholds (QST), following the recommendations of Hartman et al. (2009). Base calls, for which quality score as calculated by the algorithm implemented in the GeneChip® Sequence Analysis Software were found to be below the given QST were filtered out. Mitochondrial genome sequences obtained with QST values of 1 (most relaxed), 3, 6 and 12 (most stringent) were compared under a haploid model. The MitoChip v.2.0 chromatograms were visually inspected base by base for ambiguous base calls. Output consensus sequences of all four quality scores were aligned in GENEIOUS v5.1 (Drummond et al., 2010), from which a master consensus sequences was generated.

SNP confirmation by direct sequencing and minisequencing

Selected regions of the mtgenome were sequenced independently via single and multiplex PCR amplification followed by direct sequencing, and SNaPshot minisequencing (Haak et al., 2010) in order to:

- 1) verify the HVR-I sequence between positions 15997 and 16410;
- 2) verify the 22 coding region haplogroup diagnostic SNPs targeted by the GenoCoRe22 reaction;
- 3) interrogate the deletions at np 249, 290, 291, which are diagnostic of the CZ and C1 sub-haplogroups (Figure 3) but were not detected by the MitoChip v.2.0;
- 4) resolve two ambiguous signals produced by the MitoChip v.2.0

resequencing analysis at np 7551 and 8500;

5) confirm the new SNPs identified here in the mtgenome sequence of specimen UZOO-74, but also to determine in the two other C1 carriers from Uznyi Oleni Ostrov, UZOO-7 and UZOO-8.

PCRs for direct sequencing were carried out in the same conditions as those described in the 'Material and Methods' section of Chapter One, using the primers reported in Table S1.

Typing of coding-region SNPs using the GenoCore22 reaction was performed using the protocol described in the 'Material and Methods' section of Chapter One.

Authentication of the ancient mtDNA sequence

The aDNA sequence data was validated using three lines of evidence: 1) monitoring of contamination, 2) reproducibility, 3) phylogenetic consistency.

1) Pre-PCR DNA work was carried out at the ACAD, University of Adelaide, a purpose-built positive air pressure laboratory dedicated to aDNA studies, which is physically isolated from any molecular biology laboratory amplifying DNA. Routine decontamination of the laboratory surfaces and instruments involves exposure to UV radiation and thorough cleaning using DNA oxidants such as bleach, Decon (Decon labs) and Ethanol. In order to protect the laboratory environment from human DNA, researchers are required to wear protective clothes consisting of a whole body suit, a facemask, a face shield, gum boots, and three pairs of surgical gloves that are changed on a regular basis. Obvious large-scale contamination within the laboratory or in the reagents were monitored and controlled by blank controls (one extraction blank for every five ancient samples and two PCR/GenoCoRe22 blank controls for every six reactions). In addition, no haplotype similar to any of those possessed by laboratory members was consistently amplified from aDNA extracts.

2) For all three Uznyi Oleni Ostrov individuals, the HVR-I sequences and GenoCoRe22 profiles could be replicated from two independent samples extracted independently. For UZOO-74, the same HVR-I sequence and coding region SNPs were obtained by direct sequencing/GenoCoRe22 minisequencing and by resequencing using the MitoChip v.2.0. Private SNPs identified by the MitoChip v.2.0 were also confirmed by direct sequencing.

3) The phylogenetic consistency of the combination of variable positions in the mtgenome was an additional indicator of the authenticity of the sequence. 34

mutations were 100% consistent with the phylogenetic position of UZOO-74 on the Hg C1 branch. No deviation from this position could be detected. The combination of five additional mutations was found to be unique to the haplotype sequenced here. None of these additional mutations was found to define branches within the human mtDNA tree, thus providing little support for them arising from exogenous contamination. In addition, considering the multiple replications performed to type the SNPs of interest, jumping PCR is thought to have had little or no impact on the Hg C1 haplotype presented here.

Phylogeny of the C1 clade

In order to generate a phylogeny of the Hg C1 HVR-I sequences displaying the C1 mutational pattern 16223T-16298C-16325C-16327T in Eurasia were gathered from the literature (Table S2). Sequences of whole mtgenomes belonging to Hg C1 were also compiled from the online GenBank database on the basis of the list published in Ebenesersdóttir et al., 2011 (Table S2). Sequences were corrected for known sequencing errors and ambiguities and, in particular, sequence length polymorphisms following recommendations in Bandelt et al., 2006; Bandelt et al., 2008. Mutations at positions 16182C, 16183C, 16193.1C and 16519 were systematically ignored, as these positions are thought to represent mutational hotspots and/or recurrent sequencing artefacts, which create reticulations in the phylogenetic analysis (van Oven & Kayser, 2009). A tree was then constructed manually for complete mtgenome sequences on the basis of the tree constructed in Ebenesersdóttir et al., 2011.

RESULTS/DISCUSSION

Use of MitoChip v.2.0 for resequencing of ancient mitochondrial genomes

The approach - consisting in aDNA enrichment followed by MitoChip v.2.0 resequencing - implemented in this study allowed the unambiguous sequencing of 99.32% (16457 out of 16569 bp) of the UZOO-74 mtgenome dated around 7,500 yBP.

The novel method used here appeared particularly well suited to the taphonomic properties of aDNA for the following reasons: 1) very small DNA fragments (in the 30 – 60 bp size range) are targeted, providing preferential access to

short aDNA molecules over longer modern DNA contaminants, 2) it reduces common polymerase artefacts, e.g., ‘jumping PCR’ by relying on small number of amplification cycles, 3) it specifically targets DNA sequences of the species of interest over metagenomic DNA (bacterial, environmental) present in DNA extracts.

Missing data (i.e., allele drop-outs called ‘N’) in the Hg C1 mtgenome constituted 112 bp, spread throughout the genome for which a reliable sequencing signal could not be obtained (Table 1). A comparison across all mtgenomes generated at ACAD showed that regions of the mtgenomes for which no reliable sequencing signal could be retrieved were virtually identical from one mtgenome to the other. This suggests that missing data is non-randomly distributed across the genome. It was shown that this distribution was not due to an under-representation of the probes targeting the corresponding regions during the hybridisation steps, as DNA libraries have been constructed using different probe batch, i.e., independently generated. Conversely, a closer inspection of the missing regions suggested that the sequence of these regions represent template likely to be energetically sub-optimal for hybridisation, e.g., A-T rich sequences or homopolymeric stretches. As a result, further work is needed to improve the efficiency of probe hybridisation in the problematic regions (Brotherton et al., manuscript in preparation).

The ancient mitochondrial haplotype of individual UZOO-74 obtained with the MitoChip v.2.0 displayed nucleotide differences at 41 positions from the rCRS. Among these mutations, 34 represented all substitutions defining the C1 clade. However, sequencing using the MitoChip v.2.0 does not allow the detection of indels. As a consequence, the deletions at np 249, 290 and 291, which define the C1 clade (np 290, np 291) and the CZ (np 249) phylogenetic group to which C1 belong, could not be detected. The MitoChip v.2.0 yielded ambiguous signals (‘N’ calls) for two positions defining the C1 branch: Hg M8 diagnostic np 8584 and Hg C diagnostic np 9545. Consequently, the substitutions at these positions, as well as diagnostic deletions in HVR-II for Hg CZ and C, were verified by direct sequencing.

Once the 34 Hg C1 specific mutations out of the total 41 nucleotide differences detected by the MitoChip v.2.0 confirmed the phylogenetic position of the Uznyi Oleni Ostrov haplotype, seven nucleotide differences at np 247, 7551, 8500, 8577, 11605, 12217 and 16189 remained. These seven additional nucleotide differences were interrogated by direct sequencing in order to verify whether they represented true private mutations defining a novel C1 sub-clade. Two of the seven substitutions (np

7551 and 8500) could not be confirmed by direct sequencing in any of the three Uznyi Oleni Ostrov aDNA extracts (individuals UZOO-7, UZOO-8, and UZOO-74). Accordingly, they were considered as false positives of the MitoChip v.2.0 resequencing analysis. In contrast, the three transitions 247A, 8577G, 12217G, and the transversion 11605t in the coding region, as well as the transition 16189C in HVR-I could be confirmed by direct sequencing and were taken into account in further phylogenetic analysis of the C1 clade (Table 2). Importantly, we observed no SNPs characteristic of other haplogroups, nor any mixed signals that could indicate DNA degradation or the participation of DNA from multiple sources.

Table 1. Positions of missing bases ('N' calls) in the C1 complete mitochondrial genome when compared to the revised Cambridge Reference sequence.

Positions	Positions	Positions
238 - 246	7,332	9,536 - 9,537
4,763 - 4,768	7,519 - 7,521	9,545 - 9,546
4,770 - 4,772	7,523 - 7,524	9,999
5,085 - 5,090	7,526	10,001
5,094	7,531	10,450 - 10,453
5,304	7,535	10,455 - 10,456
5,315	7,545	14,497 - 14,499
5,498 - 5,499	7,547	14,501 - 14,504
5,501 - 5,511	7,549 - 7,550	14,775 - 14,782
5,514 - 5,517	7,552 - 7,553	14,784
7,300	7,556	
7,307 - 7,309	7,569	

Table 2. Positions and nucleotide changes in the Uznyi Oleni Ostrov C1f haplotype when compared to the revised Cambridge Reference Sequence.

Position¹	Mutation	Position¹	Mutation
73	A to G	10398	A to G
247	G to A	10400	C to T
249	A deletion	108713	T to C
263	A to g	11605	A to t
290	A deletion	11719	G to A
291	A deletion	11914	G to A
489	T to C	12217	A to G
750	A to G	12705	C to T
1438	A to G	13263	A to G
2706	A to G	14318	T to C
3552	A to a	14766	C to T
4715	A to G	14783	T to C
4769	A to G	15043	G to A
7028	C to T	15301	G to A
7196	C to a	15326	A to G
8577	A to G	15487	A to t
8584	G to A	16189	T to C
8701	A to G	16223	C to T
8860	A to G	16298	T to C
9540	T to C	16325	T to C
9545	A to G	16327	C to T

¹ Transitions are reported with upper case letters and transversions with lower case letters. Nucleotide changes in bold represent mutations in the Uznyi Oleni Ostrov haplotype that are new within the C1 clade.

Resequencing base call rate

It has been argued that the MitoChip v.2.0 base call rate may depend on the genetic distance between the probed genome sequence and the European rCRS (Hg H), whose sequence was used to construct the MitoChip v.2.0 probes (Hartmann et al., 2009). Probed genomes of African origin have been shown to yield lower base call rates than genomes of European origin. This lower average base call rate is likely explained by the impairment of hybridisation efficiency in flanking regions due to the presence of SNPs, which are more numerous in African genomes than in European genomes when compared to the (European) rCRS sequence of the chip probes. Where proximal SNPs cause lower hybridisation lower average base call rates result. The

more genetically distant from rCRS the probed DNA is, the larger the number of SNPs and the lower the base call rate.

In this regard, Hg C1 and Hg H can be considered relatively distant genetically, each belongs to one of the two non-African macro-haplogroups N and M. A possible consequence of any loss of signal in the resequencing of the present Hg C1 haplotype is that private mutations that may have been located in 122 positions with low signal strength were not identified. Potentially, this may have led to a partial phylogenetic reconstruction of the C1 haplotype. In the eventuality that mutations in the missing regions of this novel Hg C1 sequence remained undetected, these mutations are thought to be rare within the global human mtDNA tree. To support this idea, 200 complete human mtgenomes randomly drawn from the 8,731 mtgenomes currently available on the PhyloTree website (www.phylotree.org; build dated to the 11th of February 2011; van Oven & Kayser, 2009) were searched for potential SNPs located in the missing regions listed in Table 1. This search returned no candidate SNPs. The retrieval of 99.32% of the mtgenome haplotype for the Uznyi Oleni Ostrov individual has been sufficient to obtain a clear confirmation of its phylogenetic position within the human mitochondrial tree and to identify private mutations defining a new sub-clade.

Phylogeny of the mitochondrial C1 lineage in Mesolithic Uznyi Oleni Ostrov

After confirmation of the five novel mutations 247A, 8577G, 11605t, 12217G and 16189C in the C1 mtgenome of individual UZOO-74, it was possible to search for the same SNPs in the two other C1 carrier individuals UZOO-7 and UZOO-8. The method confirmed the presence of identical SNPs, suggesting that the three C1 individuals from Uznyi Oleni Ostrov were closely maternally related. Their mtgenomes may be entirely identical or display differences in the form of additional private SNPs at coding region positions that have not been sequenced in the remaining UZOO-7 and UZOO-8 individuals.

The five variants identified in this C1 lineage represented novel sub-clade defining mutations that have not been reported together within a single Hg C1 haplotype before. Therefore, the mitochondrial haplotype identified by MitoChip v.2.0 resequencing represents a distinct new clade that we designate 'C1f' following the conventional nomenclature (Figure 3). Therefore Hg C1 is now characterised by six monophyletic clades: C1a, C1b, C1c, C1d, C1e and C1f. The tree typology suggests

that the C1f branch split from the most recent common ancestor of the C1 clades and evolved independently.

The split of the six clades from the C1 root could theoretically be dated using mtgenome sequences. However, the use of molecular data for dating divergence times can be problematic. The most commonly used method in modern mtDNA studies is based on the calculation of the ρ statistic (Saillard et al., 2000). However, this method has been shown to produce inaccurate dates, especially when the genetic data were collected from populations with complex demographic histories (e.g., Cox et al., 2008). The haplogroup C1 phylogeny suggests that this is likely to be the case for the C1 clades. The tree obtained is indeed very imbalanced: with the three American clades showing signals of a recent expansion, contrary to the three other Asian, Icelandic and Mesolithic clades. Moreover, a limitation of molecular dating methods is the inaccuracy of the estimation of the human mitochondrial substitution rate and the fact that they rely on the use of a constant mutation rate through time (see section General Introduction/Dating using molecular data, and its limits). Improvements of the estimation of the mitochondrial substitution rate have been made, especially by correcting for purifying selection (Soares et al., 2009). Using this method, the modern clades C1a, C1b and C1c were calculated to have diverged around 17,100 yBP (95% Confidence Interval: 12,000 – 22,500 yBP). However, in the improved method, ancient mtgenome sequences cannot be used to calculate divergence times. The program BEAST (Drummond & Rambaut, 2007) can calculate divergence times from mtgenomes through the estimation of a substitution rate that is allowed to vary in time. BEAST can also deal with temporally heterogeneous mtgenome sequences. Haplogroup C1 is problematic because its tree is polytomous, i.e., all the C1 clades branch out a single common C1* node, and BEAST only deals with bifurcating trees. In addition; some biases of molecular dating associated with incomplete sampling of the genetic diversity were shown (Fagundes et al., 2010), where the coalescent age of the C1d clade was recalculated after adding new data. With only one C1f sequence and a very few of the Icelandic C1e having been reported, it is likely that the sampling of the genetic diversity for these clades is not extensive enough to use these methods and date divergence time.

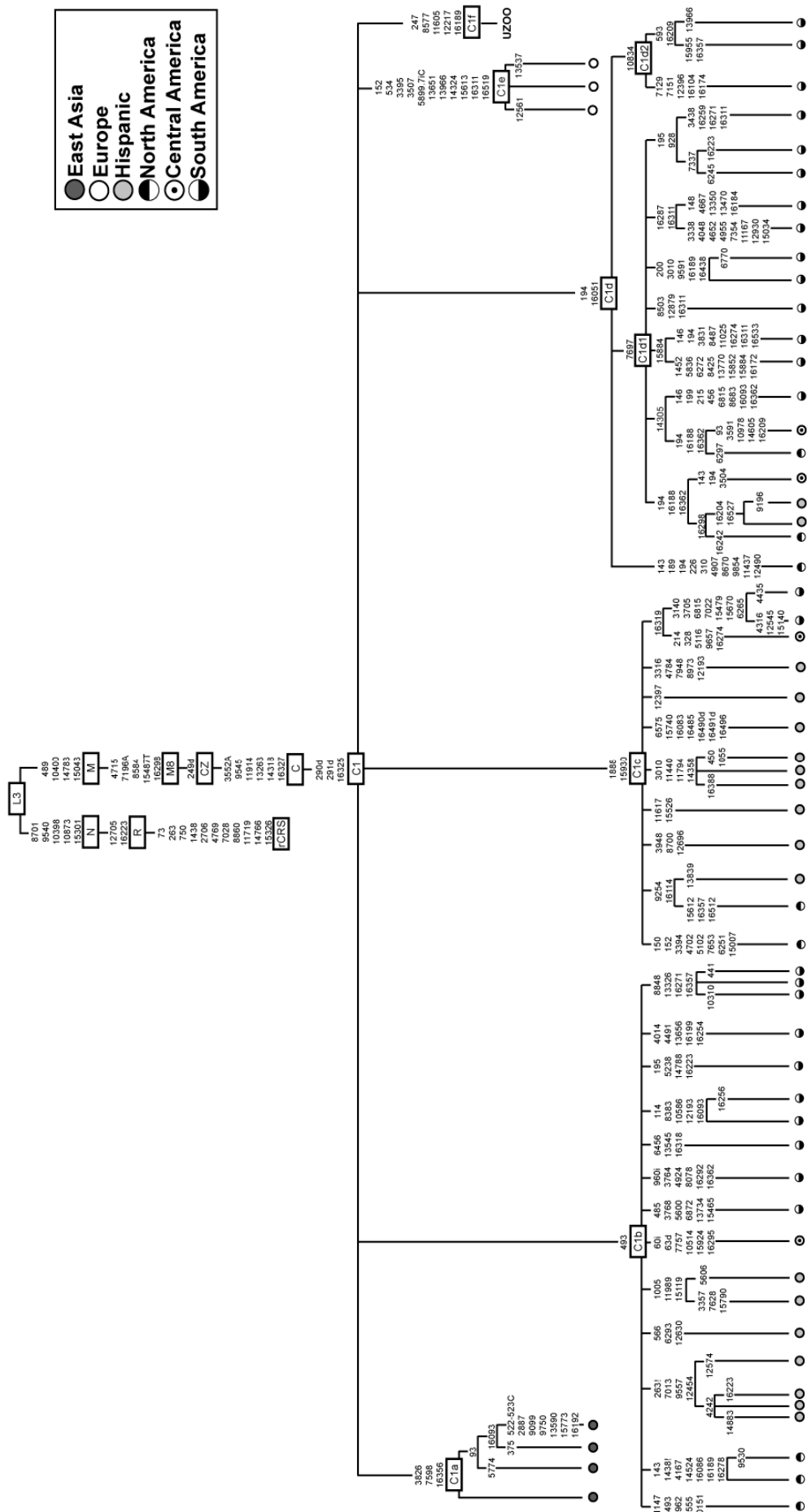


Figure 3. Median joining phylogenetic tree of haplogroup C1 complete mitochondrial genomes. A haplogroup L3 sequence was chosen as the root of the tree. Mutations are reported according to the revised Cambridge reference sequence. “i” represents deletions. “d” represents insertions. See Supplementary Table S2 for description and references for the corresponding sequences.

No match was found for the C1f mitochondrial haplotype in the database of modern complete mtgenomes (currently 8731 entries in the PhyloTree database; van Oven & Kayser, 2009). This absence could be explained by either extinction or near extinction of the lineages since the Mesolithic, or by under-sampling of mitochondrial genomes in modern human populations.

Under-sampling of whole mitochondrial genomes in Eurasia

The number of published modern *Homo sapiens* complete mtgenomes is still small compared to that of HVR-I sequences (around 8,731 and 200,000, respectively) and new studies still regularly report the discovery of new clades and lineages (e.g., Hg C1e Ebenesersdóttir et al. 2011; within Hg C1d, Perego et al., 2010). To date, the coverage of modern-day populations for complete mtgenome sequencing has been geographically heterogeneous and the sampling has focused on few specific populations. As a consequence, mtgenomes available from the literature still give an incomplete picture of the existing mitochondrial diversity.

Under-sampling of complete mtgenomes can explain the gap in the knowledge of the Hg C1 diversity in Eurasia, despite the fact that some Eurasian populations are densely sampled for HVR. Considering the HVR-I diversity described in modern-day Eurasia, close matches for the HVR-I sequence of C1f do not display the mutation at np 16189 (Figure 2) and hence, none matched the C1f HVR-I haplotype exactly. However, np 16189 is one of the most recurrent variable positions in human mtDNA (Bandelt et al., 2008) and its mutational instability gives it little phylogenetic discrimination power. It could then be possible that these haplotypes belong to the C1f clade without harbouring the mutation at nt 16189. Therefore, additional SNPs in the coding region would be needed to definitely rule out these Eurasian C1 haplotypes as potential members of the C1f clade. These potential C1f candidates in Eurasia are rare but widespread in Eurasia, hence clear inferences regarding the origin and evolutionary history of the C1f clade based on modern-day frequencies are difficult to infer. In the absence of identified geographical location for source populations for C1 in Eurasia, it is conceivable that these regions might not have been sampled for whole mtgenomes. Prehistoric populations of north east Europe (Mesolithic Uznyi Oleni Ostrov and Bronze Age Bolshoy Oleni Ostrov) were shown to exhibit mitochondrial affinities with modern populations of western and southern Siberia, the Altai region, or Mongolia (Chapter One), potential source populations could therefore be expected

in these regions. This is supported by the fact that most of the diversity of Hg C is found in present-day populations of East Eurasia. Even though no HVR haplotype closely related to the C1f Uznyi Oleni Ostrov sequence has been found in these populations so far, they have not been sampled as densely and extensively as, for example, European populations. Thus, if C1f is rare and has a reduced geographical distribution today, scant sampling of some modern Eurasian populations could explain the absence of direct matches for the C1f haplotype detected in Uznyi Oleni Ostrov in the literature. Given the current absence of exhaustive mtgenome data, that is well-sampled geographically and/or temporally, the absolute number of C1 clades and their phylogenetic relationships cannot be established. At present, it seems also unreasonable to reconstruct the timing of the arrival of Hg C1 lineages in Europe via coalescence age dating and/or determine whether all European C1 lineages reached Europe as part of the same migration as the Uznyi Oleni Ostrov C1f branch or as part of other movements from the East.

Effect of post-Mesolithic population dynamics

The absence of the C1f matrilineage in modern populations inhabiting the geographical area around the graveyard of Uznyi Oleni Ostrov could be explained by the extinction or near extinction of this European matrilineage since the Mesolithic. Previous studies reported the mtDNA structure of Mesolithic populations of central/eastern Europe (Bramanti et al., 2010) and Scandinavia (Malmström et al., 2010). Like the population of Uznyi Oleni Ostrov, previously described European Mesolithic populations were characterised by high frequencies of Hg U (Chapter One). The mtDNA structure observed in European Mesolithic populations is very dissimilar to the rather homogeneous mitochondrial makeup of present-day Europeans, which emerged during the Neolithisation and subsequent periods. This implies little direct genetic continuity between Mesolithic and modern-day Europeans as a result of significant population dynamics. In addition, Hg C1 could not be detected in any of the European Mesolithic populations, suggesting under-sampling of Mesolithic populations for aDNA, relatively low frequencies and geographical distribution for Hg C1f in the Mesolithic, and/or mating isolation of the Uznyi Oleni Ostrov population. The low frequency and restricted geographical distribution of Hg C1 may have made this matrilineage particularly vulnerable to the demographic processes, such as bottlenecks, genetic drift, large-scale migrations or population

replacements that may have occurred since the Mesolithic. Eventually, Hg C1 may have reached extremely low frequencies or have gone extinct, thus preventing it from being detected in present-day European populations.

Under-sampling of whole mitochondrial genomes in the Americas

Similarly to the present C1f situation, no individual carrying the European ‘Icelandic’ sister-clade C1e could be found outside Iceland (Ebenesersdóttir et al., 2011), providing equally few clues about the origin of this particular clade. Ebenesersdóttir et al, 2011, proposed an American origin for the C1e clade in Iceland through mating of Viking explorers with Native American women sometime before 300 years ago. The fact that the overwhelming majority of Hg C1 diversity is found on the American continent was used as an argument to support this hypothesis. Admixture with Vikings must have been limited as no other American-specific lineage (e.g., Hg A2, B2, D1, C1b, C1c, C1d) has been found in Iceland to give further support to this hypothesis. The absence of C1e in the Americas was suggested to result from the still scarce sampling of whole mtgenome diversity in the Americas.

In contrast, the possibility of a prehistoric genetic influence from the Americas into Mesolithic Europe is highly unlikely. However, in the eventuality that further sampling of complete mtgenomes in the Americas reveals the presence of additional haplotypes belonging to C1f, it would suggest an evolutionary history similar to that of mtDNA Hg X2. Like Hg C1, Hg X2 displays relatively low frequencies and a broad distribution (in the Near East, Europe, Central Asia, Siberia as well as North America for clade X2a; Reidla et al., 2003). One model for the present distribution of Hg X2 suggests that the X2a clade split early from the rest of the X2 lineages in the Near East, and spread to reach east Siberia before participating in the second wave of migration into the Americas through admixture with Beringian populations (Perego et al., 2009). A similar scenario involving an early split of the different C1 clades in the Near East or Central Eurasia followed by their spread and isolated evolution could be considered as an explanation for the wide geographical distribution of Hg C1. However, this scenario currently lacks substantial support.

A proposed shared genetic history for the Icelandic-specific C1e and the Mesolithic C1f European sub-clades

Rather than an American origin, the Icelandic-specific C1e clade could have had a recent origin in northern Europe and a shared history with the C1f sub-clade of Mesolithic north east Europe. This hypothesis is relevant with regard to the origins of the Icelandic population. Iceland was discovered and first settled by Scandinavian Vikings around 1,130 years ago. Vikings raids extended as far from their homeland in Scandinavia as France, Spain and Sicily, but their main expansion range comprised western Russia, the Baltic region, Scandinavia, and the British Isles (Helgason et al., 2000). The study of the mtDNA diversity of present-day Icelanders identified that most of the Icelandic mtDNA lineages had Norse (from Scandinavia) or Gaelic origins (from the British Isles) and that the Icelandic gene pool had strongly been impacted by genetic drift (Helgason et al., 2000; Helgason et al., 2003).

Considering the population history of Icelanders, as well as the identification of the monophyletic C1f clade in Mesolithic north east Europe, the following alternative scenario for the history of the C1e and C1f sub-clades can be proposed. The Icelandic-specific C1e and Mesolithic C1f lineages may have both split from the common ancestors of the C1 lineages somewhere in Eurasia and later reached northern Europe during independent or similar migrations (before the Mesolithic for C1f). The rare occurrence of the C1e and C1f sub-clades in Europe could be the result of their dilution within the pre-existing European mtDNA diversity when these lineages reached Europe. Of note, a contrasting pattern of elevated frequency and diversity was observed for the American sub-clades of Hg C1 (C1b, C1c and C1d), which signals important population expansion during the initial peopling of the continent. The presence of a sub-clade (C1f) closely related to the Icelandic-specific C1e sub-clade in a region neighbouring the homeland of the Vikings give support to the hypothesis that Hg C1e might have been part of the gene pool of the Vikings who first colonised Iceland. The C1e sub-clade might have been preserved at detectable frequencies in the Icelandic population under the effect of founder event, but have virtually gone extinct from the source population in northern Europe as a consequence of its low frequency.

CONCLUSION

The specific enrichment of ancient human mtDNA by hybridisation prior to complete mtgenome resequencing using the MitoChip v.2.0 has been shown to be an effective alternative to next generation sequencing approaches (Brotherton, manuscript in preparation). The method proved to be powerful at overcoming problematic characteristics of aDNA, such as high fragmentation and low concentration of target human mtDNA compared to co-extracted environmental DNA (mainly of bacterial origin).

This method recovered 99.32% of the mtgenome from a 7,500 year old ancient human specimen from the Mesolithic site of Uznyi Oleni Ostrov, western Russia. This mtgenome is unique and belongs to Hg C1, which is widely distributed outside Africa but rare in Europe. The C1 haplotype defines a new sub-clade, named here 'C1f'. Future sampling of ancient and modern populations for complete mtgenomes might help localising, in time and space, the origins of the C1f lineages, thus providing insights into the origins of the Mesolithic population of Uznyi Oleni Ostrov. Future access to additional complete mitochondrial sequences from ancient or modern populations could also find further important relevance for the understanding of early human migrations in Eurasia, but also potentially of the colonisation of the Americas.

The retrieval of the Hg C1f mtgenome is exceptional in many regards. The C1 mtgenome was dated back 7,500 yBP, belongs to a Hg C1 sub-clade that had not been detected previously in modern-day humans, and finally was found outside the geographical range of Hg C1.

LIST OF SUPPLEMENTARY MATERIALS

Table S1. Sequences of primers used in this study.

Table S2. Description and reference for the C1 complete mitochondrial genomes used to construct HVRI and mitochondrial genome C1 phylogenies.

ACKNOWLEDGMENTS

We acknowledge Paul Brotherton of the Australian Centre for Ancient DNA, University of Adelaide, for designing the approach used in this study, developing the hybridisation-based method for ancient human mtDNA and performing the aDNA enrichment, library construction and preparation of the sample for analysis on the

MitoChip v.2.0. We also thank Jennifer Templeton of the Australian Centre for Ancient DNA, University of Adelaide, for technical support.

REFERENCES

1. Achilli, A., Perego, U., Bravi, C., Coble, M., Kong, Q., Woodward, S., Salas, A., Torroni, A., Bandelt, H. (2008). The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* 3, e1764.
2. Andrews, R., Kubacka, I., Chinnery, P., Lightowlers, R., Turnbull, D., Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23, 147.
3. Bandelt, H., Kivisild, T. (2006). Quality assessment of DNA sequence data: autopsy of a mis-sequenced mtDNA population sample. *Ann Hum Genet* 70, 314-326.
4. Bandelt, H., Parson, W. (2008). Consistent treatment of length variants in the human mtDNA control region: a reappraisal. *Int J Legal Med* 122, 11-21.
5. Bermisheva, M., Tambets, K., Villems, R., Khusnutdinova, E. (2002). [Diversity of mitochondrial DNA haplotypes in ethnic populations of the Volga-Ural region of Russia]. *Mol Biol (Mosk)* 36, 990-1001.
6. Bramanti, B., Thomas, M., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M., Jankauskas, R., Kind, C., Lueth, F., Terberger, T., Hiller, J., Matsumara, S., Forster, P., Burger, J. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326, 137-140.
7. Brotherton, P., Endicott, P., Sanchez, J., Beaumont, M., Barnett, R., Austin, J., Cooper, A. (2007). Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35, 5717-5728.
8. Cox, M.P. (2008). Accuracy of molecular dating with the rho statistic: deviations from coalescent expectations under a range of demographic models. *Hum Biol* 80(4):335-57.
9. 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(12), e15214.
10. Drummond, A., Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7, 214.
11. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A. (2010). Geneious v5.1, Available from <http://www.geneious.com>
12. Ebenesersdóttir, S.S., Sigurdsson, A., Sanchez-Quinto, F., Lalueza-Fox, C., Stefansson, K., 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.
13. Ermini, L., Olivieri, C., Rizzi, E., Corti, G., Bonnal, R., Soares, P., Luciani, S., Marota, I., De Bellis, G., Richards, M.B., Rollo, F. (2008). Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr Biol* 18, 1687-1693.

14. Fagundes, N., Kanitz, R., Bonatto, S. (2008). A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization of Beringia. *PLoS One* 3, e3157.
15. Forster, P., Harding, R., Torroni, A., Bandelt, H. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59, 935-945.
16. Gilbert, M.T., Kivisild, T., Grønnow, B., Andersen, P.K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Götherström, A., Campos, P.F., Rasmussen, M., Metspalu M.; Higham, T.F., Schwenniger, J.L., Nathan, R., De Hoog, C.J., Koch, A., Møller, L.N., Andreasen, C., Medgaard, M., Villems, R., Bendixen, C., Willerslev, E. (2008). Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* 320, 1787-1789.
17. Green, R. E., Krause, J., Ptak, S. E., Briggs, A. W., Ronan, M. T., Simons, J. F., Du, L., Egholm, M., Rothberg, J. M., Paunovic, M., Pääbo, S. (2006). Analysis of one million base pairs of Neanderthal DNA. *Nature* 444, 330-336.
18. Grousset, R. (1970) *The Empire of the Steppes: History of Central Asia*. Ed. Rutgers University Press.
19. Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K.W., Burger, J. (2005). Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310, 1016-1018.
20. Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., Der Sarkissian, C.S., Brandt, G., Schwarz, C., Nicklisch, N., Dresely, V., Fritsch, B., Balanovska, E., Villems, R., Meller, H., Alt, K.W., Cooper, A., Genographic consortium. (2010). Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. *PLoS Biol* 8, e1000536.
21. Hartmann, A., Thieme, M., Nanduri, L.K., Stempfl, T., Moehle, C., Kivisild, T., Oefner, P.J. (2009). Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum Mutat* 30, 115-122.
22. Helgason, A., Sigurdardottir, S., Nicholson, J., Sykes, B., Hill, E.W., Bradley, D.G., Bosnes, V., Gulcher, J.R., Ward, R., Stefansson, K. (2000). Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. *Am J Hum Genet* 67, 697-717.
23. Helgason, A., Nicholson, G., Stefansson, 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.
24. Ho, S., Phillips, M., Cooper, A., Drummond, A. (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol* 22, 1561-1568.
25. Ho, S., Endicott, P. (2008). The crucial role of calibration in molecular date estimates for the peopling of the Americas. *Am J Hum Genet* 83, 142-146; author reply 146-147.
26. Ingman, M., Kaessmann, H., Pääbo, S., Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* 408, 708-713.
27. Ingman, M., Gyllensten, U. (2007). Rate variation between mitochondrial domains and adaptive evolution in humans. *Hum Mol Genet* 16, 2281-2287.
28. Just, R.S., Diegoli, T.M., Saunier, J.L., Irwin, J.A., Parsons, T.J. (2008). Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals. *Forensic Sci Int Genet* 2, e45-48.

29. Kong, Q. P., Yao, Y. G., Sun, C., Bandelt, H. J., Zhu, C. L., Zhang, Y. P. (2003) Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet* 73: 671–676.
30. Krause, J., Briggs, A., 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.
31. Maitra, A., Cohen, Y., Gillespie, S. E., Mambo, E., Fukushima, N., Hoque, M. O., Shah, N., Goggins, M., Califano, J., Sidransky, D., Chakravarti, A. (2004). The Human MitoChip: a high-throughput sequencing microarray for mitochondrial mutation detection. *Genome Res* 14, 812-819.
32. Malmström, H., Gilbert, M., Thomas, M., Brandström, M., Storå, J., Molnar, P., Andersen, P., 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.
33. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380.
34. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D. M., Gilbert, M. T., Endicott, P., Mastana, S., Papiha, S. S., 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.
35. Metspalu, M., Kivisild, T., Bandelt, H.-J., Richards, M., Villems, R. (2006) The pioneer settlement of modern humans in Asia. In: Bandelt, H.-J., Macaulay, V., Richards, M., eds. *Human mitochondrial DNA and the evolution of Homo sapiens*. Berlin: Springer-Verlag. pp 181–199.
36. Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., Sukernik, R.I., Olckers, A., Wallace, D.C. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A* 100, 171-176.
37. Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annu Rev Genet* 38, 645-679.
38. Patel M., Sive H.L. (1996). Subtractive cDNA cloning. in *Protocols in Molecular Biology*. Unit 5.11.
39. Perego, U. A., Achilli, A., Angerhofer, N., Accetturo, M., Pala, M., Olivieri, A., Kashani, B. H., Ritchie, K. H., Scozzari, R., Kong, Q. P., Myres, N. M., Salas, A., Semino, O., Bandelt, H. J., Woodward, S. R., Torroni, A. (2009). Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol* 19, 1-8.

40. Perego, U. A., Angerhofer, N., Pala, M., Olivieri, A., Lancioni, H., Kashani, B. H., Carossa, V., Ekins, J. E., Gómez-Carballa, A., Huber, G., Zimmermann, B., Corach, D., Babudri, N., Panara, F., Myres, N. M., Parson, W., Semino, O., Salas, A., Woodward, S. R., Achilli, A., Torroni, A. (2010). The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res* 20, 1174-1179.
41. Pfeiffer, H., Brinkmann, B., Hühne, J., Rolf, B., Morris, A. A., Steighner, R., Holland, M. M., 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.
42. Rando, J.C., Cabrera, V.M., Larruga, J.M., Hernandez, M., Gonzalez, A.M., Pinto, F., Bandelt, H.J. (1999). Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* 63, 413-428.
43. Reidla, M., Kivisild, T., Metspalu, E., Kaldma, K., Tambets, K., Tolk, H. V., Parik, J., Loogväli, E. L., 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, J. P., Chaventré, A., Khusnutdinova, E., Osipova, L., Stepanov, V., Voevoda, M., Achilli, A., Rengo, C., Rickards, O., De Stefano, G. F., Papiha, S., Beckman, L., Janicijevic, B., Rudan, P., Anagnou, N., Michalodimitrakis, 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.
44. Saillard, J., Forster, P., Lynnerup, N., Bandelt, HJ, Nørby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67(3):718-26.
45. Sagerström, C.G., Sun, B.I., Sive, H.L. (1997). Subtractive cloning: past, present, and future. *Annu Rev Biochem* 66, 751-783.
46. Sampietro, M., Lao, O., Caramelli, D., Lari, M., Pou, R., Martì, M., Bertranpetit, J., Lalueza-Fox, C. (2007). Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc Biol Sci* 274, 2161-2167.
47. Shields, G., Schmiechen, A., Frazier, B., Redd, A., Voevoda, M., Reed, J., Ward, R. (1993). mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am J Hum Genet* 53, 549-562.
48. Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A., Salas, A., Oppenheimer, S., Macaulay, V., Richards, M. (2009). Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *American Journal of Human Genetics*, 740-759.
49. Starikovskaya, E.B., Sukernik, R.I., Derbeneva, O.A., Volodko, N.V., Ruiz-Pesini, E., Torroni, A., Brown, M.D., Lott, M.T., Hosseini, S.H., Huoponen, K., Wallace, D.C. (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.
50. Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D. G., Mulligan, C. J., Bravi, C. M., Rickards, O., Martinez-Labarga, C., Khusnutdinova, E. K., Fedorova, S. A., Golubenko, M. V., Stepanov, V. A., Gubina, M. A., Zhadanov, S. I., Ossipova, L. P., Damba, L., Voevoda, M. I., Dipierri, J. E., Villems, R., Malhi, R. S. (2007). Beringian standstill and spread of Native American founders. *PLoS One* 2, e829.

51. van Oven, M., Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30, E386-394. www.phylotree.org/
52. Volodko, N., Starikovskaya, E., Mazunin, I., Eltsov, N., Naidenko, P., Wallace, D., Sukernik, R. (2008). Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. Am J Hum Genet 82, 1084-1100.
53. Wang, Z., Brown, D.D. (1991). A gene expression screen. Proc Natl Acad Sci U S A 88, 11505-11509.
54. Zhou, S., Kassaei, K., Cutler, D.J., Kennedy, G.C., Sidransky, D., Maitra, A., Califano, J. (2006). An oligonucleotide microarray for high-throughput sequencing of the mitochondrial genome. J Mol Diagn 8, 476-482.

SUPPLEMENTARY MATERIALS

Table S1. Sequences of the primers used in this study.

	Primer name	Primer sequence (5'-3')	Positions of the targeted fragment	Fragment size (base pairs)	Reference
HVR I	L15996	CTCCACCAATTAGCACCCAAAGC	15977 - 16161	185	Vigilant et al., 1991
	H16142	ATGTACTACAGGTGGTCAAG			Stone & Stoneking, 1998
	L16117	TACATTACTGCCAGCCACCAT	16097 - 16237	141	Haak et al., 2005
	H16218	TGTGTAGATTGAGGGTTG			Handt et al., 1996
	L16209	CCCATGCTTACAAAGCAAGT	16190 - 16322	133	Handt et al., 1996
	H16303	TGGCTTTATGTACTATGTAC			Handt et al., 1996
	L16287	CAC TAGGATACCAACAACC	16268 - 16429	162	Handt et al., 1996
	H16410	GCGGGATATTGATTTACCGG			Handt et al., 1996
HVR II	L00205	ACAGGGCGAACATACATTACTAAAGTG	00181 - 00261	105	This study
	H00261	GTTATGATGTCGTGGAAAGTGG			This study
	L00270	TGAATGTCGCACAGCCACTTTCCA	00246 - 00350	105	This study
	H00327	TTGGCAGAGATGTGTTAAGTGCT			Haak et al., 2008
Coding Region	Primer name	Primer sequence (5'-3')			Reference
	M8	TTGCGTTTCATTTCATGGCCCC	08542 - 08624	110	This study
	C1f	AGTCGGTTGTTGATGAGATATTTGGAGG			This study
	C	CTGAGCCCTTTTACCACCTCCAGCCTA	09500 - 09559	83	This study
	C1f	GATTTAGCGGGGTGATGCCCTGTG			This study
	C1f	ACAAAGCTCCATCTGCCCTACGACA	11573 - 11631	80	This study
	C1f	ACTACGAGGGCTATGTGGCTGA			This study
	L12197	TGACAAACAGAGGCTTACGACCCC	12175 - 12197	68	This study
	H12223	TGGGGGCATGAGTTAGCAGT			This study

Table S2. Description and reference for the C1 complete mitochondrial genomes used to construct HVRI and mitochondrial genome C1 phylogenies.

C1 clade	Geographic / Ethnic origin	Reference	GenBank accession number
HVR-I sequences			
C1?	Lebanese	Balanovsky, personal communication	
C1?	Lebanese	Balanovsky, personal communication	
C1?	Bashkir	Bermisheva et al., 2002	
C1?	Bashkir	Bermisheva et al., 2002	
C1?	Bashkir	Bermisheva et al., 2002	
C1?	Icelander	Helgason et al., 2000	
C1?	Icelander	Helgason et al., 2000	
C1?	Icelander	Helgason et al., 2003	
C1?	Thai	Kampuansai et al., 2006	
C1?	Indian	Metspalu et al., 2004	
C1?	Indian	Metspalu et al., 2004	
C1?	Indian	Metspalu et al., 2004	
C1?	Indian	Metspalu et al., 2004	
C1?	Indian	Metspalu et al., 2004	
C1?	Germany	Pfeiffer et al., 1999	
C1?	Canary Islander	Rando et al., 1999	
C1a	Altaiian	Balanovsky, personal communication	
C1a	Altaiian	Balanovsky, personal communication	
C1a	Orok	Bermisheva et al., 2005	
C1a	Orok	Bermisheva et al., 2005	
C1a	Orok	Bermisheva et al., 2005	
C1a	Orok	Bermisheva et al., 2005	
C1a	Orok	Bermisheva et al., 2005	
C1a	Orok	Bermisheva et al., 2005	
C1a	Orok	Bermisheva et al., 2005	
C1a	Highland Kirghiz	Comas et al., 1998	
C1a	Japanese	Horai et al., 1990	
C1a	Japanese	Imaizumi et al., 2002	
C1a	Japanese	Imaizumi et al., 2002	
C1a	Mongolian	Kolman et al., 1996	
C1a	Mongolian	Kolman et al., 1996	
C1a	Daur	Kong et al., 2003	
C1a	Korea	Lee et al., 2006	
C1a	Kalmyk	Nasidze et al., 2005	
C1a	Kalmyk	Nasidze et al., 2005	
C1a	Buryat	Shimada et al., GenBank	
C1a	Buryat	Shimada et al., GenBank	
C1a	Buryat	Shimada et al., GenBank	
C1a	Buryat	Shimada et al., GenBank	
C1a	Buryat	Shimada et al., GenBank	

C1 clade	Geographic / Ethnic origin	Reference	GenBank accession number
C1a	Buryat	Shimada et al., GenBank	
C1a	Ulchi	Starikovskaya et al., 2005	
C1a	Starikovskaya	Starikovskaya et al., 2005	
C1a	Japanese	Tajima et al., 2004	
C1a	Buryat	Tajima et al., 2004	
C1a	Japanese	Tanaka et al., 2004	
Complete mitochondrial genomes			
C1a	Japanese	Tanaka et al., 2004	AP008311
C1a	Ulchi, Siberia	Starikovskaya et al., 2005	AY519496
C1a	Buryat, Siberia	Derenko et al., 2007	EF153779
C1a	Nanaitci, Siberia	Ingman & Gyllensten, 2007	EU007858
C1b	Canary	Maca-Meyer et al., 2001	AF382009
C1b	North America	Mishmar et al., 2003	AY195759
C1b	Hispanic	Just et al., 2008	DQ282447
C1b	Hispanic	Just et al., 2008	DQ282448
C1b	Hispanic	Just et al., 2008	DQ282449
C1b	Hispanic	Just et al., 2008	DQ282450
C1b	Hispanic	Just et al., 2008	DQ282451
C1b	Hispanic	Just et al., 2008	DQ282452
C1b	Hispanic	Just et al., 2008	DQ282453
C1b	Hispanic	Just et al., 2008	DQ282454
C1b	Hispanic	Just et al., 2008	DQ282455
C1b	Hispanic	Just et al., 2008	DQ282456
C1b	Hispanic	Just et al., 2008	DQ282457
C1b	Hispanic	Just et al., 2008	DQ282458
C1b	Hispanic	Just et al., 2008	DQ282461
C1b	Hispanic	Just et al., 2008	DQ282464
C1b	Hispanic	Just et al., 2008	DQ282469
C1b	Hispanic	Just et al., 2008	DQ282475
C1b	Hispanic	Just et al., 2008	DQ282476
C1b	South America, Zoro	Fagundes et al., 2008	EU095223
C1b	South America, Zoro	Fagundes et al., 2008	EU095224
C1b	South America, Quechua,	Fagundes et al., 2008	EU095225
C1b	South America, Quechua	Fagundes et al., 2008	EU095226
C1b	Brazil, Arara	Fagundes et al., 2008	EU095227
C1b	South America, Poturujara,	Fagundes et al., 2008	EU095228
C1b	Brazil/Venezuela, Yanomamö	Fagundes et al., 2008	EU095229
C1b	Brazil/Venezuela, Yanomamö	Fagundes et al., 2008	EU095230
C1b	Brazil/Venezuela, Yanomamö,	Fagundes et al., 2008	EU095231
C1b	Colombia/Venezuela, Wayuu	Tamm et al., 2007	EU095549
C1b	United States	Achilli et al., 2008	EU431085
C1b	Mexico, Pima	Hartmann et al., 2009	EU597545
C1b	Mexico, Pima	Hartmann et al., 2009	EU597557
C1c	South America, Arsario	Tamm et al., 2007	EU095527
C1c	Colombia, Kogui	Tamm et al., 2007	EU095544

C1 clade	Geographic / Ethnic origin	Reference	GenBank accession number
C1c	Dominican Republic	Achilli et al., 2008	EF079875
C1c	Canada	Achilli et al., 2008	EU431086
C1c	United States	Achilli et al., 2008	EU431087
C1c	Hispanic	Just et al., 2008	DQ282459
C1c	Hispanic	Just et al., 2008	DQ282460
C1c	Hispanic	Just et al., 2008	DQ282462
C1c	Hispanic	Just et al., 2008	DQ282463
C1c	Hispanic	Just et al., 2008	DQ282465
C1c	Hispanic	Just et al., 2008	DQ282466
C1c	Hispanic	Just et al., 2008	DQ282467
C1c	Hispanic	Just et al., 2008	DQ282468
C1c	Hispanic	Just et al., 2008	DQ282470
C1c	Hispanic	Just et al., 2008	DQ282471
C1c	Unknown	Family Tree DNA	EU327891
C1c	Unknown	Family Tree DNA	EU327973
C1c	Unknown	Family Tree DNA	EU617323
C1d	Venezuela/Guyana, Warao	Ingman et al., 2000	AF347012
C1d	Venezuela/Guyana, Warao	Ingman et al., 2000	AF347013
C1d	Colombia, Corequaje	Tamm et al., 2007	EU095537
C1d	Hispanic	Just et al., 2008	DQ282472
C1d	Hispanic	Just et al., 2008	DQ282473
C1d	Hispanic	Just et al., 2008	DQ282474
C1d	Brazil/Guyana, WaiWai	Fagundes et al., 2008	EU095222
C1d	Canada, British Columbia	Malhi et al., 2010	GU215075
C1d	Colombia, Mestizos	Perego et al., 2010	HM107313
C1d	Colombia, Mestizos	Perego et al., 2010	HM107314
C1d	Colombia, Mestizos	Perego et al., 2010	HM107315
C1d	USA, Montana	Perego et al., 2010	HM107319
C1d	Mexico, Zacatecas	Perego et al., 2010	HM107321
C1d	Mexico, Sonora	Perego et al., 2010	HM107322
C1d	Argentina, Salta	Perego et al., 2010	HM107323
C1d	Argentina, Buenos Aires,	Perego et al., 2010	HM107327
C1d	Argentina, Buenos Aires	Perego et al., 2010	HM107329
C1d	Mexico, Oaxaca	Perego et al., 2010	HM107334
C1d	USA, Texas,	Perego et al., 2010	HM107335
C1d	Argentina, Buenos Aires	Perego et al., 2010	HM107346
C1d	Uruguay	Perego et al., 2010	HM107348
C1d	Chile, Bio-bio	Perego et al., 2010	HM107350
C1e	Icelandic	Ebenesersdottir et al., 2011	n/a
C1e	Icelandic	Ebenesersdottir et al., 2011	n/a
C1e	Icelandic	Ebenesersdottir et al., 2011	n/a

In Chapters One and Two, I showed temporal patterns of mitochondrial influences from both western (Europe) and eastern (Siberia) Eurasia in north eastern Europe. In Chapter Three, I investigate the mitochondrial gene pool of a prehistoric population of south east Europe, the Scythians, who were horse-riding nomads of the Iron Age.

Chapter Three:

The Mitochondrial Gene Pool of Scythians of the Rostov Area, Russia: A Melting Pot of Eurasian Influences

Abbreviations: ACAD, Australian Centre for Ancient DNA; aDNA, ancient DNA; ALT, Altaians; BA, Bronze Age; Ct, threshold cycle; dNTP, deoxynucleoside triphosphate; EG, Egyin Gol; Exo, exonuclease; F_{ST} , fixation index; HVR-I, hypervariable region I; IA, Iron Age; KAZ, Kazakhs; KUR, Kurgans; LOK, Lokomotiv; mtDNA, mitochondrial DNA; PCA, principal component analysis; qPCR, quantitative real-time Polymerase Chain Reaction; rCRS, revised Cambridge reference sequence; RPM, revolutions per minute; RSA, Rabbit Serum Albumin; SAP, Shrimp Alkaline Phosphatase; SAR, Sar; SBE, single base extension; SCY, Scythians; SNP, single nucleotide polymorphism; TAR, Tarim; UPF, University of Pompeu Fabra; yBP, years Before Present.

ABSTRACT

Scythians are known from written sources to have been a horse-riding nomadic people who inhabited the region north of the Black Sea around 2,700 to 2,800 years Before Present (yBP). However, their genetic origins remain mysterious. The nature of the cultural and genetic relationships between Scythians and other Bronze Age (2,800 – 5,500 yBP) and Iron Age (2,300 – 2,800 yBP) nomadic populations of the Eurasian Steppe are still debated. Due to climatic conditions favourable for DNA preservation in parts of the Eurasian Steppe, the mitochondrial diversity of a range of Bronze Age and Iron Age populations has already been described. Here, in order to shed light on the origins of the Black Sea Scythians and their genetic affinity with other ancient nomadic populations of the Eurasian Steppe, we characterized the mitochondrial DNA (mtDNA) structure of a population of 16 Scythian individuals of the Rostov-on-Don area, Russia (2,200 - 2,600 yBP). Mitochondrial data from the hypervariable region I and the coding region was compared to data from both ancient and modern-day Eurasian populations using principal component analysis, classical multi-dimensional scaling and haplotype sharing analysis. The results of these analyses showed that the mitochondrial gene pool of Scythians was under diverse genetic influences. The gene pool of Scythians was characterised by the concomitant presence of lineages of both western and eastern Eurasian origins, a genetic feature shared with modern-day populations of Central Asia. We also revealed genetic influences from Siberia and the Central Asian corridor (Iraq, Iran, Pakistan, India). Previously described Iron Age populations of the Eurasian Steppe were found to display distributions of mtDNA lineages similar to those observed in contemporaneous Scythians. These mitochondrial similarities among nomadic populations indicate either a recent common origin or a significant amount of gene flow on the maternal side among ancient nomadic populations. The comparison of ancient Eurasian populations demonstrates the power of ancient DNA sampled in time and space to reconstruct past human population processes and events.

INTRODUCTION

The Scythians

Scythians are mainly known as ancient horse-riding warriors of the Black Sea area (2,700 to 2,800 years Before Present, yBP) based on descriptions by the Greek historian Herodotus (2,500 yBP). Archaeological investigation of burial mounds, called 'kurgans', has revealed the existence of nomadic populations in the steppe north of the Black Sea dated from the Bronze Age (~2,800 – 5,500 yBP) to the Iron Age (~2,300 – 2,800 yBP; Yablonsky, 2000). Some Iron Age nomad burials were archaeologically associated with the Scythians depicted by Herodotus (Yablonsky, 2000; Ricaut et al., 2004a). Artefacts harbouring similarities with the material culture of nomads from the steppe north of the Black Sea were also found in funerary monuments across the Eurasian Steppe, an 8,000 kilometre-long strip of grassland, which extends from modern-day Hungary to Mongolia (Bashilov & Yablonsky, 2000).

Many questions have arisen from the archaeological study of Scythians and other nomadic populations of the Eurasian Steppes and are still intensely discussed by archaeologists, historians, anthropologists, and linguists. The debate has mainly focused on questions relating to the origins and disappearance of the Scythians, as well as their cultural and genetic relationships to other nomadic populations of the Eurasian Steppe.

The Bronze Age in the central Eurasian Steppe

In the Bronze Age diverse nomadic populations lived occupied the Eurasian Steppe, of which the Black Sea area represents the westernmost extremity. Their subsistence strategy was a mixture of foraging activities (hunting, gathering and river fishing) and, to a lesser extent, farming economy (crop culture and animal breeding; Bashilov & Yablonsky, 2000). The two main Bronze Age cultures of Eurasian Steppe nomads were the Andronovo culture of southern Russia, modern-day Kazakhstan and western Central Asia (3,000 – 4,300 yBP) and the Timber-Grave, or Srubna, culture of the area between the northern Black Sea shore and the Volga River (present-day western Russia, 3,100 – 3,800 yBP; Koryakova & Epimakhov, 2007). Culturally similar nomadic Bronze Age material cultures were found over a large geographical area ranging from the Caucasus Mountains to south Siberia and to the south of present-day Uzbekistan, Tajikistan and Turkmenistan. Archaeological and

anthropological studies have suggested that long-distance migrations might have underpinned this wide distribution (Mandelstam, 1966; Mandelshtam, 1967; Potemkina, 1987; Pyankova, 1974; Pyankova, 1987; Yablonsky, 1996). Factors that may have enabled such migrations include the homogeneity of the ecological environment of the Eurasian Steppe and the lack of significant geographical barriers, enabling contacts between nomadic groups, as well as the emergence and spread of horse riding and of the spoke-wheeled technology around 3,000 – 4,000 yBP (Bokovenko, 2000; Renfrew, 2002; Levine, 2004; Drews, 2004; Anthony, 2007). In particular, incursions of groups and/or trade items characterized by traits belonging to the Andronovo culture were proposed to have reached the Xinjiang region in present-day north west China (Kuz'mina, 2008; Mei, 2000; Mei & Shell, 2002). There, human remains exhibiting European physical traits were found in association with non-local textiles of possible western origins (Mallory & Mair 2000; Mair 1998; Barber, 1999).

The Iron Age in the central Eurasian Steppe

Rapid and radical changes marked the transition to the Iron Age in nomadic Eurasian Steppe populations (Hanks, 2010), including Black Sea Scythians. With the advent of the Iron Age around 2,700 – 2,800 yBP, Eurasian Steppe nomads underwent an increased specialisation in livestock breeding as well as rapid developments in their politics, ideology and technology (Bashilov & Yablonsky, 2000). The first Scythian funerary monuments were excavated in the area north of the Black Sea (Bashilov & Yablonsky, 2000). Similarities in material/cultural elements of these burials were also found in the Altai (south Siberia) and led to the definition of the 'Scythian Triad': weapons, horse harnesses, and items decorated with zoomorphic elements constituting the 'Scythian Animal Art' (Grakov & Melukova, 1954). The 'Scythian Triad' was subsequently used as an archaeological 'Scythian marker' to identify cultures of the Eurasian Steppe (Bashilov & Yablonsky, 2000). The main Iron Age nomadic cultures of the Eurasian Steppe were the Tagar culture of present-day Khakassia, south Siberia (3,000 – 3,100 yBP), the Tashtyk culture of the Yenisei River, present central Russia (2,000 - 2,400 yBP), and the European Scythian culture (2,200 – 2,800 yBP; Keyser et al., 2009). Scythians inhabited a region delimited by present-day Ukraine in the West, the Black Sea and the Caucasus Mountains in the South, Kazakhstan in the East and the southern Ural in the North (Marcenko & Vinogradov, 1989).

The origins of Scythians

Two main hypotheses have been proposed for the origins of Scythians. The first hypothesis suggests an Asian origin, whereas the second advocates a local Scythian ethnogenesis involving a genetic continuity with Bronze Age nomads of the north of the Black Sea.

The Asian origin hypothesis was developed on the basis of texts by Herodotus, according to which Scythians had reached the Black Sea from Central Asia (Terenozhkin, 1971). The Asian origin hypothesis relies on two main arguments. First, the oldest burials, where harness elements typical of the Scythian culture were found, are located in Asia, not in Europe. Second, the animal patterns characteristic of the Scythian material art have not been observed in the material culture of preceding Bronze Age nomads of the Black Sea but were common elements of Late Bronze Age artefacts of nomadic groups ranging from the Yenisei River Basin (North West Siberia) to Mongolia (Bashilov & Yablonsky, 2000).

The local Scythian hypothesis proposes that Scythians originated from the Black Sea area and were genetically linked to local Bronze Age populations (Grakov, 1977). This idea was supported by palaeoanthropological studies (Debets, 1948; Alexeev, 1980). It was proposed that the cultural similarities observed among nomadic populations of the central Eurasian Steppe had been acquired by Scythians in the course of their military campaigns in the eastern part of the Steppe (Bashilov & Yablonsky, 2000). Furthermore, the reassessment of the age of the Scythian site of Arzhan in Tuva (south Siberia) provided additional support for the indigenous origin hypothesis. This site was indeed initially thought to corroborate the Asian origin hypothesis by establishing older dates for Scythian sites in the east – 2,700 – 2,900 yBP (Gryaznov, 1980) - than in the west of the Steppe – 2,600 – 2,800 yBP (Murzin, 1986). However, younger dates (2,500 – 2,700 yBP) from a more recent redating of Arzhan are currently more accepted for this site (Chlenova, 1997) removing it as providing support to the Asian origin hypothesis.

Cultural and genetic homogeneity among ancient nomads of the Eurasian Steppe

Related to the question of the Scythian origins is the broader question of the cultural and genetic homogeneity among ancient nomads of the Eurasian Steppe. It was traditionally accepted that nomads of the Eurasian Steppe constituted an ethnic and cultural unity. This idea led to the establishment, in archaeology and

anthropology, of the terms ‘Scythian World’, ‘Scytho-Siberians’ or ‘Eurasian Cultural Continuum of the Scythian Epoch’ (Raevsky, 1993) to describe these temporally and geographically distinct populations. The unifying element among ancient central Eurasian nomads was the ‘Scythian Triad’. Similarities in lifestyle and economy (nomadic pastoralists and cattle breeders) were also used to justify the homogeneity among these populations.

However, this view was challenged by further and more detailed examination of the material culture of nomads from the Eurasian Steppe. Archaeological work revealed that the apparent homogeneity of artefacts observed among various groups of the Eurasian Steppe was in fact only superficial. Elements of the ‘Scythian Triad’ were found to be temporally and spatially diverse. When examining other cultural markers such as pottery, a significant heterogeneity was observed among nomadic groups; as a consequence, they could no longer be considered one homogeneous cultural entity. The ostensible similarity in material culture was instead explained by migrations or frequent contacts among small nomadic groups (Bashilov & Yablonsky, 2000). Moreover, the study of biological features also challenged the concept of unity among populations. Significant differences in craniometric traits have been observed among Scythians of the Black Sea, i.e. between ‘steppe Scythians’ and ‘forest-steppe Scythians’ (Kozintsev, 2007).

Genetic diversity of present-day Eurasian populations

The debate about the Scythian origins and the suggested homogeneity among nomadic groups of the Eurasian Steppe, raise the questions of, first, the genetic continuity between Scythians and nomadic Bronze Age populations of the Black Sea area or Central Asia and, second, the amount of gene flow between ancient nomadic populations.

The study of uniparentally -inherited genetic markers, such as mitochondrial DNA (mtDNA, maternally) and Y-chromosome (paternally) markers sampled in modern-day populations, has previously provided means to reconstruct the origins of human groups and other aspects of human population history, for example the peopling of Europe (Richards et al., 1996; Richards et al., 1998; Richards et al., 2000; Semino et al., 2000; Rosser et al., 2000; Scozzari et al., 2001). However, such genetic approaches based on present-day genetic data alone are not sufficient to investigate the origins of Scythians and their relationship to other nomadic populations of the

Eurasian Steppe. The main reason is that modern data lack temporal resolution to answer specific archaeological questions, i.e., the genetic structure of ancient cultural groups cannot be inferred with certainty from modern genetic data. Moreover, the verification of hypotheses about the Scythian origins and the genetic homogeneity among Eurasian Steppe nomads would require the identification of potential modern-day descendants of the Scythians and other ancient nomadic groups. Even if these modern-day populations could be identified, the time discrepancy between ancient and modern-day populations (e.g., genetic drift) and dynamic population processes (e.g., population replacement) would seriously confound the reconstruction of the genetic history of Eurasian Steppe nomads.

Central Eurasia comprises a vast area ranging from the Volga region in the West to Mongolia in the East and from Afghanistan in the South to southern Siberia in the North (Figure 1D). The reconstruction of past population dynamics in central Eurasia is recognised as a particularly complex task due to the issues mentioned above. Modern-day central Eurasians are indeed characterized by a significant amount of diversity at the cultural, linguistic and genetic levels, which may reflect a complex population history. This was also evidenced by the important heterogeneity observed among populations for different genetic markers: blood groups (Fiori et al., 2000), nuclear markers (autosomal and X-linked loci in Segurel et al., 2008; microsatellites in Martinez-Cruz et al., 2011), Y-chromosome (Wells et al., 2001; Zerjal et al., 2002), mtDNA (Comas et al., 1998; Comas et al., 2004; Quintana-Murci et al., 2004; Irwin et al., 2010). At the mtDNA level, central Eurasians are characterised by a mixture of geographically well differentiated lineages considered to be of western (mtDNA haplogroups H, HV, I, J, K, T, U, V, W, and X) and eastern (mtDNA haplogroups A, B, C, D, E, F, G, Y, Z) Eurasian origin on the basis of their current distribution and diversity (e.g., Wallace et al., 1999; Ingman et al., 2000; Maca-Meyer et al., 2001; Herrnstadt et al., 2002; Mishmar et al., 2003). Studies of monoparental genetic markers (mtDNA and Y-chromosome) in Central Asia – including present-day Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan – suggest contrasting female and male genetic histories. Evidence from modern-day mtDNA tends to suggest that the mixed genetic makeup of Central Asians has been produced by the admixture of differentiated populations (Comas et al., 2004; Quintana-Murci et al., 2004) more than by the expansion of an ancestral mixed population after a 'maturation' phase as proposed on the basis of Y-chromosome data (Wells et al., 2001;

Zerjal et al., 2002). Founder effects, bottleneck events and mating isolation were proposed to have played an important role in the heterogeneity observed among modern central Eurasian populations for both the mtDNA and Y-chromosome markers. The distribution of genetic variability in central Eurasia may also have been greatly impacted by super-imposed migrations along a west-east axis over the last millennia. These migrations range from the first settlement of anatomically modern humans in Eurasia in the Palaeolithic, trade along the Silk Road (from the Mediterranean Sea to the Pacific Ocean, 600 - 2,100 yBP), prehistoric and historical movements of pastoralist nomads of the Eurasian Steppe, and military expansions. In the light of their present mtDNA diversity and complex population history, no central Eurasian population in particular can be identified as the direct descendents of the Scythians.

Ancient DNA from central Eurasia

An alternative approach to test the hypotheses about Scythian origins, and potential genetic homogeneity among ancient nomadic populations of the Eurasian Steppe, is the direct genetic characterisation of ancient Scythians and comparison with genetic data from other ancient and present-day populations. Bronze Age and Iron Age populations of central Eurasia representing various cultures have previously been analysed using ancient DNA (aDNA). The climatic conditions in parts of this region, arid and/or cold, are thought to be optimal for the preservation of bio-molecules such as DNA. As a result, the current ancient DNA sampling of central Eurasia, although still far from being comprehensive, has provided detailed insights into the temporal and geographical distribution of human mtDNA diversity in this region.

The mtDNA characterization of Bronze Age and Iron Age cultures of Kazakhs (Lalueza-Fox et al., 2004) and south Siberian Kurgans (Keyser et al., 2009) were consistent with archaeology and anthropology proposing a shift from a European influence in the Bronze Age to a mixed western/eastern Eurasian influence in the Iron Age. In the case of ancient Kazakhstan, the input of eastern Eurasian lineages was proposed to have originated in Siberia and/or Mongolia and to have reached this region during the westward Xiongnu expansions (2,600 - 2,700 yBP, Lalueza-Fox et al., 2004). A Xiongnu cemetery in the Egyin Gol Valley, Mongolia (2,200 – 2,300 yBP) was also characterised at the mtDNA level, providing data for comparison (Keyser-Tracqui et al., 2003). In the Altai, ‘western’ lineages were found in Bronze

Age individuals (Chikisheva et al., 2007) and both ‘western’ and ‘eastern’ lineages were observed in Iron Age individuals of the Pazyryk culture (Ricault et al., 2004a; Ricault et al., 2004b; Chikisheva et al., 2007; Pilipenko et al., 2010). The aDNA study of individuals of the Sargat culture (1,500 - 2,500 yBP; Bennett & Kaestle, 2010) suggested that western Siberia was the westernmost limit for the spread of eastern mtDNA lineages in the Iron Age.

Here, we expand the knowledge of the mtDNA diversity within nomadic populations of the Eurasian Steppe by generating 15 new mtDNA hypervariable region I (HVR-I) sequences from Scythian individuals. These individuals were collected from sites in the Rostov-on-Don area, Russia, and were estimated to range in age from around 2,200 to 2,600 yBP based on cultural affinities and archaeological background. The first objective of this study was 1) to describe the mtDNA structure of Scythians; then, 2) to compare the data with modern-day Eurasian mtDNA data to localise the current distribution - in terms of frequency and diversity - of lineages detected in Scythians and identify genetic affinities, and lastly, 3) to use mtDNA data from pre-Iron Age and Iron Age populations of Eurasia to search for direct evidence for Scythian origins. Similarly, the genetic homogeneity among nomadic populations of the Eurasian Steppe was examined by comparing Scythian mtDNA sequences with data from contemporaneous Iron Age populations of Eurasia.

MATERIAL AND METHODS

Sample description and archaeological context

The Southern Research Centre of the Russian Academy of Sciences (anthropologist Elena Batieva), Russia, provided a total of 34 human teeth representing 17 individuals for the present study. The samples were collected from nine archaeological sites located in the vicinity of Rostov-on-Don, Rostov Oblast, south west Russia (Figure S1). Rostov-on-Don is situated in the south eastern part of the East European Plain, on the right bank of the Don River, around 46 km North of the Black Sea.

Sample preparation and DNA extraction

Ancient DNA extraction and amplification was carried out at the Australian Centre for Ancient DNA (ACAD), University of Adelaide, Australia. In order to control for potential DNA contaminants arising from ACAD experiments, samples from individuals RD-3 and RD-12 were sent to the University of Pompeu Fabra (UPF), Barcelona, Spain (Carles Lalueza-Fox, Oscar Ramirez) for independent replication of the results.

The protocols used for sample decontamination, preparation, digestion and silica-based DNA isolation were identical as those described in the 'Material and Methods' section of Chapter Two.

Hypervariable-Region I sequencing and coding region GenoCore22 typing

The HVR-I sequencing (positions 15997 to 16409) and GenoCore22 typing of coding region Single Nucleotide Polymorphisms (SNPs) were performed according to the same protocols as those described in Chapter One. Results of the GenoCore22 typing can be found in Table S1.

Cloning

Two PCR products obtained for individuals RD-3 and RD-12 were cloned at UPF. The TOPO-TA cloning kit (Invitrogen®) was used and the protocol described in Lalueza-Fox et al., 2007.

Quantitative Real-Time PCR

Quantitative Real-Time PCR experiments were carried out in the same conditions as those described in Chapter One. Results of DNA quantification can be found in Table S2.

To compare the copy-number of the 133 bp (L16209/H16303) and the 179 bp (L16209/H16348) fragments, the Shapiro-Wilk W test was first used to verify that the number of copies for each fragment followed a normal distribution ($p = 0.2215$ for the L16209/H16303 short fragment and $p = 0.5381$ for the long L16209/H16348 fragment). A significantly larger number of copies for the shorter compared to the larger fragment was statistically confirmed by a one-tailed paired t-test ($p = 0.04337$) in R version 2.12 (R Development Core Team, <http://www.R-project.org>). This result

is in accordance with relatively low levels of modern DNA contaminants on the aDNA extracts produced in this study.

Authentication of the mtDNA data

The possibility that sequences obtained from ancient remains were not impacted by contamination or post-mortem damage can never be totally dismissed. However, strict authenticity criteria were followed in order to monitor contamination and artefactual mutations caused by post-mortem DNA damage (Pääbo et al., 2004). Here, the authenticity of the sequences was assessed according to the seven authentication criteria presented in the 'Material and Methods' section of Chapter One. Low level of laboratory contamination was assessed by the replication of sequences for individual RD-12 sent to UPF for independent replication. The discrepancies observed between sequences obtained at UPF and ACAD for individual RD-3 are discussed in the text. The sequences of the clones obtained at UPF for individuals RD-3 and RD-12 are reported in Figure S2.

Ancient populations used in comparative analyses

Mitochondrial DNA data generated from the Scythians of the Rostov-on-Don area (Table S3) was compared to data obtained from other ancient as well as modern-day populations.

The Scythians were compared to ancient populations of central Eurasia from Kazakhstan, south Siberia and the Altai Mountains previously reported in the literature. Genetic homogeneity over a very large geographical area was tested for these mobile ancient populations. Ancient populations located at the periphery of the Eurasian Steppe, in Mongolia, the peri-Baikal area (South East Siberia) and central/eastern Europe, were also included in the comparison. These were ancient populations that were not culturally associated with Steppe nomads or the so-called 'Scythian World'. Previous aDNA work having shown genetic shifts in central Eurasia occurring during the Iron Age, ancient populations were segregated into two groups for analysis: pre-Iron Age and Iron Age/post-Iron Age populations. The amount of ancient genetic data collected from central Eurasia was large enough to allow this temporal segregation (Table 1).

Table 1. Description of the ancient populations used for comparison.

Description	Abbreviation	Reference	N	Date range (yBP ^a)
PRE-IRON AGE				
Central and East European hunter-gatherers	HG	Bramanti et al., 2009; Krause et al., 2010	22	4,250 - 30,000
Scandinavian Pitted-Ware Culture individuals	PWC	Malmstrom et al., 2009	19	4,500 - 5,300
Bronze Age individuals of Kazakhstan	KAZ-BA	Lalueza-Fox et al., 2004	13	2,700 - 3,400
Bronze Age Kurgan individuals of South Siberia	KUR-BA	Keyser et al., 2009	11	2,800 - 3,800
Bronze Age individuals of the Altai	ALT-BA	Chikisheva et al., 2007	3	3,500 - 4,000
Bronze-Age individuals of the Tarim Basin	TAR	Li et al., 2010	20	3,980
Kitoid Neolithic of Lake Baikal	LOK	Mooder et al., 2005	30	6,130 - 7,140
IRON AGE				
Iron Age individuals of Kazakhstan	KAZ-IA	Lalueza-Fox et al., 2004	12	2,100 - 2,800
Iron Age Kurgan individuals of South Siberia	KUR-IA	Keyser et al., 2009	15	1,600 - 2,800
Pazyryk culture individuals of the Altai	ALT-IA	Ricaut et al, 2004a; Ricaut et al, 2004b; Chikisheva et al., 2007; Pilipenko et al., 2010	11	2,300 - 2,500
Sargat individuals of South West Siberia	SAR	Benett & Kaestler, 2010	5	1,500 - 2,500
Xiongnu individuals of the Egyin Gol Valley, Mongolia	EG	Keyser-Tracqui et al., 2003	46	2,200 - 2,300

^ayears Before Present

The pre-Iron Age (4,500 - 30,000 yBP) dataset was composed of the following populations (Figure 1A): Palaeolithic/Mesolithic hunter-gatherers of central/eastern Europe (Bramanti et al., 2009; Krause et al., 2010), Scandinavian Pitted-Ware Culture individuals (Malmström et al., 2009), Bronze Age individuals of Kazakhstan (Lalueza-Fox et al., 2004), Bronze Age Kurgan nomads of the Andronovo and Karasuk cultures (south central Siberia; Keyser et al., 2009), nomads of the Altai Karakol culture (south central Siberia; Chikisheva et al., 2007), Bronze Age individuals of the Xiaohe tomb complex of the Tarim Basin (north west of present-day China; Li et al., 2010), and Neolithic Kitoid individuals of the peri-Baikal area (south east Siberia; Mooder et al., 2005).

The populations included in the Iron Age/post Iron Age (1,500 -2,800 yBP) dataset were (Figure 1B): Iron Age individuals of Kazakhstan (Lalueza-Fox et al., 2004), Iron Age Kurgan individuals of the Tagar and Tachtyk cultures (south central Siberia; Keyser et al., 2009), Iron Age individuals of the Altai Pazyryk culture (Ricaut et al., 2004a; Ricaut et al., 2004b; Chikisheva et al., 2007; Pilipenko et al; 2010), confederated nomadic tribes of the Xiongnu period of Mongolia (Keyser-Tracqui et

al., 2003), individuals of the Sargat culture of west Siberia (Bennett & Kaestle, 2010), and the Scythians of the Rostov area presented here.

Other ancient genetic data for prehistoric European populations was available, notably for early Neolithic (e.g., Haak et al., 2005; Haak et al., 2010; Sampietro et al., 2007; Deguilloux et al., 2010) and subsequent Bronze Age/Iron Age populations (e.g., Caramelli et al., 2007; Melchior et al., 2010). They were not included in the present analyses because of their peripheral location compared to ancient Central Eurasians and the fact that their genetic structure might reflect demographic events related to the Neolithic transition in Europe (6,000 – 10,000 yBP). The Neolithisation having possibly involved a significant gene flow from the Near East to Europe (Haak et al., 2010), including these populations in the comparative dataset would have added a level of complexity that was not relevant to the questions discussed here.

Another temporal dimension was added with a third dataset containing mtDNA data from modern-day populations extensively sampled over Eurasia. These three datasets allowed the examination of the genetic relationships within contemporaneous populations and of genetic changes through time by comparison between time-periods.

Present-day populations used in comparative analyses

Data for extant populations were compiled in the MURKA mtDNA database and integrated software, which currently contains 168,000 HVR-I records from published studies and curated by Oleg Balanovsky, Valery Zaporozhchenko and Elena Balanovska of the Russian Academy of Medical Sciences. A sub-sample of 97 ancient and modern-day Eurasian populations (~60,350 individuals) was used for haplogroup-based comparative analysis. Names of modern-day populations were abbreviated using ISO codes in capital letters, and in small letters when ISO codes were not available. Unless specified otherwise, the same population codes were used for all plots, maps and analyses in this study, i.e. PCA, MDS, Slatkin's fixation index F_{ST} calculations. On the PCA plot (Figure 2), populations were separated into six color-coded groups according to their age and geographical location. The Scythians ancient mtDNA in this study are represented in red and abbreviated as SCY. Populations previously sampled for ancient mtDNA are indicated in black and references are given in Table 1. Grey symbolises modern-day populations of the Near East and the Caucasus regions, which are indicated by: ARM, Armenia; AZE, Azerbaijan; IRN, Iran; IRQ, Iraq; JOR, Jordania; kab, Kabardians; KAZ, Kazakhstan; kur, Kurds; nog,

Nogays; PSE, Palestine; SE, Ossets; SAU, Saudi Arabia; SYR, Syria; TUR, Turkey. Yellow corresponds to modern-day populations of north east Europe and are referenced as: ALB, Albania; AUT, Austria; aro, Arorums; bas, Basques; BEL, Belarus; BGR, Bulgaria; BIH, Bosnia; CHE, Switzerland; CU, Chuvash; CYP, Cyprus; CZE, Czech Republic; DEU, Germany; ESP, Spain; EST, Estonia; FIN, Finland; FRA, France; GBR, United-Kingdom; GEO, Georgia; GRC, Greece; HRV, Croatia; HUN, Hungary; IRL, Ireland; ISL, Iceland; IT-88, Sardinia; ITA, Italy; KO, Komis; KR, Karelians; LTU, Lithuania; LVA, Latvia; ME, Maris; MO, Mordvinians; NOR, Norway; POL, Poland; PRT, Portugal; ROU, Romania; RUS, Russia; SVK, Slovakia; SVN, Slovenia; SWE, Sweden; TA, Tatars; UD, Udmurts; UKR, Ukraine. Present-day populations of eastern Eurasia and western/central Siberia are represented in blue and abbreviated as follows: ale, Aleuts; alt, Altaians; BA, Bashkirs; BU, Buryats; CHU, Chukchi; esk, Eskimos; eve, Evenks; evn, Evens; ket, Kets; KK, Khakhassians; kham, Khamnigans; khan, Khants; kor, Koryaks; man, Mansi; MNG, Mongolians; NEN, Nenets; nga, Nganasans; niv, Nivkhs; SA, Yakuts; sel, Selkups; sho, Shors; Tel, Telenghits; tof, Tofalars; tub, Tubalars; tuv, Tuvinians; ulc, Ulchi; yuk, Yukaghirs. Modern-day populations of Central Asia were represented in brown and abbreviated as follows: buk, Bukharan Arabs; dun-haz, Dungans-Hazaras; TJK, Tajiks; TKM-kar, Turkmen- Karakalpak; kal, Kalash; kho, Khoreman Uzbek; shu, Shugnan; UZB, Uzbeks. See description and references of the populations in Supplementary Table S4.

Map of haplogroup frequencies

Haplogroup frequencies in both ancient and modern-day populations were summed into pools according to their present distribution in western Eurasia (haplogroups H, HV, I, J, K, T, U, V, W, X), eastern Eurasia (haplogroups A, B, C, D, F, G, Y, Z) or ‘other’ (L, M*, N*, N1). Percentages of ‘western’, ‘eastern’ and ‘other’ haplogroups were represented by a bar plot placed on a map of Eurasia at the average geographical position of the corresponding population. Mapping of haplogroup pool frequencies offers a very crude representation of the distribution of the mtDNA diversity, in part because some haplogroups do not display strictly restricted distributions, e.g. U4, which is found in western Europe and at high frequencies in populations of western Siberia (Malyarchuk et al., 2004). However, this approach

allows a rapid and global visualization of the distribution of mtDNA lineages at the scale of Eurasia as well as major genetic shifts through time.

Principal Component Analysis (PCA)

In order to compare the population of Scythians with modern-day Eurasians and ancient populations in more resolution than provided by frequencies of ‘western’ and ‘eastern’ haplogroups, two Principal Component Analyses (PCA) were performed for frequencies of 21 haplogroups: A, C, D, F, H, HV, I, J, K, N1, T, U2, U4, U5, U7, V, W, X. Frequencies of the five ‘east Eurasian’ haplogroups were pooled into the ‘EAS’ group: B, E, G, Y, Z. Frequencies of haplogroups U1, U6, U8 were pooled in ‘Uother’ and finally frequencies in three haplogroups found at lower frequencies in Eurasia and pooled into the ‘other’ group: L, M*, N*. Pooling and removal of rare haplogroups (with frequencies below 1%) allowed statistical noise to be reduced. On the PCA plots (Figure 2), a yellow circle delimited the positions of the following European populations: aro, AUT, BIH, BEL, CHE, CU, CZE, DEU, EST, FRA, GBR, HUN, HRV, ISL, LTU, LVA, NOR, POL, ROU, SVK, SWE, UKR. The first PCA compared the Scythians with modern-day Eurasian and pre-Iron Age populations (4,500 – 30,000 yBP). The second PCA compared the Scythians with modern-day Eurasian populations and Iron Age/post-Iron Age populations (1,500 – 2,800 yBP).

Haplotype-based analyses of the mtDNA data

In order to confirm the genetic affinities between Scythians and other ancient and modern-day populations that have emerged from the analysis of mtDNA haplogroup data with more resolution, haplotype-based analyses were performed: Network (Bandelt et al., 1995), Multidimensional Scaling (MDS) of Slatkin’s fixation indexes (F_{ST} , Slatkin, 1995), Analysis of the Molecular Variance (AMOVA; Excoffier, 1992), for ancient mtDNA data; and haplotype sharing for both ancient and modern mtDNA data.

Fixation index (F_{ST}) calculations and Analysis of the Molecular Variance (AMOVA)

Analyses of haplogroup frequency data through calculation of Slatkin’s F_{ST} (Slatkin et al., 1995) and AMOVA were performed using the software Arlequin version 3.11 (Excoffier, Laval & Schneider, 2005).

Classical Multi-Dimensional Scaling (MDS)

Classical MDS was performed using F_{ST} values calculated between ancient populations. PCA and MDS were carried out using in house scripts written in R version 2.12 (R Development Core Team, <http://www.R-project.org>) and using the 'prcomp' function for PCA.

Haplotype-sharing analysis

A database containing 168,000 HVR-I haplotypes modern-day populations (Oleg Balanvosky, Elena Balanovska) was searched for sequences identical to those obtained for Scythians (Table S5). The T1a 16126C-16163G-16186T-16189C-16294T (individuals RD-6 and RD-11) and H revised Cambridge reference sequence (rCRS; individual RD-9) haplotypes were excluded from this search because of their very wide distribution that renders them uninformative in terms of genetic affinities. Populations for which a match could be found were pooled into 14 main groups according to their geographical location: Near East (including Iraq, Iran, Saudi Arabia, Uzbekistan), Middle East (including Turkey, Palestine, Jordania), Caucasus, western Europe, Balkans, eastern Europe, Volga-Ural, western Siberia, eastern Siberia, Central Asia, east Asia, Far east Asia, south east Asia, southern Asia (see Table S5 for population allocation within geographical groups).

First, for each population pool, the number of haplotypes in these populations found to be similar to one of the informative Scythian haplotypes was divided by the population size of the corresponding pool. Second, in order to allow a better visualisation of population affinities, the pools of populations used in the previous analysis were pooled further: Near East, Middle East, Caucasus into the 'Near East/Caucasus' pool, west Europe, Balkans into the 'west Europe' pool, east Europe, Volga-Ural into the 'east Europe' pool, west Siberia, east Siberia, into the 'Siberia' pool, east Asia, Far east Asia, south east Asia, south Asia into the 'east Eurasia' pool. For each Scythian, the number of individuals sharing the same haplotype in the population pool was divided by the corresponding population size of the pool. The results of the haplotype sharing analyses were represented as percentages of shared haplotypes, in bar plots constructed in R version 2.12.

RESULTS

Success rate for the amplification of authenticated ancient mtDNA

A high amplification success rate was obtained for the present Scythian sample set, with 16 individuals out of 17 (94.1%) yielding reliable mtDNA sequence data. The high amplification success rate was in accordance with the good macroscopic preservation of the Scythian samples (Figure S1).

One individual (RD-4) was excluded from the analysis because several authentication criteria could not be met. Amplification success for this sample was low and the few sequences were ambiguous and not replicable.

Problems associated with the independent replication of one ancient mtDNA haplotype

Two samples from individuals RD-3 and RD-12 were sent for independent replication to an aDNA laboratory in Barcelona (UPF). For individual RD-12, the sequences obtained in both laboratories matched. For RD-3, the replication of the sequencing yielded different haplotypes. The RD-3 haplotype obtained at UPF was 16189C-16270T-16274A and belongs to haplogroup U5b1 whereas the RD-3 haplotype obtained at the ACAD was 16256T-16318t and belongs to haplogroup U7.

Although we cannot entirely exclude contamination, we believe that appropriate measures have been taken to reduce the risk of contamination in both laboratories for five reasons.

1) Contamination originating from personnel working in both ancient DNA laboratories was ruled out, as none of their HVR-I sequences matched any of the RD-3 haplotypes.

2) The RD-3 sample extracted initially at the ACAD was re-extracted independently in the same laboratory by another operator (Bastien Llamas), and this extraction led to the same U7 haplotype as previously obtained. In addition, replication of the sequences for the 15 other Scythian individuals within the ACAD was successful. Independent extractions carried out on different days and simultaneously with samples from different geographical regions indeed yielded identical haplotypes for the 15 other Scythian individuals.

3) The U7 haplotype obtained for individual RD-3 did not match any haplotype sequenced at the ACAD during the last four years. As a consequence, the

possibility of cross-contamination can be considered unlikely, as well as the possibility of sample exchange between sample sets. Similarly, the RD-3 U5b1 haplotype obtained at the UPF could not be found in the database of all sequences obtained at the ACAD and in particular for Scythian individuals, thus providing additional evidence against mix-up among sample sets.

4) The sequences of the clones obtained at the UPF laboratory did not provide any evidence that contamination or DNA damage could be the cause of the mismatch between the RD-3 haplotypes obtained from the two different laboratories (Figure S2).

5) In further support for the low-contamination conditions in both laboratories, no extraction or PCR control ever yielded amplification products.

In situations where independent replication fails to confirm ancient sequences, the likelihood that one or the other non-matching haplotype represents the ‘real’ haplotype can sometimes be estimated. This is the case when one or both haplotypes display a restricted distribution in the modern-day human population that makes the presence of the haplotype in the studied sample set more or less likely given the context of the archaeological find. Here, the observation of both the U5b1 and U7 haplotypes in Scythians of the Black Sea area was not irrelevant with regard to the human population history in this region. When searching the comparative database of modern haplotypes, matches closely related to the U5b sequence were found in Spain and the Canary Islands (ten haplotypes), in north Africa (two haplotypes) and in France (one haplotype). Close matches to the U7 haplotype had a very different distribution as they were found in Iraq (three haplotypes), Iran (two haplotypes), India (one haplotype) and Pakistan (one haplotype). Although the U5b haplotype displays a modern-day distribution that encompasses the location of the laboratory (Barcelona) where it has been extracted, it could not be fully proven that it originated from a modern contaminant. The location, in eastern Europe, of the Scythian burials that yielded the samples under examination was not inconsistent with the finding of both the U5b and U7 haplotypes in the Scythian gene pool.

It was possible that one or even both samples were contaminated before their arrival at the ACAD and the UPF. Another possibility was that the two samples provided for individual RD-3 were not collected from the same individual by archaeologists and anthropologists. In order to test these hypotheses, a third sample from the RD-3 individual was later obtained from the Russian anthropologists of the Southern Research Centre of the Russian Academy of Sciences (Elena Batieva) and

sent to the UPF laboratory for replication. Extraction and re-amplification of the third sample confirmed the U7 haplotype 16256T-16318t initially obtained at the ACAD. The U5b haplotype 16189C-16270T-16274A originally obtained at the UPF could not be retrieved and was consequently removed from further analyses. The U7 haplotype 16256T-16318t was obtained twice for individual RD-3 in independent laboratories, therefore it was kept for further analyses.

Scythian sample set used in this study

In this study, we analysed a dataset composed of 16 haplotypes: 14 sequences replicated internally at the ACAD (using two independent samples from each individual) and two sequences replicated in an independent laboratory (individuals RD-3 and RD-12; Table 2).

Table 2. Result overview for ancient mitochondrial DNA typing.

Samples	Chronology ^a	HVRI sequence (np 15,997-16,409) ^b 16,000+	Hg ^c (HVRI)	Hg ^d (coding region)	Analyses ^e
RD-1	Scythian, 2,300 – 2,400 yBP	189C-232A-249C-304C-311C	F1b	R9	E(3), Q
RD-2	Scythian, 2,300 – 2,400 yBP	223T-258G-298C-327T	C	C	E(2)
RD-3	Sarmatian	256T-318t	U7	U	E(2), I, Q
RD-5	Scythian, 2,400 – 2,600 yBP	192T-256T-263C-270T-399G	U5a	U	E(2)
RD-6	Scythian	126C-163G-186T-189C-294T	T1a	T	E(2)
RD-7	Scythian, 2,500 – 2,600 yBP	078G-126C-294T-296T	T2	T	E(2)
RD-8	Scythian, 2,300 – 2,500 yBP	223T-290T-319A-362C	A4	A	E(2)
RD-9	Scythian	rCRS	H	H	E(2), Q
RD-10	Scythian	354T	H2a1	H	E(2)
RD-11	Scythian	126C-163G-186T-189C-294T	T1a	T	E(2)
RD-12	Scythian	051G-129C-189C-362C	U2e	U	E(2), I, C(8)
RD-13	Scythian, 2,400 – 2,300 yBP	223T-362C	D	D	E(2)
RD-14	Scythian, 2,200 – 2,500 yBP	086C-129A-223T-391A	I3	I	E(2)
RD-15	Scythian, 2,200 – 2,400 yBP	086C-129A-223T-391A	I3	I	E(2)
RD-16	Scythian, 2,200 – 2,400 yBP	192T-256T-270T-311C	U5a	U	E(2)
RD-17	Scythian, 2,300 – 2,400 yBP	223T-239T-243C-319A-362C	D4b1	D	E(2), Q

^aDates are estimated on the basis on the archaeological artefacts associated with the remains in year Before Present (yBP). ^bVariable nucleotide positions (np) when compared to the revised Cambridge Reference Sequence (rCRS, Andrews et al., 1999). Transitions are reported with upper-case letters, transversions with lower-case letters. ^cHaplogroup (Hg) assigned on the basis of Hypervariable Region I (HVRI). ^dHaplogroup (Hg) determined by the coding-region GenoCoRe22 assay. ^eE(), number of samples from which DNA was independently extracted; I, results replicated in an independent laboratory; C() number of HVRI clones; Q, HVRI DNA quantification performed.

Mitochondrial haplogroup structure of the Scythians and comparison with modern-day populations of Eurasia

Mitochondrial haplogroups detected in Scythians were distributed into ‘western’ and ‘eastern’ haplogroups as follows: 62.4% ‘western’ (12.4% H, 12.4% I, 18.9% T, 6.3% U2, 12.4% U5), 31.3% ‘eastern’ (6.3% A, 6.3% C, 12.4% D, 6.3% F) and 6.3% ‘other’ (U7). Distributions of ‘western’ and ‘eastern’ lineages similar to the Scythian mtDNA makeup were found in modern populations of Bukharan Arabs and Khoremiens of Uzbekistan, both located in Central Asia, as well as in Udmurts of east Europe (Figure 1C; Table 3). When analysed with PCA, the mtDNA haplogroup frequencies observed in Scythians (SCY) fell close to modern-day Europeans on the PCA biplots (Figure 2). Scythians were found at the periphery of the homogeneous cluster of western/central Europeans, close to populations of eastern Europeans such as Tatars (TA2), Pomors (pom), Udmurts (UD) of western Russia, as well as Bulgarians (BGR) and Albanians (ALB) of the Balkans. The position of the Scythians on the biplot was also close to present-day Central Asian Shugnans of Tajikistan (shu), which are characterised by a high frequency of ‘western Eurasian’ haplogroups, in particular H (29.4%).

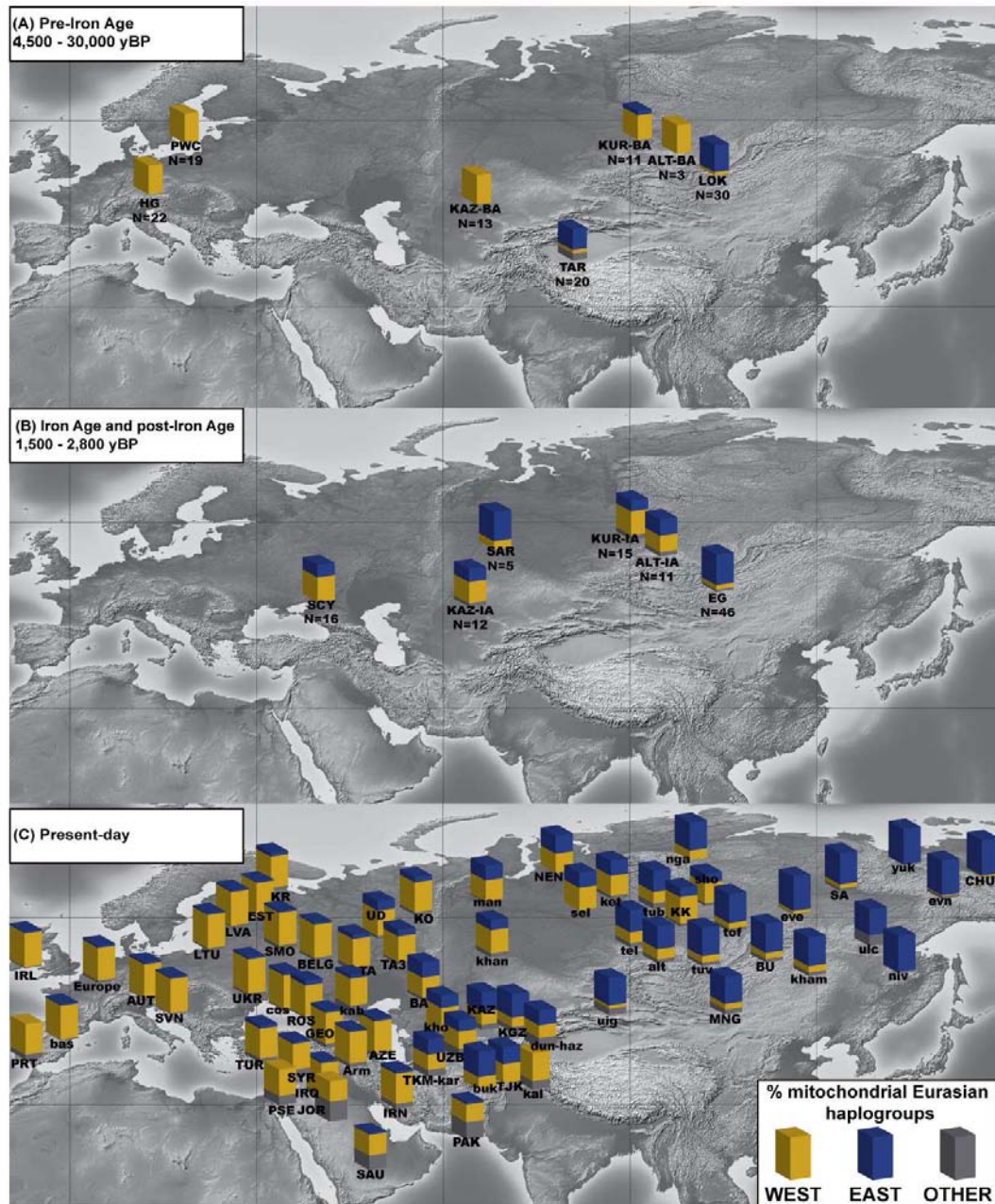


Figure 1. Maps showing the distribution of Western (yellow), Eastern (blue) and other (grey) mitochondrial haplogroups in pre-Iron Age (4,500 yBP – 30,000 yBP; A), Iron Age and post-Iron Age (1,500 – 2,800 yBP; B), present-day (C) Eurasian populations, as well as the location of central Eurasia, Central Asia and the central Asian corridor (D). See Material and Methods for abbreviations of populations.

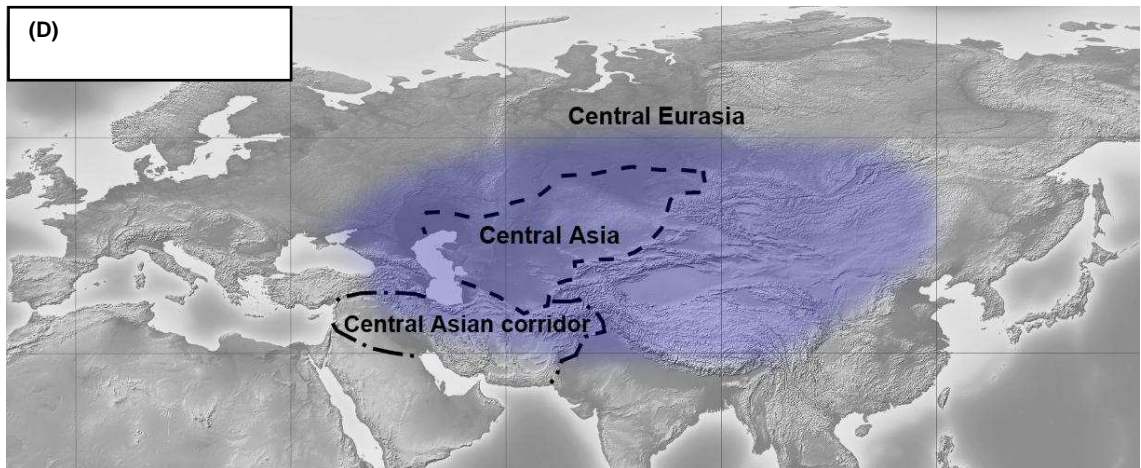


Figure 1 (continued). Maps showing the distribution of Western (yellow), Eastern (blue) and other (grey) mitochondrial haplogroups in pre-Iron Age (4,500 yBP – 30,000 yBP; A), Iron Age and post-Iron Age (1,500 – 2,800 yBP; B), present-day (C) Eurasian populations, as well as the location of central Eurasia (blue), Central Asia and the central Asian corridor (D). See Material and Methods for abbreviations of populations.

Table 3. Distribution of Western, Eastern and other mitochondrial haplogroups in ancient and modern Eurasian populations.

	Abbreviation	Population ^a	N	Frequency of mitochondrial haplogroup from:		
				West Eurasia ^b	East Eurasia ^c	other ^d
EUROPE	aro	Aromuns	133	97.0	0.0	3.0
	AUT	Austrians	117	98.4	0.8	0.8
	BA	Bashkirs	207	55.1	39.3	5.6
	bas	Basques	106	97.2	0.0	2.8
	belg	Belgorod Russians	148	96.1	0.6	3.3
	CHE	Swiss	230	97.8	0.4	1.8
	cos	Cossacks (Russia)	132	95.6	2.2	2.2
	EST	Estonians	662	97.5	0.2	2.6
	Europe	European pool ^e	133	93.8	1.7	4.2
	IRL	Irish	300	97.9	0.3	1.7
	ISL	Icelanders	448	98.7	0.6	0.6
	KO	Komis	127	84.3	11.1	4.8
	KR	Karelians	305	87.2	6.0	6.9
	LTU	Lithuanians	180	97.7	0.6	1.7
	LVA	Lativians	413	97.8	0.5	1.7
	pom	Pomors (Russia)	81	93.7	1.2	4.9
	PRT	Portuguese	848	86.3	0.0	11.1
	ros	Rostov Russians	111	88.2	9.0	2.7
	SMO	Smolensk Russians	147	93.3	0.7	6.1
	SVN	Slovens	233	98.8	0.4	0.9
	TA	Tatars ^e	197	82.2	8.9	8.8
	TA2	Tatars ^g	225	80.6	14.7	4.5
	UD	Udmurts	109	70.7	23.0	6.4
	UKR	Ukrainians	610	96.7	1.3	2.0
NEAR EAST - CAUCASUS	ARM	Armenians	192	90.5	0.5	8.8
	GEO	Georgians	158	89.4	1.9	8.9
	IRN	Iranians	517	85.5	3.6	10.9
	IRQ	Irakians	168	78.1	0.6	20.3
	JOR	Jordanians	146	72.6	0.0	27.4
	kab	Kabardians	163	79.0	5.6	15.3
	kur	Kurds	73	89.0	2.7	8.2
	nog	Nogays	206	57.3	38.3	4.4
	PSE	Palestinians	117	75.3	0.0	23.9
	SAU	Saudians	325	58.7	0.3	39.3
	SE_N	North Ossets	106	88.6	7.5	3.7
	SE_S	South Ossets	183	91.3	2.2	6.6
	SYR	Syrians	169	74.0	1.2	22.5
	TUR	Turks	608	87.6	4.5	7.9
CENTRAL ASIA	buk	Bukharan Arabs	20	70.0	30.0	0.0
	dun_haz	Dungans - Hazara	39	35.8	35.8	28.2
	kal	Kalash	44	77.2	0.0	22.7
	KAZ	Kazakhhs	125	35.2	51.2	13.6
	KGZ	Kyrgyz	20	25.0	65.0	10.0
	kho	Khoremians	20	70.0	30.0	0.0
	PAK	Pakistani	366	49.4	1.9	48.6

Table 3 (continued). Distribution of Western, Eastern and other mitochondrial haplogroups in ancient and modern Eurasian populations.

	Abbreviation	Population ^a	N	Frequency of mitochondrial haplogroup from:		
				West Eurasia ^b	East Eurasia ^c	other ^d
CENTRAL ASIA	PAK	Pakistani	366	49.4	1.9	48.6
	shu	Shugnans	44	77.1	20.5	2.3
	TJK	Tajiks	20	55.0	45.0	0.0
	TKM-kar	Turkmens-Karakalpak	40	42.5	40.0	17.5
	UZB	Uzbeks	62	50.1	27.4	22.6
	ale	Aleuts	199	0.0	100.0	0.0
	alt	Altaians	90	28.9	62.2	8.9
	BU	Buryats	411	14.2	79.0	6.5
	CHU	Chukchi	262	0.4	99.7	0.0
	esk	Eskimos	85	0.3	99.5	0.2
	eve	Evens	100	4.9	94.7	0.3
	evn	Evenks	307	1.0	96.0	3.0
CENTRAL NORTH – EAST EURASIA	ket	Kets	104	54.9	38.4	6.7
	kham	Khamnigans	99	14.1	78.8	7.1
	khan	Khants	318	66.0	23.6	10.3
	KK	Khakassians	110	20.9	76.3	2.7
	kor	Koryaks	147	0.0	100.0	0.0
	man	Mansi	161	60.2	37.8	1.9
	MNG	Mongolians	262	11.7	74.1	14.2
	NEN	Nenets	207	51.8	48.3	0.0
	nga	Nganansans	118	22.7	77.0	0.0
	niv	Nivakhs	113	0.0	99.2	0.9
	SA	Yakuts	770	9.9	87.5	2.6
	sel	Selkups	120	61.6	38.3	0.0
	sho	Shors	82	23.1	73.2	3.7
	tel	Telenghits	71	29.5	59.1	11.2
	tof	Tofalars	104	11.5	80.7	7.7
	tub	Tubalars	645	30.7	56.9	12.5
	tuv	Tuvinians	72	15.2	78.7	6.2
	uig	Uighurs	16	6.3	75.0	18.8
	ulc	Ulchi	166	0.0	72.3	27.7
	yuk	Yukaghirs	153	0.0	100.0	0.0

^asee supplementary materials for description of populations and corresponding references (Table S4); ^bsum of haplogroups H, HV, I, J, K, T, U, V, W, X frequencies; ^csum of haplogroups L, M*, N1, R frequencies; ^dsum of haplogroups A, B, C, D, E, F, G, Y, Z frequencies; ^esee supplementary materials for the definition of the European pool and references (Table S4); ^fMalyarchuk et al., 2010, ^gBermisheva et al., 2002.

Mitochondrial haplogroup structure of Iron Age populations of Eurasia

Both ‘western’ and ‘eastern’ haplogroups were observed in all Iron Age/post-Iron Age populations of central Eurasia (SCY, KAZ-IA, KUR-IA, ALT-IA, SAR, EG; Figure 1B). However, PCA of haplogroup frequencies showed that this shared feature was not indicative of genetic homogeneity among Iron Age populations when haplogroup frequencies data was analysed with more resolution by PCA (Figure 2B). The Scythians did not show particular genetic affinities for any of the contemporaneous populations of Eurasia on the PCA biplot, which, unlike Scythians, did not cluster with present-day European populations. Altaians (ALT-IA), Kazakhs (KAZ-IA) and Kurgans (KUR-IA) of the Iron Age instead shared an increase in frequency of ‘eastern’ haplogroups with modern-day populations of eastern Eurasia and Central Asia.

Haplotype-based analyses

Compared to haplogroup frequency-based analyses, examination of haplotypic data (haplotype sharing and networks) provided a more resolving power to assess the Scythian population affinities with both present-day and ancient Eurasian populations (Figures 3, 4, 5).

Haplotype sharing analyses found no exact match in the comparative modern-day and ancient populations for two out of twelve haplotypes retrieved from Scythians: haplotype U5a 16192T-16256T-16263C-16270T-16399G (individual RD-5) and haplotype U7 16256T-16318t (individual RD-3). For haplotype U7, one and two-step derivatives were found to be restricted to present-day Iran (Nasidze et al., 2006), Iraq (Richards et al., 2000; Al-Zahery et al., 2003; Behar et al., 2008), India (Kivisild et al., 1999) and Pakistan (Cordaux et al., 2003), thus showing a genetic affinity of Scythians with the Central Asian corridor.

The exact matches found for the ten remaining informative haplotypes detected in Scythians displayed a very wide distribution across present-day Eurasia (Figure 3A). The percentages of matches increased along a west to east axis (1.25% – 2.01% in the Near/Middle East and 1.28% – 3.51% in Europe), reached maximal values in present-day populations of Central Asia (7.12%) and east Asia (8.44%), and decreased further east (1.74% – 3.09% in east/south-east/south Asia).

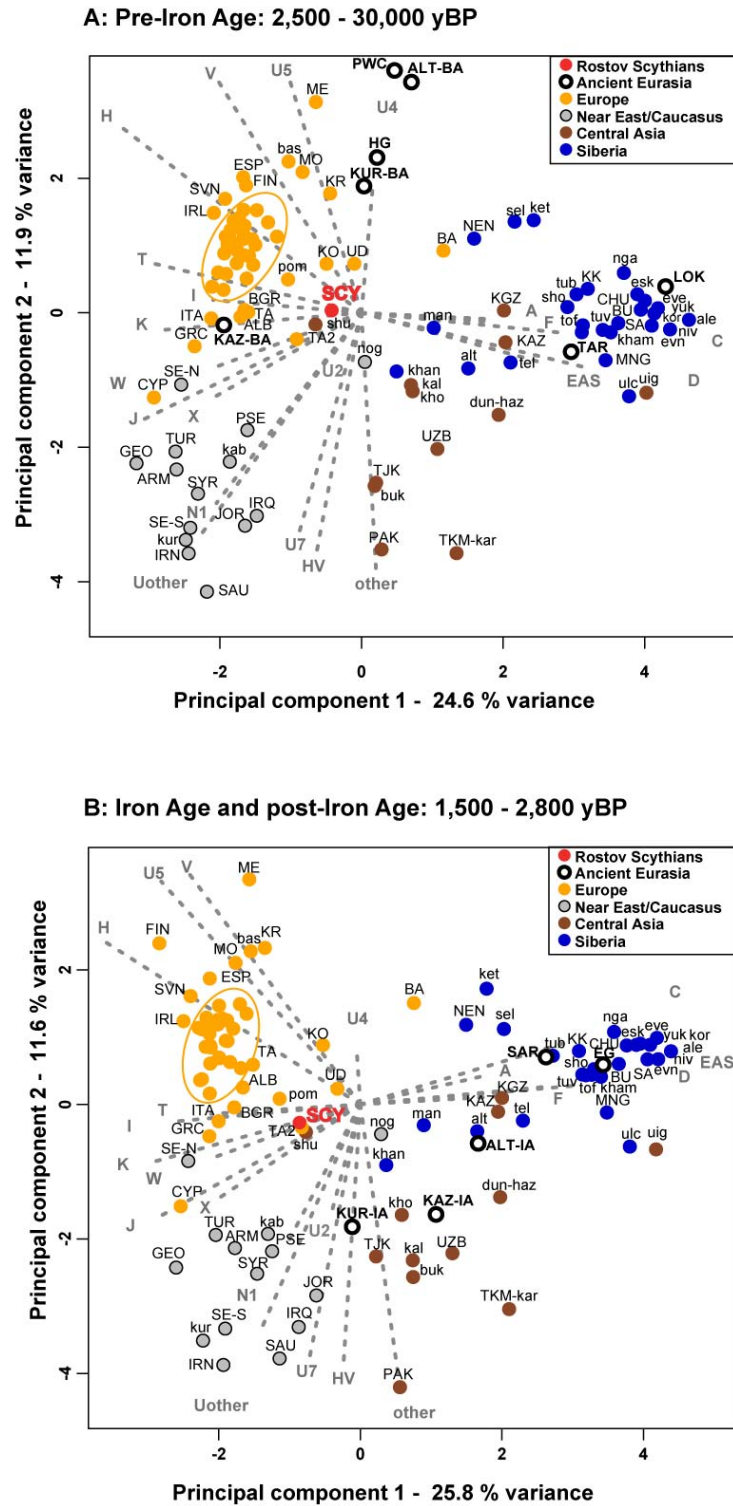


Figure 2. Principal component analysis of 21 mitochondrial haplogroup frequencies comparing modern Eurasian populations and Rostov Scythians with (A) pre-Iron Age populations (30,000 – 1,600 yBP) and (B) Iron Age/post-Iron Age populations (2,800 – 1,500 yBP). PCA axes 1 and 2 account for 24.6% and 11.9% of the total variance, respectively for (A) and 25.8% and 11.6% of the total variance for (B). Arrows represent haplogroup vectors. European Populations within the yellow circles are European populations, as described in Material and Methods, see also for haplogroup pooling and population abbreviations.

Informative haplotypes of Scythians shared with Central Asians

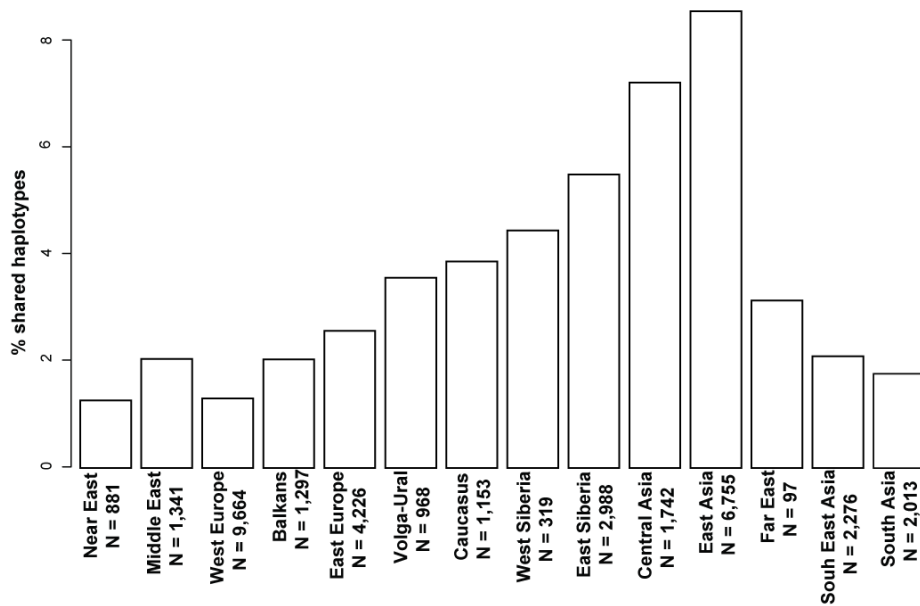
When the distribution of each of the informative haplotypes found in Scythians was examined among present-day Eurasian pools, Central Asia appeared as a common denominator (Figure 3B). Despite the small sample-size of the comparative database available for Central Asians (N = 1,742) in comparison to east (N = 14,448) and west (N = 16,155) Eurasians, six out of ten of the informative haplotypes detected in Scythians could be observed in modern-day populations of Central Asia (U5a, U2e, H2a1, A4, F1b, D*). Interestingly, none of the haplotypes reported here was shared concomitantly among Central Asians, Eastern Eurasians and Western Eurasians. The haplotypes found in Scythians displayed instead an interesting pattern of distribution where matches were found either (1) in both Central Asians and Western Eurasians (H2a, U2e, U5a haplotypes) or (2) in both Central Asians and Eastern Eurasians (A4, F1b and D* haplotypes).

(1) Haplotypes shared with modern-day populations of Central Asia and west Eurasia (western and eastern Europe) were haplotypes H2a1 (RD-10), U2e (individual RD-12), and U5a (individual RD-16). When considering present-day populations, haplotypes H2a1 and U2e displayed wide geographical distributions over Europe (Figure 3B). In contrast, U5a exhibited a distribution, which was mainly restricted to east Europe (Ukraine, Poland, Lithuania and Russians of the Vladimir, Smolensk and Novgorod districts, west Russia; see Table S5 for all population descriptions and associated references).

When considering past populations, haplotype H2a1 had several closely related matches in ancient populations over a large period of time (from the Mesolithic to the Iron Age; Figure 4). Exact matches and/or close relatives for U5a and U2e lineages were found among populations of the pre-Iron Age (HG, KAZ-BA, KUR-BA, ALT-BA) and the Iron Age (ALT-IA, KUR-IA) of Europe, Siberia, the Tarim basin and Central Asia.

(2) Scythian haplotypes shared with modern-day populations of Central Asia and east Eurasia include haplotypes A4 (individual RD-8), F1b (individual RD-1) and D* (individual RD-13). In present-day populations, exact matches for these three haplotypes detected in Scythians individuals were found Central Asia, Siberia and east Asia (Figure 3B).

A. Geographical distribution of all haplotypic matches



B. Geographical distribution of matches by haplotype

Geographical location of the haplotypic matches

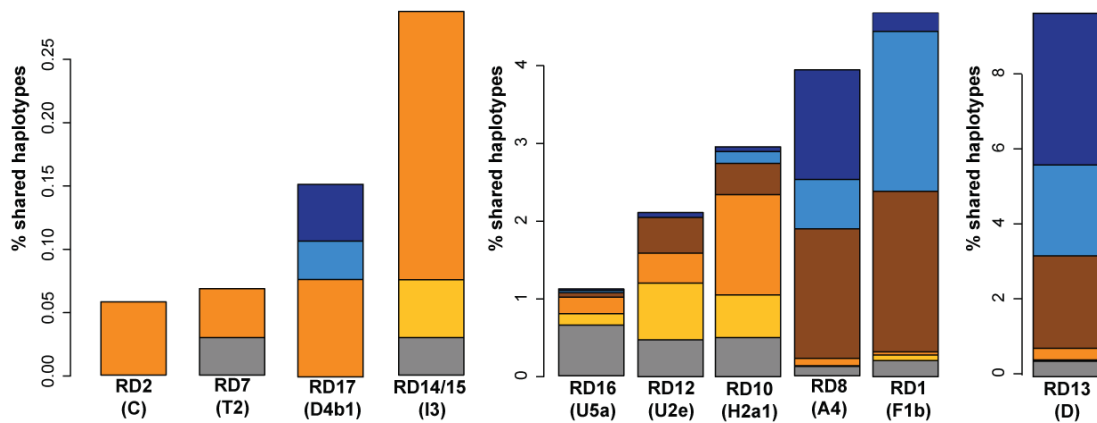
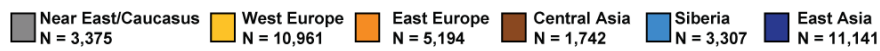


Figure 3. Percentage of informative Rostov Scythian haplotype matches in modern Eurasian populations. Figure A represents the geographical distribution of the matches for all the informative haplotypes sequenced in the Rostov Scythian individuals. The haplotypes obtained for individual RD5 found no match in the modern comparative database. Figure B represents the geographical distribution of the matches for each informative haplotype.

In past populations, exact matches could be found for haplotypes A4 and F1b: in Iron Age individuals of Kazakhstan (KAZ-IA) and Mongolia (EG) for haplotype A4, and in individuals of the Neolithic peri-Baikal area (LOK) and Iron Age Mongolia (EG) for haplotype F1b (Figure 4). None of three haplotypes found in Scythians and present-day Central Asians, Siberians and East Asians could be detected in Bronze Age nomadic populations of Kazakhstan (KAZ-BA), Kurgans (KUR-BA) and Altai (ALT-BA). In contrast, for all Iron Age populations of the same geographic areas (KAZ-IA, KUR-IA and ALT-IA) exact matches and close derivatives were reported.

In summary, lineages detected in Scythians and found today in western Eurasia and Central Asia (1) were previously reported in many pre-Iron Age populations from western Eurasia to the peri-Baikal area (Siberia). Lineages found today in east Eurasia and Central Asia (2) were found in pre-Iron Age populations of east Eurasia (but not in central Eurasia) and in Iron Age populations from east Eurasia to the Black Sea (including central Eurasia).

Informative haplotypes of Scythians absent from modern-day Central Asians

Haplotype sharing analysis could not detect the four informative haplotypes I3 (individuals RD-14 and 15), T2 (individual RD-7), C (individual RD-2) and D4b1 (individual RD-17) from the Scythian dataset in present-day populations of Central Asia (Figure 3B). The lack of matches in this region could be explained by the smaller comparative sample size available for Central Asians but might also indicate specific genetic links with modern-day populations of Europe, where they could be detected in a small number of individuals.

Haplotype I3 (individuals RD-14 and 15) had the widest distribution in modern-day populations with exact matches found in Poland, Russia, Hungary, Turkey, and Italy. In past populations, closely related derived I3 lineages displayed a wide distribution including Bronze Age Kazakhs and Iron Age Kurgans of southern Siberia. Of note, the I3 haplotype was shared with two Scythian individuals from the same graveyard (graveyard Glinishe I, Starocherkasskaya, Aksaysky district). As a consequence, it cannot be excluded that these two individuals were maternally closely related. This is the only instance of possible maternal kinship detected in the present study.

Haplotype T2 (individual RD-7) could be found in present-day Tatars of the Volga-Ural region (west Russia), Western Russians of the Yaroslavl district, Cherkessians of the Caucasus. One or two-step derivatives of this Scythian haplotype were detected in both the Bronze and Iron Age populations of Kazakhstan, exemplifying the close genetic relationship between these populations.

Haplotype C (individual RD-2) was found to be restricted to modern-day Hungarians and Bashkirs of the Volga-Ural region (west Russia). While only being found in Europe today, this particular C haplotype could not be observed in any ancient populations of Europe. In addition, the presence of one-step derivatives (16223T-16298C-16327T) in Neolithic populations of the peri-Baikal area (LOK), Bronze Age populations of the Tarim Basin (TAR) and Iron Age individuals of the Sargat culture in western Siberia (SAR) suggests an origin of this C lineage in eastern Eurasia. The C lineage must have been brought into Europe by a migration from Siberia before the Iron Age and might later have gone virtually extinct in Siberia. No haplotype closely related to this C lineage could be found in the populations of ancient central Eurasian Kurgans, Altaians and Kazakhs either in the Bronze Age or in the Iron Age.

Haplotype D4b1 (individual RD-17) might have had an east Eurasian origin, as closely related haplotypes were found in Iron Age populations of Mongolia (EG) and the Altai (ALT-IA), but not in ancient populations of Europe. It was detected in modern-day populations of Siberia and eastern Eurasia (Tuvinians, Koreans and Tibetans), but also in eastern European Bashkirs. No match was found in Central Asia, in present-day or in past populations.

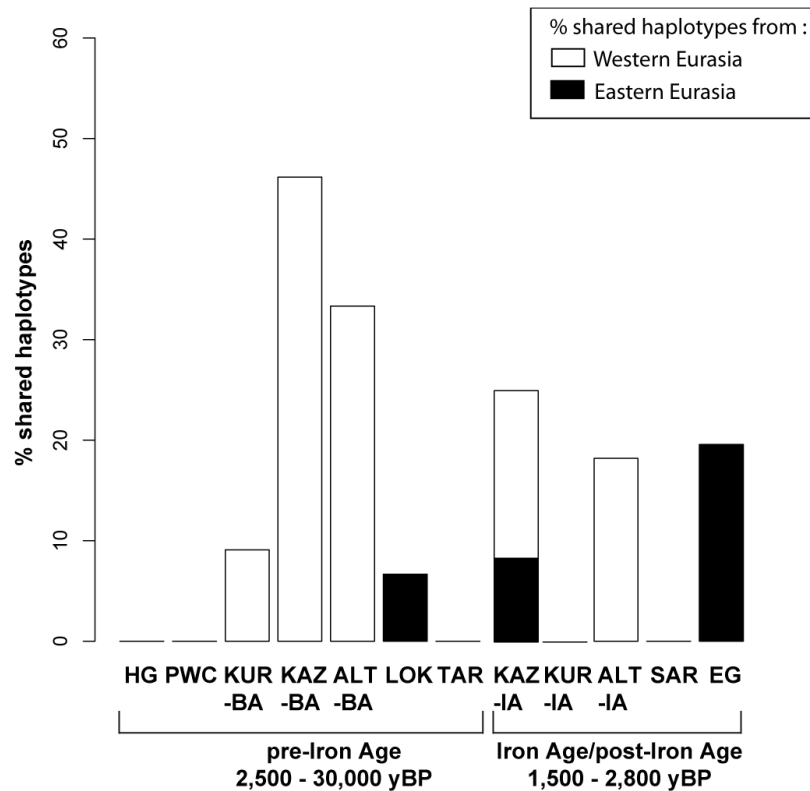


Figure 5. Haplotype sharing analysis. Percentage of haplotypes shared with Rostov Scythians.

Mitochondrial homogeneity among ancient populations of central Eurasia

The network analysis showed that the HVR-I sequences obtained for the Scythians were phylogenetically closely related to haplotypes previously reported in pre-Iron Age and Iron Age populations of Eurasia (Figure 4). In order to visualise ancient population differentiation, we used MDS of Slatkin's F_{ST} calculated among all ancient (Figure 6A), Pre-Iron Age (Figure 6B) and Iron Age/post-Iron Age (Figure 6C) populations using haplotypic data. In the pre-Iron Age (Figure 6B), the populations of Kazakhs (KAZ-BA) Kurgans (KUR-BA) and inhabitants of the Altai (ALT-BA) appeared more closely related to each other than to European hunter-gatherers (HG) or Scandinavian Pitted-Ware Culture individuals (PWC). In the Iron Age (Figure 6C), Scythians (SCY) and Iron Age Kazakhs (KAZ-IA) were shown to be equally distant to Altaians (ALT-IA) and to Kurgans (KUR-IA). MDS revealed more significant homogeneity in the pre-Iron Age (KUR-BA, ALT-BA and KAZ-BA) than in the Iron-Age (KUR-IA, ALT-IA and KAZ-IA), with Iron Age Kazakhs showing the most important differentiation from the other pre-Iron Age and Iron Age populations.

Mitochondrial continuity in eastern Siberian populations

The phylogenetic network of mtDNA haplotypes (Figure 4) revealed a striking genetic continuity between eastern Siberian populations of the Bronze Age (TAR, LOK) and the Iron Age (EG). PCA plots of haplogroup frequency data also suggested genetic continuity between populations of eastern Siberia (Figure 2). Populations of the peri-Baikal Neolithic (LOK), the Bronze Age Tarim Basin (TAR), and Iron Age individuals of Mongolia (EG) could be placed within the mtDNA variability of eastern Eurasia as a result of their high frequencies in ‘eastern’ haplogroups. Genetic similarities identified on the basis of haplogroup frequencies data in ancient populations of east Eurasia were tested for haplotypic data by AMOVA. If the three populations (LOK, TAR and EG) are considered part of a single meta-population, the ‘within population’ variance represented 80.84% of the total variance ($p < 0.001$). Based on the observation of human remains exhibiting ‘Caucasoid’ morphological features (as opposed to ‘Mongoloid’ morphological features in present-day eastern Eurasia and central Eurasia) and the detection of ‘western’ lineages in the population of the Tarim Basin (Li et al., 2010), these were also tested for the presence of genetic structure segregating the Bronze Age population of the Tarim Basin (TAR) from other eastern Eurasian populations. The ‘within population’ variance of 77.76% of the total variance ($p < 0.001$) was smaller than the one observed in the absence of genetic structure, thus supporting the genetic differentiation of the Bronze Age population of the Tarim Basin suggested from phenotypic observations.

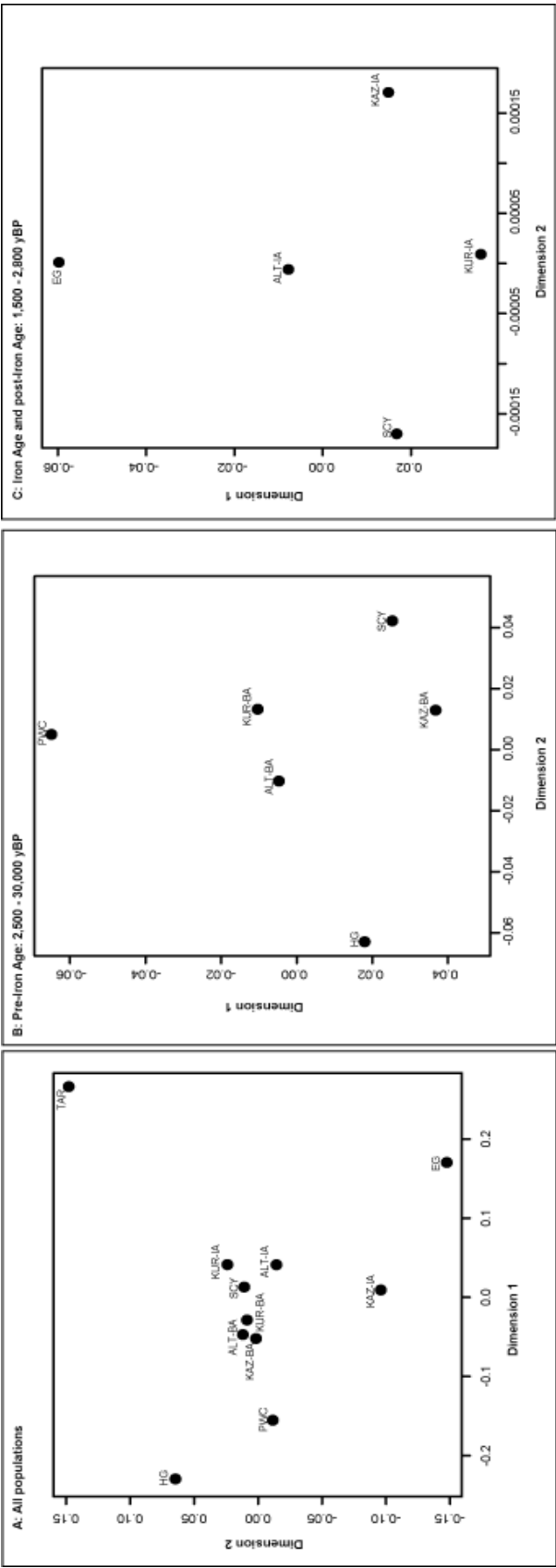


Figure 6. Classical Multi-Dimensionnal Scaling of Slatkin's fixation indexes F_{ST} .

DISCUSSION

Mitochondrial makeup of Scythians

Ancient DNA revealed here the mtDNA structure of Scythians of the Rostov-on-Don region (Black Sea area, south west Russia). While haplogroup frequency-based PCA showed a genetic affinity between Scythians and present-day Europeans, the examination of the current distribution of the mtDNA haplogroups revealed a pattern of mixed western (haplogroups H, I, T and U) and eastern (haplogroups A, C, D and F) Eurasian origins similar to that observed in present-day Central Asians and Siberians. Analyses at the haplotypic level confirmed the genetic affinity of the Scythians with present-day Central Asians. Haplotype-sharing analyses also detected genetic influences from Siberia (C and D4b1 haplotypes) and the Central Asian corridor, i.e. Iran, Iraq, Pakistan, India (U7 haplotype). Two main scenarios can be proposed to explain the presence of the Siberian (C and D4b1) and the Central Asian (U7) lineages in the gene pools of Scythians:

1) The C, D4b1 and U7 lineages could have been brought from Siberia and the Central Asian corridor into the Black Sea area through migrations before or during the Iron Age.

2) The C, D4b1 and U7 lineages could have been incorporated into the Scythian gene pool through admixture with Siberian and Central Asian populations during Scythian migrations into these regions.

Overall, the aDNA analysis of the Scythians revealed multiple genetic influences from western and eastern Eurasia. Specific geographic areas such as Siberia and the Central Asian corridor could also be identified as the source of mitochondrial influences in the Scythian mtDNA gene pool. Previous aDNA analyses of Eurasian populations provided evidence to reconstruct the formation of the pattern of mixed origin observed in Scythians.

Western mtDNA substratum in Bronze Age nomads of central Eurasia

In the Bronze Age, the gene pools of nomadic populations of central Eurasia – Kazakhs (KAZ-BA), south Siberian Kurgans (KUR-BA) and Altaians (ALT-BA) – were mainly constituted of ‘western’ haplogroups (H, HV, K, I, T and U). ‘Western’ haplotypes (H, I3, T1a, T2, and U5a) similar or closely related to sequences observed in Bronze Age populations of central Eurasia (KAZ-BA, KUR-BA, and ALT-BA)

were detected in Scythians. These haplotypic relationships suggest a common origin for the mtDNA ‘western’ component of Scythians and Bronze Age central Eurasians. This genetic makeup was also shared with European hunter-gatherers (HG) and Scandinavian Pitted-Ware Culture individuals (PWC), as shown by examination of haplogroup frequencies and mtDNA sequences. For example, the haplotype U4 16356C was found in Palaeolithic/Mesolithic foraging populations of Europe (HG and PWC) and in Bronze Age Kurgans. Haplotypes closely related to sequences observed in ancient European foragers (HG) were found in Bronze Age Kazakhs, e.g., the 16311C derivative of the central/eastern European hunter-gatherer (HG) haplotype U5a 16192T-16256T-16270T. These two haplotypes (U4 and U5a) are basal and are expected to be more widely distributed in space and time than their derivatives. Accordingly, the U5a 16192T-16256T-16270T haplotype exhibited a wide distribution from Scandinavian Pitted-Ware culture individuals (PWC) in the West to the peri-Baikal area in Neolithic populations (LOK) in the East. Two explanations can be proposed for the wide distribution of ‘western’ haplogroups in the Bronze Age:

(1) The distribution of mtDNA lineages observed in central Eurasia in the Bronze Age is the result of genetic continuity with a Palaeolithic/Mesolithic substratum of ‘western’ mtDNA lineages that extended further east prior to the Iron Age than observed today. ‘Western’ lineages were found as far east as the peri-Baikal area, as exemplified by the Neolithic population of the Lokomotiv archaeological site (LOK), which displayed 6.7% of ‘western’ lineages, represented by a single U haplotype (Mooder et al., 2005).

(2) The ‘western’ lineages have reached the Eurasian Steppe through more recent migrations possibly by Yamna-derived Afanasievo culture individuals from the steppe North of the Black Sea (Hemphill & Mallory, 2004).

Genetic input from the East into ancient nomadic populations of central Eurasia

Genetic discontinuity between Bronze Age (KAZ-BA, KUR-BA, and ALT-BA) and Iron Age (KAZ-IA, KUR-IA, and ALT-IA) nomadic populations of central Eurasia was previously suggested to have been introduced through genetic input from eastern Eurasia prior to or during the Iron Age (Lalueza-Fox et al., 2004; Keyser et al., 2009). The (pre-) Iron Age introduction of lineages of ‘eastern’ origin produced mixed mtDNA gene pools similar to those observed today in ‘contact zones’ of Eurasia: western Siberia and Central Asia. The western mtDNA haplogroups observed in

Bronze Age populations (haplogroups H, T, U) remained in Iron Age populations of central Eurasia (KAZ, KUR and ALT). These Iron Age populations also showed evidence of haplotypic continuity with Bronze Age populations of central Eurasia. Comparison of these nomadic Bronze Age and Iron Age populations of central Eurasia supports the idea that carriers of 'eastern' mtDNA lineages did not completely replace populations of the corresponding areas but were incorporated into these populations. Analysis of the Y-chromosome diversity in Kurgans of south Siberia suggested an equal maternal and paternal contribution from the East and proposed a 'whole population' rather than 'war-like' model (causing sex-biased distributions of genetic markers) to describe past gene flow into south Siberia (Keyser et al., 2009).

Scythians displayed haplotypic similarities with other admixed Iron Age populations of central Eurasia (KUR-IA, KAZ-IA and ALT-IA), and both Scythians and Iron Age central Eurasians shared common haplotypes with ancient populations of eastern Eurasia (peri-Baikal Neolithic individuals, LOK; Bronze Age individuals of the Tarim Basin, TAR; Iron Age Xiongnu individuals of Mongolia, EG). These haplotypic similarities suggest that the genetic inputs that affected populations of central Eurasia (Kazakhs, south Siberian Kurgans and individuals of the Altai), and the ancestral populations of Scythians were genetically similar. The geographic origin of this gene flow could not be precisely identified as the eastern Asian lineages detected in Scythians, are currently widely distributed in eastern Eurasia (Siberia, east/far east/south/south east Asia).

None of the C and D4b1 lineages detected in Scythians has been observed in Bronze Age populations of central Eurasian nomads, whereas the only 'eastern' haplogroup detected in these populations was haplogroup Z in Bronze Age Kurgans (KUR-BA). Haplogroup Z was also detected in Iron Age individuals of the Sargat culture (SAR) of western Siberia. While haplogroups C and D are the two most common haplogroups over the whole of Siberia today, where they display a homogeneous distribution (Derenko et al., 2010), the distribution of haplogroup Z is more irregular over this vast area and reaches its highest frequencies and diversity in populations of north east Siberia (Volodko et al., 2008). The presence of haplogroup Z in eastern Europe, in populations of the Volga-Ural region, the Baltic area and in Saami of Fennoscandia was proposed to be a signal of past migrations from Siberia into north east Europe (Tambets et al., 2004; Ingman & Gyllensten, 2007; see also Chapter One). Because haplogroup Z has not been found in any ancient population of

eastern Eurasia (LOK, TAR) or any Iron Age nomadic group of central Eurasia (EG) so far, it can be proposed that the ‘eastern’ Z lineages were not part of the westward spread of east Eurasian genetic influences into central Eurasia that could be identified from the geographical and temporal distribution of the ‘eastern’ A, C, D, F and G lineages.

The genetic information recorded through time by previous aDNA studies of Eurasian populations seems to contradict the previously proposed ‘maturation phase’ model, based on Y-chromosome data (Wells et al., 2001). According to this model the mixed genetic makeup of the ancestral population of Central Asians was formed during a ‘maturation phase’ that preceded the initial Palaeolithic dispersal of humans in Central Asia ~40,000 – 50,000 yBP. The presence of ancient populations characterized by high frequencies of ‘western’ haplogroups in central Eurasia rather supports the model proposing that a later admixture between genetically differentiated populations from the East and the West has given rise to the genetic makeup of populations of central Eurasia. This was first shown by the study of mtDNA diversity in modern-day Eurasians and later confirmed by several ancient DNA studies (Comas et al., 2004; Quintana-Murci et al., 2004; Lalueza-Fox, 2004; Keyser et al., 2009).

Homogeneity among Iron Age populations of the central Eurasian Steppe

Genetic homogeneity among ancient populations of the Eurasian Steppe has been challenged previously, in particular among culturally similar groups of nomads (Yablonski et al., 2000). Previous results from aDNA studies allowed the direct testing of mtDNA homogeneity among ancient nomadic populations of central Eurasia.

In the Bronze Age, nomadic populations of Kazakhstan (KAZ-BA), the Altai (ALT-BA) and southern Siberia Kurgans (KUR-BA) shared common genetic features. The gene pools of Bronze Age populations of central Eurasia were indeed dominated by ‘western’ lineages and close haplotypic relationships could be observed among sequences reported for these populations. Differentiation among these Bronze Age populations was observed when comparing haplogroup frequencies and could be the result of sampling biases introduced by the small amount of ancient data available. Alternatively, these differences could reflect the impact of different genetic influences on these geographically distant populations (Kurgans and Altaians in the north of the central Eurasian Steppe versus Kazakhs in the South). However, Bronze Age nomadic populations of central Eurasia (KAZ-BA, KUR-BA and ALT-BA) were shown to be

more closely related to each other than to other pre-Iron Age populations of the Neolithic peri-Baikal area (LOK) and Bronze Age Tarim Basin (TAR), which fell within the mtDNA diversity of present-day Eastern Siberians.

In the Iron Age, the mtDNA gene pools of nomadic populations of central Eurasia (KAZ-IA, KUR-IA and ALT-IA) were all made up of a 'western' and an 'eastern' component. In that regard, populations of Kurgans (KUR-IA) and Kazakhstan (KAZ-IA) showed a genetic affinity with populations of Central Asia. Genetic affinity between Iron Age Scythians and present-day Central Asians was also demonstrated through the examination of haplogroup and haplotype data. However, varying haplogroup frequencies appeared to cause genetic heterogeneity among Iron Age central Eurasian nomads (KAZ-IA, KUR-IA and ALT-IA) but also among modern-day populations of Central Asia. The heterogeneous pattern of mtDNA lineage distribution among Central Asian populations could be explained by the randomising effect of small sample size or by the action of population processes such as, reduction in population size, varying amounts of external gene flow, reproductive isolation, and founder events. Demographic processes could explain the discrepancy observed between the results obtained for Scythians, on the one hand, from PCA (affinity between Scythians and present-day Europeans), and on the other hand, from examination of haplotypic data and of the current haplogroup distribution in Eurasia (affinity between Scythians and present-day Central Asians). The randomising effects of small population sizes and/or demographic processes on haplogroup frequencies could also be at the origin of the population differentiation observed between the Iron Age nomadic populations of Scythians (SCY), Kazakhs (KAZ-IA), south Siberian Kurgans (KUR-IA) and Altaians (ALT-IA), as haplotypes reported from these populations were shown to be closely related. Shared haplotypes among populations of (pre Iron Age and Iron Age) nomads of central Eurasia could indicate similar origins and/or the transfer of mtDNA lineages through admixture favoured by the nomadic lifestyle of these populations.

Another sign of genetic differentiation between Scythians and other Iron Age nomadic populations of central Eurasia was the presence of haplogroup U7, which was exclusively found in Scythians. The presence of the U7 haplotype in Scythians indicated a limited genetic input from the Central Asian corridor: the Middle East (Iran and Iraq), India and Pakistan. Of note, haplogroup U7 was also detected in modern-day western Siberian populations of Khants (14.2 %) and Mansi (3.2 – 5%;

Pimenoff et al., 2008), but no direct genetic link between Scythians and Western Siberians could be identified. Haplogroup U7 is indeed represented in modern-day western Siberians by a unique U7 haplotype (16309C-16318t) that is different from the one observed here in Scythians (16256T-16318t). Because of the lack of genetic diversity in the west Siberian U7 haplogroup, its presence in modern-day Khants and Mansi at detectable frequencies was interpreted as the result of recent gene flow and founder effect and/or population bottleneck. As a consequence, the migrations that brought haplogroup U7 into Scythians and Western Siberians might have been distinct.

‘Western’ genetic influence in the Bronze Age Tarim Basin

Mitochondrial haplogroup frequency and haplotypic data from individuals of the Neolithic peri-Baikal area (LOK), the Iron Age Xiongnu culture in Mongolia (EG), and present-day south east Siberia indicated a long-term continuity in this area. Bronze Age mummies of the Tarim Basin were previously shown to exhibit ‘Caucasoid’ morphological features, an influence from western Eurasia that was confirmed by the detection of ‘western’ lineages at low frequencies in the Bronze Age Tarim Basin (Li et al., 2010). AMOVA of haplotypic data in ancient eastern Eurasian populations showed in this study that the Bronze Age population of the Tarim Basin could be differentiated from the Neolithic population of the peri-Baikal area (LOK) and the Iron Age population of Mongolia (EG). Based on archaeological evidence, a Kurgan origin in the Andronovo culture (~3,000 – 4,000 yBP) of the south Russian steppe was suggested for the western genetic component of the Bronze Age Tarim Basin (Hemphill & Mallory, 2004). This was proposed in support for the ‘Steppe hypothesis’ for a south Siberian origin of the ‘Caucasoid’ populations of the Tarim Basin. In contrast, the alternative ‘Bactrian oasis hypothesis’ that favours a Central Asian origin could not be supported genetically. Like the population of the Tarim Basin, individuals of the Bronze Age Andronovo culture (3,000 – 4,000 yBP) of south Siberia (KUR-BA) were shown to display both ‘western’ Eurasian mtDNA lineages and phenotypic traits. The presence of ‘Caucasoid’ phenotypes, i.e. light brown hair, fair skin and blue or green eyes, was indeed proposed in Andronovo individuals on the basis of the typing of predictive SNPs in human pigmentation genes (Keyser et al., 2009).

CONCLUSION

In Central Asia, previous studies comparing genetic data from the mtDNA genome, the Y-chromosome and autosomes in modern-day populations revealed sexually asymmetric demographic processes and the importance of social structures and mating patterns (Chaix et al., 2007; Chaix et al., 2008; Ségurel et al., 2008; Heyer et al., 2009). The presence of a hierarchical social structure in the Scythians presented here could probably introduce a bias in the present reconstruction of the population history of Scythians. The mtDNA structure of the Scythians only represents the portion of the population present in Scythian burials, which were usually members of a higher social class. In the case where ancient nomadic populations of central Eurasia, like present-day Central Asians, were organised in patriarchal societies, it would be important to examine the Y-chromosome diversity of ancient nomadic populations of central Eurasia in order to determine whether genetic homogeneity is present on the paternal side. This work is necessary to complete the picture of the population history of humans in central Eurasia reconstructed owing to a significant amount of diachronically sampled mtDNA in this region.

This study is the first to describe the mtDNA structure of Scythians of the Black Sea. We showed here that Scythians were the recipients of genetic input from multiple sources. The gene pool of Scythians was influenced by the western Palaeolithic Eurasian mtDNA substratum, which prior to the Iron Age spread further east than today. The genetic makeup of Scythians was also influenced by 'eastern' lineages that may have spread westward from East Eurasia prior to or during the Iron Age. Traces of genetic influence from Siberia and the Central Asian corridor were also detected. In order to provide answers to the question of the Scythian ethnogenesis and to test the hypothesis of the local origin of Scythians, it will be necessary to obtain aDNA data from the Black Sea area dated prior to the Iron Age.

Haplogroup and haplotype similarities could be identified between Scythians and other Iron Age nomadic populations of central Eurasia. The apparent genetic similarity between ancient populations of central Eurasian nomads could indicate a relatively recent common origin. Alternatively, close genetic relationships could have been the result of homogenisation through contact and intermarriage between nomadic groups that were facilitated by the mobility of nomads across the Eurasian Steppe. The same processes could explain the similarities observed in the material culture of

Eurasian Steppe nomads. Cultural traits (art and technologies) may have been exchanged as part of long-standing trading networks among nomadic populations of central Eurasia, as could have been genetic material.

LIST OF SUPPLEMENTARY MATERIALS

Figure S1: Pictures of selected samples from Scythians, exemplifying levels of preservation. Samples RD-2 and RD-6 yielded consistent haplotypes. Sequences obtained from sample RD-3 could be replicated from two different extracts at the Australian Centre for Ancient DNA, University of Adelaide, and from another sample independently at the University of Pompeu-Fabra, Barcelona. RD-4 yielded ambiguous non reproducible.

Figure S2A-C: Clone sequences for individual RD-3 (replicated U7 haplotype; A), individual RD-3 (non-replicated U5b haplotype; B), and individual RD-12 (B). The mutation at position 16183C was not considered for statistical analyses and phylogenetic reconstruction.

Table S1: Results of SNP typing in the mtDNA coding region using the GenoCoRe22 SNaPshot assay. SNPs typed on the L-strand are reported in capital letters in the reference rCRS profile, whereas SNPs typed on the H-strand are reported in small letters. Missing data signifies allelic dropout or fluorescence signal below the background threshold (100 relative fluorescent units, rfu). ‘g/a’ indicates the presence of a mixed signal for the position interrogated. A mixed signal was repeatedly obtained at position 8994 (haplogroup W) with the detection of an additional G base. However, the rest of the profile never could phylogenetically support the presence of the G base at this particular position. Despite of the allelic dropout consistently observed at position 10034 (haplogroup I), the remaining profile of individuals RD-14 and RD-15 and their HVR-I sequence were consistent to their assignment within haplogroup I. For each individual, profiles were obtained from two independent extracts, except for the individuals replicated at the University of Pompeu-Fabra, Barcelona: RD-3, RD-12.

Table S2: Results of quantitative PCR.

Table S3: Details of the Scythian archaeological sites of the Rostov-on-Don area sampled for ancient DNA.

Table S4: Details and references of the comparative population datasets used in Principal Component Analyses.

Table S5: Details and references of the comparative populations used in shared haplotype analyses.

ACKNOWLEDGMENTS

We thank Elena Batieva for providing the samples. We warmly acknowledge Oleg Balanovsky, Valery Zaporozhchenko and Elena Balanovska of the Russian Academy of Medical Sciences for compiling the comparative database of modern-day populations. We thank Valery Zaporozhchenko for searching the comparative modern-day database. We acknowledge Carles Lalueza-Fox and Oscar Ramirez of the University of Pompeu-Fabra, Barcelona, and Bastien Llamas of the Australian Centre for Ancient DNA, University of Adelaide for the independent replication of the analysis. We thank Jeremy Austin and Bastien Llamas for linguistic revision and helpful comments.

REFERENCES

1. Adler, C. J., Haak, W., Donlon, D. & Cooper, A. (2011). Survival and recovery of DNA from ancient teeth and bones. *J Archaeol Sci* 38, 956-964.
2. Alexeev, V. P. (1980). *Diskussionniiye problemy otechestvennoi skifologii. Narodi Azii i Afriki* 6, 81-2 (Discussions of the Problems of Native Scythology. Population of Asia and Africa).
3. Al-Zahery, N., Semino, O., Benuzzi, G., Magri, C., Passarino, G., Torroni, A., Santachiara-Benerecetti, A.S. (2003). Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol Phylogenet Evol* 28, 458-472.
4. Andrews, R., Kubacka, I., Chinnery, P., Lightowlers, R., Turnbull, D., Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23, 147.
5. Anthony, D. W. (2007). *The Horse, the Wheel, and Language: How Bronze Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton, NJ: Princeton University Press.
6. Bandelt, H.J., Forster, P., Sykes, B.C., Richards, M.B. (1995). Mitochondrial portraits of human populations using median networks. *Genetics* 141, 743-753.
7. Barber, E. (1999). *The Mummies of Ürümchi*. New York: Norton.
8. Bashilov, V.A., Yablonsky, L.T. (2000). in *Kurgans, Ritual Sites, and Settlements: Eurasian Bronze and Iron Age*. Davis-Kimball, J., Murphy, E.M., Koryakova, L., Yablonsky, L.T. (eds.). BAR International Series 890. Oxford: Archeopress.
9. Behar, D., M., 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., Villems, R., Skorecki, K. (2008). Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. *PLoS One* 3, e2062.

10. Bennett C.C., Kaestle F.A. (2010). Investigation of Ancient DNA from Western Siberia and the Sargat Culture. *Human Biology*. 82 (2).
11. Bokovenko, N., Yablonsky, L.T. (2000). in *Kurgans, Ritual Sites, and Settlements: Eurasian Bronze and Iron Age*. Davis-Kimball, J., Murphy, E.M., Koryakova, L., Yablonsky, L.T. (eds.). BAR International Series 890. Oxford: Archeopress.
12. Drews, R. (2004). *Early Riders: The Beginnings of Mounted Warfare in Asia and Europe*. London: Routledge.
13. Bramanti, B., Thomas, M., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M., Jankauskas, R., Kind, C., Lueth, F., Terberger, T., Hiller, J., Matsumara, S., Forster, P., Burger, J. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326, 137-140.
14. Caramelli, D., Vernesi, C., Sanna, S., Sampietro, L., Lari, M., Castri, L., Vona, G., Floris, R., Francalacci, P., Tykot, R., Casoli, A., Bertranpetit, J., Lalueza-Fox, C., Bertorelle, G., Barbujani, G. (2007). Genetic variation in prehistoric Sardinia. *Hum Genet* 122(3-4) 327-336.
15. Chaix, R., Austerlitz, F., Hegay, T., Quintana-Murci, L., Heyer, E. (2008). Genetic traces of east-to-west human expansion waves in Eurasia. *Am J Phys Anthropol* 136, 309-317.
16. Chaix, R., Quintana-Murci, L., Hegay, T., Hammer, M.F., Mobasher, Z., Austerlitz, F., Heyer, E. (2007). From social to genetic structures in Central Asia. *Curr Biol* 17, 43-48.
17. Chikisheva, T.A., Gubina, M.A., Kulikov, I.V., Karafet, T.M., Voevoda, M.I., Romaschenko A.G. (2007) A paleogenetic study of the prehistoric populations of the Altai. *Anthropology*. 4(32).
18. Chlenova, N. L. (1997). *Tsentralnaya Azia i skifi. Data kurgana Arzhan i ego mesto v sisteme kultur skiaskogo mira. Moskva* (Central Asia and the Scythians. Chronological data of the Arzhan Kurgan and its Place in the Cultural System of the Scythian World).
19. Comas, D., Calafell, F., Mateu, E., Pérez-Lezaun, A., Bosch, E., Martínez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., Pettener, D., Bertranpetit, J. (1998). Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. *Am J Hum Genet* 63, 1824-1838.
20. Comas, D., Plaza, S., Wells, R.S., Yuldaseva, N., Lao, O., Calafell, F., Bertranpetit, J. (2004). Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur J Hum Genet* 12, 495-504.
21. Cordaux, R., Saha, N., Bentley, G.R., Aunger, R., Sirajuddin, S.M., Stoneking, M. (2003). Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur J Hum Genet* 11, 253-264.
22. Debets, G. F. (1948). *Paleoantropologia SSSR (Trudi Instituta Etnografii. Novaya serie IV). Moskva - Leningrad : AN SSSR* (Paleoanthropology of the USSR. Research of the Institute of Ethnography. New Series).
23. Deguilloux, M.F., Soler, L., Pemonge, M.H., Scarre C., Jousaume, R., Laporte, L. (2010). News from the west: Ancient DNA from a French megalithic burial chamber. *Am J Phys Anthropol* 144 (1), 108-118.
24. Excoffier, L., Laval, G., Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1, 47-50.
25. Excoffier, L., Smouse, P.E., Quattro, J.M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human

- mitochondrial DNA restriction data. *Genetics* 131, 479-491.
25. Fiori, G., Facchini, F., Pettener, D., Rimondi, A., Battistini, N., Bedogni, G. (2000). Relationships between blood pressure, anthropometric characteristics and blood lipids in high- and low-altitude populations from Central Asia. *Ann Hum Biol* 27, 19-28.
 26. Grakov, B. N., Melukova, A. I. (1954). *Ob etnicheskikh i kul'turnih razlichiyakh v stepnikh i leso-stepnikh oblastiakh Yevropeiskoi chasti SSSR v skifskoe vremya, p. 93 in Shelov, D. B. (eds.), Voprosi skifo-sarmatskoi arkheologii. Moskva: Nauka* (Ethnic and Cultural Differences in the Steppes and Forrest-Steppe Areas of the European Part of the USSR during the Scythian Period. Questions of Scytho-Sarmatian Archaeology).
 27. Grakov, B. N. (1977). *Rannii zheleznii vek. Moskva: Moscow State University* (The Early Iron Age).
 28. Gryaznov, M. P. (1980). *Arzhan: tsarskii kurgan skifskogo vremeni. Leningrad: Nauka*. (Arzhan: A Tzar's Kurgan of the Scythian Period).
 29. Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K.W., Burger, J. (2005). Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310, 1016-1018.
 30. Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., Der Sarkissian, C.S., Brandt, G., Schwarz, C., Nicklisch, N., Dresely, V., Fritsch, B., Balanovska, E., Villems, R., Meller, H., Alt, K.W., Cooper, A., Genographic consortium. (2010). Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. *PLoS Biol* 8, e1000536.
 31. Hanks, B. (2010) *Archaeology of the Eurasian Steppe Iron Age*, accepted for publication with Cambridge University Press, World Archaeology Series.
 32. Hemphill, B.E., Mallory, J.P. (2004). Horse-mounted invaders from the Russo-Kazakh steppe or agricultural colonists from western Central Asia? A craniometric investigation of the Bronze Age settlement of Xinjiang. *Am J Phys Anthropol* 124:199–222.
 33. Herrnstadt, C., Elson, J., L., Fahy, E., Preston, G., Turnbull, D., M., Anderson, C., Ghosh, S., S., Olefsky, J., M., Beal, M., F., Davis, R., E., Howell, N. (2002). Reduced-median-network analysis of complete mitochondrial DNA coding region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* 70, 1152-1171.
 34. Heyer, E., Balaesque, P., Jobling, M.A., Quintana-Murci, L., Chaix, R., Segurel, L., Aldashev, A., Hegay, T. (2009). Genetic diversity and the emergence of ethnic groups in Central Asia. *BMC Genet* 10, 49.
 35. Ingman, M., Kaessmann, H., Pääbo, S., Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* 408, 708-713.
 36. Irwin, J., Ikramov, A., Saunier, J., Bodner, M., Amory, S., Röck, A., O'Callaghan, J., Nuritdinov, A., Atakhodjaev, S., Mukhamedov, R., Parson, W., Parsons, T.J. (2010). The mtDNA composition of Uzbekistan: a microcosm of Central Asian patterns. *Int J Legal Med* 124, 195-204.
 37. Keyser, C., Bouakaze, C., Crubézy, E., Nikolaev, V., 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.
 38. 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.

39. Kivisild, T., Bamshad, M., J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., Laos, S., Parik, J., Watkins, W., S., Dixon, M., E., Papiha, S., S., Mastana, S., S., Mir, M., R., Ferak, V., Villems, R. (1999). Deep common ancestry of indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9, 1331-1334.
40. Koryakova, L.N., Epimakhov, A.V. (2007). *The Urals and Western Siberia in the Bronze and Iron Ages*. Cambridge University Press.
41. Kozintsev, A.G. (2007). Scythians of the North Pontic Region: between-group cranial variation, affinities, and origins. *Archaeology, Ethnology and Anthropology of Eurasia* 4 (32).
42. Krause, J., Briggs, A., 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.
43. Kuz'mina, E. (2008). *The Prehistory of the Silk Road*. Philadelphia: Univ. Penn. Press.
44. Lalueza-Fox, C., Sampietro, M., Gilbert, M., Castri, L., Facchini, F., Pettener, D., Bertranpetit, J. (2004). Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient Central Asians. *Proc Biol Sci* 271, 941-947.
45. Lalueza-Fox, C., Römpler, H., Caramelli, D., Stäubert, C., Catalano, G., Hughes, D., Rohland, N., Pilli, E., Longo, L., Condemi, S., de la Rasilla, M., Fortea, J., Rosas, A., Stoneking, M., Schöneberg, T., Bertrandoetit, J., Hofreiter, M. (2007). A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318, 1453-1455.46.
46. Levine, M. (2004). in *Traces of Ancestry: Studies in Honour of Colin Renfrew*. Jones, M. (eds.) pp. 115–26. Cambridge, UK: McDonald Inst. Monogr.
47. Maca-Meyer, N., González, A.M., Larruga, J.M., Flores, C., Cabrera, V.M. (2001). Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2, 13.
48. Mallory, J. (1989). *In Search of the Indo-Europeans: Language, Archaeology and Myth*. London: Thames and Hudson.
49. Mallory, J., Mair, V. (2000). *The Tarim Mummies: Ancient China and the Mystery of the Earliest Peoples from the West*. London: Thames and Hudson.
50. Malmström, H., Svensson, E.M., Gilbert, M.T., Willerslev, E., Götherström, A., Holmlund, G. (2007). More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA. *Mol Biol Evol* 24, 998-1004.
51. Malmström, H., Gilbert, M., Thomas, M., Brandström, M., Storå, J., Molnar, P., Andersen, P., 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.
52. Malyarchuk, B., Derenko, M., Grzybowski, T., Lunkina, A., Czarny, J., Rychkov, S., Morozova, I., Denisova, G., Miścicka-Sliwka, D. (2004). Differentiation of mitochondrial DNA and Y chromosomes in Russian populations. *Hum Biol* 76, 877-900.
53. Mandelshtam, A. M. (1966). *Kochevniki na puti v Indiu (Materiali i issledovania po arkheologii SSSR 136)*. Moskva: Nauka (The Nomads of the Route to India. Materials and Investigations of the Archaeology of the USSR).
54. Mandelshtam, A. M. (1967). *Novie pogrebenia srubnogo tipa v Yuzhnoi Turkmenii. Kratkie Soobshenia Instituta arkheologii 112, 61-5* (New Burials of the Srubnaya Culture in Southern Turkmenistan. Brief reports of the Moscow

- Institute of Archaeology).
55. Marcenko, K., Vinogradov, Y. (1989). The Scythian period in the Northern Black Sea region (750-250 BC). *Antiquity* 63, 803-13.
 56. Martínez-Cruz, B., Vitalis, R., Ségurel, L., Austerlitz, F., Georges, M., Théry, S., Quintana-Murci, L., Hegay, T., Aldashev, A., Nasyrova, F., Heyer, E. (2011). In the heartland of Eurasia: the multilocus genetic landscape of Central Asian populations. *Eur J Hum Genet* 19, 216-223.
 57. Mei, J. (2000). Copper and Bronze Metallurgy in Late Prehistoric Xinjiang: Its Cultural Context and Relationship with Neighboring Regions. BAR Int. Ser. 865. Oxford: Archaeopress.
 58. Mei, J., Shell, C. (2002). The Iron Age cultures in Xinjiang and their steppe connections. in *Ancient Interactions: East and West in Eurasia*. Boyle, K., Renfrew, C. (eds.) Cambridge, UK: McDonald Inst. Monogr.
 59. Melchior, L., Lynnerup, N., Siegismund, H., Kivisild, T., Dissing, J. (2010). Genetic Diversity among Ancient Nordic Populations. *PLoS One* 5, e11898.
 60. Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, AG., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, MD., Sukernik, RI., Olckers, A., Wallace, D.C. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A* 100, 171-176.
 61. Mooder, K.P., Weber, A.W., Bamforth, F.J., Lieverse, A.R., Schurr, T.G., Bazaliiski V.I., Savel'ev, N.A. (2005). Matrilineal affinities and prehistoric Siberian mortuary practices: a case study from Neolithic Lake Baikal. *J. Archaeol. Sci.* 32(4):619-634
 62. Murzin, V. Ju. (1990). *Proishozhdenie skifov: osnovnie etapi formirovanua skifskogo etnosa*. Kiev: Naukova dumka. (The Origin of the Scythians: The Main Stages in the Formation of the Scythian Ethnos).
 63. Nasidze, I., Quinque, D., Rahmani, M., Alemohamad, S.A., Stoneking, M. (2006). Concomitant replacement of language and mtDNA in South Caspian populations of Iran. *Curr Biol* 16, 668-673.
 64. Noonan, J.P., Hofreiter, M., Smith, D., Priest, J.R., Rohland, N., Rabeder, G., Krause, J., Detter, J.C., Pääbo, S., Rubin, E.M. (2005). Genomic sequencing of Pleistocene cave bears. *Science* 309, 597-599.
 65. Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annu Rev Genet* 38, 645-679.
 66. Pilipenko, A.S., Romaschenko, A.G., Molodin, V.I., Parzinger, H., Kobzev, V.F. (2010). Mitochondrial DNA studies of the Pazyryk people (4th to 3rd centuries BC) from northwestern Mongolia. *Archaeological and Anthropological Sciences* 2 (4): 231-236.
 67. Pimenoff, V.N., Comas, D., Palo, J.U., Vershubsky, G., Kozlov, A., Sajantila, A. (2008). Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. *Eur J Hum Genet* 16, 1254-1264.
 68. Potemkina, T. M. (1987). *K voprosu o migratsii na yug stepnikh plemen epohi bronzi. V zaimodeistvie kochevikh kultur i drevnikh tsivilizatsii, pp.76-7. Alma-Ata: Science* (On the Question of Bronze Age Steppe Tribal Migrations to the South." Interaction between Nomadic Cultures and Ancient Civilisations).
 69. Pyankova, L. T. (1974). *Mogilnik epohi bronzi Tigrovaya Balka. Sovetskaya arkheologia* 3, 171-80 (The Bronze Age cemetery of Tigrovaya Balka. Soviet Archaeology).

70. Pyankova, L. T. (1987). *K voprosu o semeinikh i obshestvennikh otnosheniyakh v epokhu pozdnei bronzi (po materialam mogilnikov Vakhshskoi kulturi)*. *Materialnaya kultura Tadjikistana* 4, 49-51 (On the Question of Familial and Public Relationships during the Bronze Age (on the basis of artifacts from cemeteries of the Vakhshskaya Culture). *Material Culture of Tajikistan*).
71. Quintana-Murci, L., Chaix, R., Wells, R., S., Behar, D., M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A., S., Coppa, A., Ayub, Q., Mohyuddin, A., Tyler-Smith, C., Qasim Mehdi, S., Torroni, A., McElreavey, K. (2004). Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet* 74, 827-845.
72. Raevsky, D. S. (1993). *Kulturno-Istoricheskoe edinstvo ili kulturnii kontinuum. Kratkie soobshenia Instituta Arkheologii* 207, 30-2 (Cultural-Historical Community or Cultural Continuum? Brief reports of the Institute of Archaeology).
73. Renfrew, C. (2002). in *Ancient Interactions: East and West in Eurasia*. Boyle, K., Renfrew, C. (eds.) Cambridge, UK: McDonald Inst. Monogr.
74. Ricaut, F.X., Keyser-Tracqui, C., Bourgeois, J., Crubézy, E., Ludes, B. (2004). Genetic analysis of a Scytho-Siberian skeleton and its implications for ancient Central Asian migrations. *Hum Biol* 76, 109-125.
75. Ricaut, F.X., Keyser-Tracqui, C., Cammaert, L., Crubézy, E., Ludes, B. (2004). Genetic analysis and ethnic affinities from two Scytho-Siberian skeletons. *Am J Phys Anthropol* 123, 351-360.
76. Richards, M., Macaulay, V., Bandelt, H., Sykes, B. (1998). Phylogeography of mitochondrial DNA in Western Europe. *Ann Hum Genet* 62, 241-260.
77. Richards, M., Côté-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H., Sykes, B. (1996). Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59, 185-203.
78. 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., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Renzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A., Bandelt, H-J. (2000). Tracing European founder lineages in the near Eastern mtDNA pool. *Am J Hum Genet*, 1251-1276.
79. Rosser, Z., Zerjal, T., Hurles, M., E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D., G., Brede, G., Cooper, G., Côté-Real, H., B., de Knijff, P., Decorte, R., Dubrova, Y., E., Evgrafov, O., Gilissen, A., Glisic, S., Gölge, M., Hill, E., W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S., A., Krumina, A., Kucinskas, V., Lavinha, J., Livshits, L., A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T., A., Mikelsaar, A., V., Mitchell, R., J., Nafa, K., Nicholson, J., Nørby, S., Pandya, A., Parik, J., Patsalis, P., C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M., J., Previderé, C., Roewer, L., Rootsi, S., Rubinsztein, D., C., Saillard, J., Santos, F., R., Stefanescu, G., Sykes, B., C., Tolun, A., Villems, R., Tyler-Smith, C., Jobling, M., A. (2000). Y-chromosomal diversity in Europe is clinal and influenced primarily by

- geography, rather than by language. *Am J Hum Genet* 67, 1526-1543.
80. Sampietro, M., Lao, O., Caramelli, D., Lari, M., Pou, R., Martí, M., Bertranpetit, J., Lalueza-Fox, C. (2007). Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc Biol Sci* 274, 2161-2167.
 81. Scozzari, R., Cruciani, F., Pangrazio, A., Santolamazza, P., Vona, G., Moral, P., Latini, V., Varesi, L., Memmi, M. M., Romano, V., De Leo, G., Gennarelli, M., Jaruzelska, J., Villems, R., Parik, J., Macaulay, V., Torroni, A. (2001). Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum Immunol* 62, 871-884.
 82. Ségurel, L., Martínez-Cruz, B., Quintana-Murci, L., Balaesque, P., Georges, M., Hegay, T., Aldashev, A., Nasyrova, F., Jobling, M., A., Heyer, E., Vitalis, R. (2008). Sex-specific genetic structure and social organization in Central Asia: insights from a multi-locus study. *PLoS Genet* 4, e1000200.
 83. Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A. S., Cavalli-Sforza, L. L., Underhill, P. A. (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290, 1155-1159.
 84. Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139, 457-462.
 85. Terenozhkin, A. I. (1971). *Skifskaya kultura. Problemi Skifskoi arkheologii. Materiali i issledovania po arkheologii SSSR 177. Moskva: Nauka* (Problems of the Scythian Culture. Materials and Investigations of the Archaeology of the USSR).
 86. Volodko, N., Starikovskaya, E., Mazunin, I., Eltsov, N., Naidenko, P., Wallace, D., Sukernik, R. (2008). Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocene peopling of the Americas. *Am J Hum Genet* 82, 1084-1100.
 87. Wallace, D.C., Brown, M.D., Lott, M.T. (1999). Mitochondrial DNA variation in human evolution and disease. *Gene* 238, 211-230.
 88. Wells, R., S., Yuldasheva, N., Ruzibakiev, R., Underhill, P., A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., Balakrishnan, K., Read, M., Pearson, N., M., Zerjal, T., Webster, M., T., Zholoshvili, I., Jamarjashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zalloua, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A., Bodmer, W. F. (2001). The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci U S A* 98, 10244-10249.
 89. Yablonsky, L. T. (1996). *Saki Yuznogo Priaralya (arkheologia I antropologia mogilnikov)*. Moskva: TIMR (The Saka of the Southern Aral Sea Area (the archaeology and anthropology of the cemeteries)).
 90. Yablonsky, L.T. (2000). in *Kurgans, Ritual Sites, and Settlements: Eurasian Bronze and Iron Age*. Davis-Kimball, J., Murphy, E.M., Koryakova, L., Yablonsky, L.T. (eds.). BAR International Series 890. Oxford: Archeopress.
 91. Zerjal, T., Wells, R.S., Yuldasheva, N., Ruzibakiev, R., Tyler-Smith, C. (2002). A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. *Am J Hum Genet* 71, 466-482.

SUPPLEMENTARY MATERIALS



Figure S1: Pictures of selected samples from Rostov Scythians. Samples RD-2 and RD-6 yielded consistent haplotypes. Sequences obtained from sample RD-3 could be replicated from two different extracts at the Australian Centre for Ancient DNA, University of Adelaide, but failed to be replicated independently at the University of Pompeu-Fabra, Barcelona. RD-4 failed to yield reproducible sequences.

A. Clone sequences for individual RD-3

[illegible]

B. - Clone sequences for individual RD-12

[illegible]

Figure S2A-B: Clone sequences. A: RD-3. B: RD-12. The mutations at position 16183C was not considered for statistical analyses and phylogenetic reconstruction.

Table S1: Results of SNP typing in the mtDNA coding region using the GenoCoRe22 SNaPshot assay. SNPs typed on the L-strand are reported in capital letters in the reference rCRS profile, whereas SNPs typed on the H-strand are reported in lower-case. Missing data signifies allelic dropout or fluorescence signal below the background threshold (100 relative fluorescent units, rfu). ‘g/a’ indicates the presence of a mixed signal for the position interrogated. A mixed signal was repeatedly obtained at position 8994 (haplogroup W) with the detection of an additional G base. However, the rest of the profile never could phylogenetically support the presence of the G base at this particular position. Despite of the allelic dropout consistently observed at position 10034 (haplogroup I), the remaining profile of individuals RD-14 and RD-15 and their HVR-I sequence were consistent to their assignment within haplogroup I. A For each individual, profiles were obtained from two independent extracts, except for the individuals replicated at the University of Pompeu-Fabra, Barcelona: RD-3, RD-12.

		13928_R9	3594_L3'4	10550_K	11467_U	4248_A	8994_W	13263_C	13368_T	11719_preHV_R0	8280delB	4580_V	6371_X	10238_N1	7028_H	10034_I	5178_D	14766_HV	12705_R	10873_N	12612_J	2758_L2'6	10400_M	Hg
rCRS		c	G	t	A	T	G	t	G	c	A	c	g	T	g	T	g	G	g	a	a	G	g	H
Sample RD- / Extract																								
1	1	G	G		A	T	G	T	G	T		C	G	T	A		G	A	G	A	g/a	G	G	R9
	2	G	G		A	T	G	T	G	T	A	C	G	T	A		G	A	G	A	A	G	G	R9
2	1	C	G	T	A	T	g/a	C	G	T	A	C	G	T	A	T	G	A	A	G	A	G	A	C
	2	C	G		A	T	G	C	G	T	A	C	G	T	A		G	A	A	G	A	G	A	C
3	1	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
5	1	C	G	T	G	T	g/a	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G		G	T	G	T	G	T	A	C	G	T	A		G	A	G	A	A	G	G	U
6	1	C	G	T	A	T	g/a	T	A	T	A	C	G	T	A	T	G	A	G	A	A	G	G	T
	2	C			A	T	G	T	A	T	A	C	G	T	A		G	A	G	g/a	A	G	G	T
7	1		G	T	A	T	g/a	T	A	T		C	G	T	A	T	G		G		A	G	G	T
	2	C			A	T	G	T	A	T	A	C	G	T	A		G	A	G	A	A	G	G	T
8	1	C	G	T	A	C	g/a	T	G	T	A	C	G	T	A	T	G	A	A	A	A	G	G	C
	2	C	G	T	A	C	g/a	T	G	T	A	C	G	T	A	T	G	A	A	A	A	G	G	C
9	1	C	G	T	A	T	g/a	T	G	C		C	G		G	T	G	G	G		A	G	G	H
	2	C			A	T	G	T	G	C	A	C	G	T	G		G	G	G	A	A	G	G	H
10	1	C	G	T	A	T	g/a	T	G	C	A	C	G	T	G	T	G	G	G	g/a	A	G	G	H
	2	C			A	T	G	T	G	C	A	C	G	T	G		G	G	G	A	A	G	G	H
11	1	C	G	T	A	T	G	T	A	T	A	C	G	T	A		G	A	G	A	A	G	G	T
	2	C			A	T		T	A	T	A	C	G	T	A		G	A	G	A	A	G	G	T
12	1	C	G	T	G	T	g/a	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
13	1	C	G		A	T	G	T	G	T	A	C	G	T	A	T	T	A	A	G	g/a	G	A	D
	2	C			A	T	G	T	G	T	A	C	G	T	A		T	A	A	G	A	G	A	D
14	1	C			A	T	G	T	G	T	A	C	G	C	A		G	A	A	A	A	G	G	N1
	2	C						T	G	T	A	C	G	C	A		G	A	A	A	A	G	G	N1
15	1	C	G	T	A	T	g/a	T	G	T			G	C	A	C	G		A	A	A	G	G	I
	2	C			A	T	G	C	G	T	A	C	G	C	A		G	A	A	A	A	G	G	N1
16	1	C	G	T	G	T	g/a	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
	2	C	G	T	G	T	g/a	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	U
17	1	C			A	T	G	T	G	T	A	C	G	T	A		T	A	A	G	A	G	A	D
	2	C			A	T	G	T	G	T	A	C	G	T	A		T	A	A	G	A	G	A	D

Table S2: Results of quantitative PCR

PCR: L16209/H16303					
Sample	Cycle Threshold (Ct)	Average Ct	Stdev	Stdev/AverageCt %	Copies/uL
RD-1	35.71	37.83	1.84	4.86	2337.10
RD-1	38.89				
RD-1	38.90				
RD-3	38.35	40.29	3.01	7.47	484.09
RD-3	38.77				
RD-3	43.76				
RD-9	38.89	38.93	0.07	0.17	1155.94
RD-9	38.90				
RD-9	39.01				
RD-17	40.32	42.03	0.51	1.22	159.30
RD-17	40.44				
RD-17	40.43				

PCR: L16209/H16348					
Sample	Cycle Threshold (Ct)	Average Ct	Stdev	Stdev/AverageCt %	Copies/uL
RD-1	35.23	35.01	0.51	1.45	4.75
RD-1	35.37				
RD-1	34.43				
RD-3	37.56	37.79	0.31	0.81	0.78
RD-3	37.68				
RD-3	38.14				
RD-9	36.30	36.39	0.32	0.89	1.95
RD-9	36.75				
RD-9	36.12				
RD-17	36.12	36.69	0.49	1.34	1.61
RD-17	36.98				
RD-17	36.96				

Table S3: Description of the Scythian archaeological sites of the Rostov-on-Don area sampled for ancient DNA.

Samples	Location
RD-1	West Rostov-on-Don, Graveyard Livenchovsky V, grave 1
RD-2	Bokovsky district, Razmetniy, graveyard Razmetniy, grave 1
RD-3	Myasnikovsky district, Nedvigovka, graveyard Tsarsky, grave 2
RD-5	Myasnikovsky district, Nedvigovka, graveyard Tsarsky, grave 2
RD-6	Semikarakorsky district, Semikarakorsk, graveyard Donskoy, grave 8
RD-7	East Aksay, graveyard Mukhin II, grave 2
RD-8	West Rostov-on-Don, graveyard Livenchovsky, grave 1
RD-9	West Rostov-on-Don city, graveyard Livenchovsky, grave 3
RD-10	West Rostov-on-Don city, graveyard Livenchovsky, grave 3
RD-11	West Rostov-on-Don city, graveyard Livenchovsky, grave 3
RD-12	Peschanokopsky district, graveyard Novo-Palestinsky II, grave 2
RD-13	Tchelinsky district, Tchelina, graveyard Tchelinsky III, grave 2
RD-14	Aksaysky district, Starocherkasskaya, graveyard Glinishe I, grave 37
RD-15	Aksaysky district, Starocherkasskaya, graveyard Glinishe I, grave 62
RD-16	Aksaysky district, Starocherkasskaya, graveyard Glinishe I, grave 89
RD-17	Belokalitvensky district, Krasnodonetskaya, graveyard Chastiye Kurgani, grave 2

Table S4: Description and references for the comparative population datasets for Principal Component Analyses. Population abbreviations followed by an asterisk signifies that the corresponding populations were part of the European pool used for comparison of ‘Western’ and ‘Eastern’ haplogroups on Figure 1.

Population name	Abbreviation	Latitude	Longitude	N	Reference
Albanians	ALB*	41.4	19.8	281	Belledi et al., 2000, Bosch et al., 2006, Balanovsky, personal communication
Aleuts	ale			199	Volodko et al., 2008
Altaians Telenghits	alt2	50.0	88.0	71	Derenko et al., 2007
Altaians Tubalars	alt3	52.0	87.0	72	Derenko et al., 2007
Altaians-Kizhi	alt1	50.5	86.0	90	Starikovskaya et al., 2005
Armenians	ARM	40.2	44.5	192	Metspalu et al., GenBank
Aromuns	aro	41.4	21.3	133	Bosch et al., 2006
Austrians	AUT	47.4	11.4	117	Parson et al., 1998, Handt et al., 1994
Azeris	AZE	40.0	48.0	88	Quintana-Murci et al., 2004, Richards et al., 2000
Bashkirs	BA	54.8	56.0	207	Bermisheva et al., 2002
Bosnians	BIH*	43.9	18.4	322	Malyarchuk et al., 2003, Harvey Unpublished
Bulgarians	BGR*	42.9	26.1	141	Calafell et al., 1996, Richards et al., 2000
Buryats	BU	53.0	110.0	411	Derenko et al., 2002, Derenko et al., 2007; Starikovskaya et al., 2005
Byelorussians Bobruisk	BEL*	53.1	29.2	92	Belyaeva et al., 2000
Byelorussians Brest		52.0	26.8	89	Balanovsky, personal communication
Byelorussians Gomel		52.6	29.7	71	Balanovsky, personal communication
Byelorussians Vitebsk		55.0	29.0	100	Balanovsky, personal communication
Chukchi	CHU	65.0	175.0	262	Volodko et al., 2008, Starikovskaya et al., 1998, Derenko et al., 2007
Chuvash	CU*	56.0	46.8	92	Bermisheva et al., 2002; Richards et al., 2000
Croatians coastal	HRV*	43.3	17.0	96	Babalini et al., 2005
Croatians isles 1		43.3	16.6	311	Babalini et al., 2005
Croatians isles 2		45.2	14.4	133	Balanovsky, personal communication
Croatians northern		45.6	18.7	294	Balanovsky, personal communication
Croatians southern		42.8	18.0	146	Balanovsky, personal communication
Cyprus	CYP	35.2	33.3	91	Irwin et al., 2007
Czech	CZE*	50.1	14.2	449	Balanovsky, personal communication, Richards et al., 2000, Vanecek et al., 2003, Malyarchuk et al., 2006
English Central	GBR*	53.0	358.0	271	Sykes et al., 2006
English Cornwall		50.5	355.8	92	Richards et al., 2000
English eastern		53.0	1.0	339	Sykes et al., 2006
English northern		56.0	357.0	403	Sykes et al., 2006
Scottish		58.0	355.0	891	Helgason et al., 2001
Wales		52.5	356.0	192	Piercy et al., 1993, Richards et al., 1996
Western isles and Skye		57.5	354.0	230	Helgason et al., 2001
Eskimo	esk			825	Volodko et al., 2008, Saillard et al., 2000, Starikovskaya et al., 1998, Helgason et al., 2005, Simonson et al., Genbank
Estonians	EST	58.4	26.7	662	Balanovsky, personal communication, Sajantila et al., et al., 1995, Sajantila et al., 1996; Pult et al., 1994; Lappalainen et al., 2008
Evenks East	eve	49.0	119.0	92	Derenko, et al., 2007
Evenks western 1		62.0	96.0	142	Starikovskaya et al., 2005, Pakendorf et al., 2006
Evenks western 2		60.0	94.0	73	Derenko et al., 2007
Evens	evn	60.0	140.0	100	Derenko et al., 2002, Tajima et al., 2004
Finns northern	FIN*	64.4	27.5	403	Meinila et al., 2001
Finns southern		62.0	26.0	105	Kittles et al., 1999, Lahermo et al., 1996
France Finistere	FRA*	48.0	355.9	142	Richard et al., 2007, Dubut et al., 2003

Table S4 (continued 1): Description and references for the comparative population datasets for Principal Component Analyses. Population abbreviations followed by an asterisk signifies that the corresponding populations were part of the European pool used for comparison of ‘Western’ and ‘Eastern’ haplogroups on Figure 1.

Population name	Abbreviation	Latitude	Longitude	N	Reference
France Perigord-Limousin		46.1	1.0	72	Dubut et al., 2003
French Bearn, Pyrennees-Atlantiques		43.3	359.5	81	Richard et al., 2007
French Brittany, Loire-Atlantique		47.5	358.3	75	Richard et al., 2007
French central		47.3	0.1	135	Richard et al., 2007, Dubut et al., 2003
French Languedoc, Hérault		43.6	3.9	85	Richard et al., 2007
French Morbihan		47.7	357.0	81	Richard et al., 2007, Dubut et al., 2003
French Normandy		49.3	0.2	85	Richard et al., 2007, Dubut et al., 2003
French Picardie, Somme		49.9	2.2	79	Richard et al., 2007
French Poitou, Vendée		46.9	358.5	80	Richard et al., 2007
French south-eastern		44.6	5.7	83	Richard et al., 2007, Dubut et al., 2003
Georgians	GEO	42.0	44.0	158	Quintana-Murci et al., 2004, Balanovsky, personal communication
Germans Lower Saxony	DEU*	52.6	9.6	700	Pfeiffer et al., 1999
Germans south	GRC*	48.3	11.5	247	Lutz et al., 1998, Richards et al., 1996
Germans west		51.7	7.3	159	Pfeiffer et al., 1999, Baasner et al., 1998
Germans Western Pomerania		54.0	13.0	300	Poetsch et al., 2003
Greek Crete		35.0	25.0	187	Balanovsky, personal communication
Greek northern		40.5	22.9	469	Richards et al., 2000, Bosch et al., 2006, Irwin et al., 2007
Hungarians	HUN*	47.5	19.0	190	Balanovsky, personal communication, Bogaszi-Szabo et al., 2006
Icelanders	ISL	64.5	337.4	448	Richards et al., 1996, Sajantila et al., 1995, Helgason et al., 2000
Iranians central	IRN	32.7	51.7	78	Metspalu et al., 2004, Quintana-Murci et al., 2004
Iranians northwest		37.3	49.6	284	Metspalu et al., 2004, Quintana-Murci et al., 2004
Iranians southwest		33.5	48.3	155	Metspalu et al., 2004, Quintana-Murci et al., 2004
Iraq	IRQ	33.3	44.4	168	Richards et al., 2000, Al-Zahery et al., 2003
Irish	IRL	53.5	350.9	300	McEvoy et al., 2004
Italians central	ITA*	41.9	12.5	183	Babalini et al., 2005, Richards et al., 2000, Tagliabracci et al., 2001
Italians eastern		41.9	14.3	73	Babalini et al., 2005
Italians southern		41.1	15.5	74	Babalini et al., 2005
Italians Tuskany		43.8	11.2	432	Achilli et al., 2007, Francalacci et al., 1996, Varesi et al., GenBank
Italians Veneto		46.0	11.0	68	Mogentale-Profizi et al., 2001
Sicily		38.0	12.9	106	Cali et al., 2001
Jordanians	JOR	31.8	35.8	146	Cabrera et al., GenBank
Kabardians	kab	43.0	43.0	163	Richards et al., 2000, Balanovsky, personal communication
Karelians Aunus	KR	62.0	32.0	218	Lappalainen et al., 2008
Karelians Viena		66.0	32.0	87	Lappalainen et al., 2008
Kazakhs	KAZ	45.0	80.0	125	Comas et al., 1998, Comas et al., 2004; Yao et al., 2000
Kets	ket	45.8	88.0	104	Derbeneva et al., 2002, Balanovsky, personal communication
Khakassians	KK	53.0	90.0	110	Derenko et al., 2002, Derenko et al., 2007; Starikovskaya et al., 2005

Table S4 (continued 2): Description and references for the comparative population datasets for Principal Component Analyses. Population abbreviations followed by an asterisk signifies that the corresponding populations were part of the European pool used for comparison of ‘Western’ and ‘Eastern’ haplogroups on Figure 1.

Population name	Abbreviation	Latitude	Longitude	N	Reference
Khamnigans	kham	53.0	115.0	99	Derenko et al., 2007
Khants	KHM_khan	62.0	72.0	318	Pimenoff et al., 2008, Balanovsky, personal communication
Komi	KO	61.0	53.0	127	Bermisheva et al., 2002
Koryaks	kor	55.0	160.0	147	Schurr et al., 1999
Kurds	kur	37.6	43.1	73	Richards et al., 2000, Quintana-Murci et al., 2004
Latvians	LVA	57.0	24.0	413	Pliss et al., 2005, Lappalainen et al., 2008
Lithuanians Aukstaiciai	LTU	55.0	24.0	90	Balanovsky, personal communication
Lithuanians Zemaiciai		55.5	22.0	90	Balanovsky, personal communication
Mansi	KHM_man	60.0	66.0	161	Derbeneva et al., 2002, Pimenoff et al., 2008
Mari	ME*	56.0	48.1	136	Bermisheva et al., 2002
Mongolians	MNG	45.0	105.0	262	Yao et al., 2004; Kolman et al., 1996; Derenko et al., 2007; Kong et al., 2003
Mordvinians	MO*	54.3	44.5	99	Bermisheva et al., 2002
Morocco	MAR	31.0	353.1	336	Rando et al., 1998, Balanovsky, personal communication
Nenets Asian	NEN_A	65.0	70.0	79	Balanovsky, personal communication
Nenets European	NEN_E	69.0	49.0	128	Saillard et al., 2000, Tonks et al., 2006
Nganasan	nga	69.5	86.2	118	Volodko et al., 2008, Derbeneva et al., 2002, Osipova et al., 2005
Nivkhs	niv	52.0	142.0	113	Starikovskaya et al., 2005, Tajima et al., 2004
Nogays	nog	44.0	47.0	206	Bermisheva et al., 2004
Norwegians	NOR*	59.9	10.6	663	Helgason et al., 2001, Passarino et al., 2002, Richards et al., 2000, Dupuy et al., 1996, Opdal et al., 1998
Ossets northern	SE	43.0	44.5	106	Richards et al., 2000
Ossets southern		42.3	44.0	183	Balanovsky, personal communication
Palestinians	PSE	31.8	35.1	117	Richards et al., 2000
Poles	POL*	52.0	21.0	583	Balanovsky, personal communication, Richards et al., 2000, Malyarchuk et al., 2002
Portugal central	PRT	39.5	352.0	317	Gonzalez et al., 2003, Pereira et al., 2004
Portugal northern		41.3	351.5	271	Gonzalez et al., 2003, Pereira et al., 2004
Portugal southern		37.2	352.2	260	Gonzalez et al., 2003, Pereira et al., 2004
Romanians 1	ROU*	47.6	23.6	92	Richards et al., 2000
Romanians south		44.1	28.6	105	Bosch et al., 2006
Russians Belgorod	RUS	50.8	36.5	148	Balanovsky, personal communication
Russians Cossacks		44.5	40.2	132	Balanovsky, personal communication
Russians Oshevensk		61.8	38.8	76	Belyaeva et al., 2000
Russians Pinega		63.4	46.5	144	Balanovsky, personal communication
Russians Pomors		66.0	42.0	81	Tonks et al., 2006
Russians Rostov		47.2	39.7	111	Kornienko et al., 2004, Richards et al., 2000
Russians Smolensk		53.9	32.9	147	Balanovsky, personal communication
Russians Unja		58.3	44.8	79	Balanovsky, personal communication
Saami	saa	68.9	27.6	559	Sajantila et al., 1995, Dupuy et al., 1996, Kittles et al., 1999, Delghandi et al., 1998, Tambets et al., 2004, Tonks et al., 2006
Sardinia	IT-88	40.0	9.0	115	Richards et al., 2000
Saudi Arabia	SAU	24.6	46.5	325	Balanovsky, personal communication, Abu-Amero et al., 2007
Selkups	sel	65.8	82.5	120	Balanovsky, personal communication
Shors	sho	52.8	87.9	82	Derenko et al., 2007

Table S4 (continued 3): Description and references for the comparative population datasets for Principal Component Analyses. Population abbreviations followed by an asterisk signifies that the corresponding populations were part of the European pool used for comparison of ‘Western’ and ‘Eastern’ haplogroups on Figure 1.

Population name	Abbreviation	Latitude	Longitude	N	Reference
Slovaks	SVK*	48.8	19.3	510	Koledova et al., 2005 Unpublished
Slovenians	SVN	46.1	14.5	233	Malyarchuk et al., 2003, Zupanec et al., 2004
Spaniards Andalusia	ESP*	38.1	355.2	65	Larruga et al., 2001, Corte-Real et al., 1996
Spaniards Cantabria		43.2	356.0	242	Maca-Meyer et al., 2003a
Spaniards central		40.6	356.2	129	Larruga et al., 2001, Corte-Real et al., 1996
Spaniards Galicia		43.0	352.0	135	Gonzalez et al., 2003, Salas et al., 1998
Spaniards north-eastern		41.6	1.9	133	Crespillo et al., 2000, Corte-Real et al., 1996
Swedes central	SWE*	59.3	17.7	105	Balanovsky, personal communication
Swedes Gotland		57.7	18.1	267	Balanovsky, personal communication
Swedes northern		68.0	20.0	97	Balanovsky, personal communication
Swedes southern		55.9	12.7	177	Balanovsky, personal communication
Swiss	CHE	46.7	6.6	230	Dimo-Simonin et al., 2000, Pult et al., 1994
Syrians	SYR	33.6	36.2	169	Balanovsky, personal communication, Richards et al., 2000
Tatars	TA	65.0	52.4	225	Bermisheva et al., 2002
Tofalars	tof	54.8	99.0	104	Derenko et al., 2002, Starikovskaya et al., 2005
Turkey	TUR	39.0	33.0	608	Balanovsky, personal communication, Quintana-Murci et al., 2004, Richards et al., 2000, Calafell et al., 1996
Tuvinians	tuv	51.6	94.4	645	Derenko et al., 2002, Derenko et al., 2007, Balanovsky, personal communication, Tonks et al., 2006, Starikovskaya et al., 2005, Pakendorf et al., 2006
Udmurts	UD	56.6	53.0	109	Bermisheva et al., 2002
Ukrainians Belgorod	UKR	50.4	35.8	95	Balanovsky, personal communication
Ukrainians Cherkasy		49.4	32.1	179	Balanovsky, personal communication
Ukrainians Hmel'nitskaya		49.7	27.3	179	Balanovsky, personal communication
Ukrainians western		49.3	24.0	157	Balanovsky, personal communication
Ulchi-Udegey-Negidal	ulc	50.0	135.0	166	Starikovskaya et al., 2005
Yakuts	SA	65.0	125.0	770	Pakendorf et al., 2003, et al., 2006; Balanovsky, personal communication; Fedorova et al., 2003; Zlojutro et al., 2006; Derenko et al., 2002, Derenko et al., 2007
Yukagir	yuk	65.0	150.0	153	Pakendorf et al., 2006, Volodko et al., 2008

Table S5: Description and references for the comparative populations where haplotypes matching Rostov Scythian haplotypes.

Pool	Population	Reference	N
West Europe	Austrians	Parson et al., 1998	101
West Europe	Britain, Northern Isles	Sykes et al., 2006	18
West Europe	Britain, Northern Isles	Tonks et al., 2006	93
West Europe	Britain, Northern Isles	Helgason et al., 2001	230
West Europe	Britain, Northern Isles	Sykes et al., 2006	236
West Europe	Britain, Northern Isles	Goodacre et al., 2005	500
West Europe	Canarians	Santos et al., 2009	89
West Europe	Canarians	Rando et al., 1999	300
West Europe	English	Tonks et al., 2006	85
West Europe	English	Sykes et al., 2006	163
West Europe	English	Sykes et al., 2006	271
West Europe	Finns	Pult et al., 1994	23
West Europe	Finns	Meinila et al., 2001	403
West Europe	French	Dubut et al., 2003	39
West Europe	French	Richard et al., 2007	44
West Europe	French	Richards et al., 2000	47
West Europe	French	Rousselet et al., 1998	50
West Europe	French	Cali et al., 2001	112
West Europe	German Caucasians	Baasner et al., 2000	101
West Europe	Germans	Brandstaetter et al., 2006	100
West Europe	Germans Central	Baasner et al., 1998	50
West Europe	Germans North	Richards et al., 1996	107
West Europe	Germans North	Poetsch et al., 2003	300
West Europe	Germans North	Pfeiffer et al., 1999	700
West Europe	Germans South	Lutz et al., 1998	198
West Europe	Icelanders	Helgason et al., 2003	552
West Europe	Irish	Sykes et al., 2006	39
West Europe	Irish	McEvoy et al., 2004	300
West Europe	Italians	Turchi et al., 2007	30
West Europe	Italians Central	Babalini ¹ et al., 2005	11
West Europe	Italians Central	Varesi et al., 2005c	61
West Europe	Italians Central	Achilli et al., 2007	86
West Europe	Norse	Helgason et al., 2001	323
West Europe	Portuguese	Brehm et al., 2003	155
West Europe	Portuguese	Pereira et al., 2004	187
West Europe	Sardinians	Richards et al., 2000	115

Table S5 (continued 1): Description and references for the comparative populations where haplotypes matching Rostov Scythian haplotypes.

Pool	Population	Reference	N
West Europe	Scotts	Sykes et al., 2006	123
West Europe	Scotts	Sykes et al., 2006	182
West Europe	Scotts	Sykes et al., 2006	209
West Europe	Scotts	Helgason et al., 2001	891
West Europe	Spaniard	Pinto et al., 1996	18
West Europe	Spaniard	Alvarez et al., 2010	37
West Europe	Spaniard	Gonzalez et al., 2003	43
West Europe	Spaniard	Larruga et al., 2001	50
West Europe	Spaniard	Plaza et al., 2003	98
West Europe	Spaniard	Casas et al., 2006	108
West Europe	Spaniard	Garcia et al., 2010	113
West Europe	Spaniard	VPereira et al., 2010	160
West Europe	Spaniard	Maca-Meyer et al., 2003	242
West Europe	Spaniard	Alvarez-Iglesias et al., 2009	282
West Europe	Spaniard	Alvarez et al., 2006	312
West Europe	Swedes	Tillmar et al., 2008	40
West Europe	Swedes	Lappalainen et al., 2008	307
West Europe	Swiss	Pult et al., 1994	76
West Europe	Swiss	Dimo-Simonin et al., 2000	154
East Europe	Czechs	Vanecek et al., 2003	93
East Europe	Czechs	Malyarchuk et al., 2006	179
East Europe	Estonians	Lappalainen et al., 2008	117
East Europe	Hungarians	Irwin et al., 2007	211
East Europe	Hungarians Transilvania	Brandstaetter et al., 2008	360
East Europe	Ingrians	Lappalainen et al., 2008	38
East Europe	Kalmyks	Nasidze et al., 2005	99
East Europe	Kalmyks	Derenko et al., 2007	110
East Europe	Karelians	Lappalainen et al., 2008	218
East Europe	Latvians	Lappalainen et al., 2008	114
East Europe	Latvians	Pliss et al., 2005	299
East Europe	Lithuanians	Kasperaviciute et al., 2004	30
East Europe	Lithuanians	Lappalainen et al., 2008	163
East Europe	Poles	Grzybowski et al., 2007	87
East Europe	Poles	Grzybowski et al., 2007	166
East Europe	Poles	Malyarchuk et al., 2002	436

Table S5 (continued 2): Description and references for the comparative populations where haplotypes matching Rostov Scythian haplotypes.

Pool	Population	Reference	N
East Europe	Romanians	Richards et al., 2000	92
East Europe	Russians	Grzybowski et al., 2007	78
East Europe	Russians Central	Malyarchuk et al., 2004	71
East Europe	Russians North	Malyarchuk et al., 2004	42
East Europe	Russians South	Malyarchuk et al., 2002	201
East Europe	Russians West	Malyarchuk et al., 2004	68
East Europe	Russians West	Grzybowski et al., 2007	79
East Europe	Russians	Belyaeva et al., 2000	83
East Europe	Russians	Kornienko et al., 2004	86
East Europe	Saami	Tonks et al., 2006	86
East Europe	Slovaks	Metspalu et al., GenBank	129
East Europe	Slovenians	Malyarchuk et al., 2003	104
East Europe	Slovenians	Zupanec et al., 2004	129
East Europe	Ukrainians	Grechanina et al., 2006	240
East Europe	Ukrainians	Malyarchuk et al., 1999	18
Volga-Ural	Bashkirs	Bermisheva et al., 2002	207
Volga-Ural	Chuvash	Bermisheva et al., 2002	56
Volga-Ural	Komi Permian	Bermisheva et al., 2002	66
Volga-Ural	Mari	Bermisheva et al., 2002	136
Volga-Ural	Mordvinians	Bermisheva et al., 2002	99
Volga-Ural	Nenets	Tonks et al., 2006	70
Volga-Ural	Tatars	Bermisheva et al., 2002	225
Volga-Ural	Udmurts	Bermisheva et al., 2002	109
Central Asia	Huis	Yao et al., 2004	45
Central Asia	Jews Middle East	Behar et al., 2008a	17
Central Asia	Kazakhs	Chaix et al., 2007	50
Central Asia	Kazakhs	Comas et al., 1998	52
Central Asia	Kazakhs	Irwin et al., 2010	256
Central Asia	Kyrghyz	Comas et al., 1998	48
Central Asia	Kyrghyz	Irwin et al., 2010	249
Central Asia	Shugnans	Quintana-Murci et al., 2004	44
Central Asia	Tajiks	Comas et al., 2004	20
Central Asia	Tajiks	Derenko et al., 2007	44
Central Asia	Tajiks	Irwin et al., 2010	244
Central Asia	Turkmens	Comas et al., 2004	20
Central Asia	Turkmens	Quintana-Murci et al., 2004	41

Table S5 (continued 3): Description and references for the comparative populations where haplotypes matching Rostov Scythian haplotypes.

Pool	Population	Reference	N
Central Asia	Turkmens	Chaix et al., 2007	51
Central Asia	Turkmens	Irwin et al., 2010	249
Central Asia	Uighurs	Yao et al., 2004	47
Central Asia	Uighurs	Comas et al., 1998	55
Central Asia	Uzbeks	Comas et al., 2004	20
Central Asia	Uzbeks	Chaix et al., 2007	37
Central Asia	Uzbeks	Quintana-Murci et al., 2004	42
Central Asia	Uzbeks	Irwin et al., 2010	53
Central Asia	Uzbeks	Yao et al., 2004	58
West Siberia	Khanty	Osipova et al., 2005	212
West Siberia	Nganasans	Derbeneva et al., 2002	24
West Siberia	Nganasans	Osipova et al., 2005	54
West Siberia	Tatars Siberia	Naumova et al., 2008	29
East Siberia	Buryats	Pakendorf et al., 2003	126
East Siberia	Buryats	Shimada et al., GenBank	134
East Siberia	Buryats	Derenko et al., 2007	295
East Siberia	Evenks	Lebedeva-Seryogin-Poltaraus et al., GenBank	29
East Siberia	Evenks	Pakendorf et al., 2006	32
East Siberia	Evenks	Pakendorf et al., 2006	39
East Siberia	Evenks	Derenko et al., 2007	45
East Siberia	Evenks	Starikovskaya et al., 2005	71
East Siberia	Kazakhs	Gokcumen et al., 2008	237
East Siberia	Kets	Derbeneva et al., 2002	38
East Siberia	Khakassians	Derenko et al., 2002	53
East Siberia	Khakassians	Derenko et al., 2007	57
East Siberia	Khamnigans	Derenko et al., 2007	99
East Siberia	Negidals	Starikovskaya et al., 2005	33
East Siberia	Nivkhs	Starikovskaya et al., 2005	56
East Siberia	Sojots	Derenko et al., 2002	30
East Siberia	Telenghits	Derenko et al., 2007	71
East Siberia	Tofalars	Starikovskaya et al., 2005	46
East Siberia	Tojins	Derenko et al., 2002	48
East Siberia	Tubalars	Starikovskaya et al., 2005	72
East Siberia	Tuvinians	Tonks et al., 2006	45
East Siberia	Tuvinians	Pakendorf et al., 2006	59
East Siberia	Tuvinians	Derenko et al., 2002	90
East Siberia	Tuvinians	Starikovskaya et al., 2005	96

Table S5 (continued 4): Description and references for the comparative populations where haplotypes matching Rostov Scythian haplotypes.

Pool	Population	Reference	N
East Siberia	Tuvinians	Derenko et al., 2007	105
East Siberia	Yakuts	Pakendorf et al., 2003	117
East Siberia	Yakuts	Zlojutro et al., 2006	144
East Siberia	Yakuts	Pakendorf et al., 2006	178
East Siberia	Yakuts	Fedorova et al., 2003	191
East Asia	Ainu	Horai et al., 1996	51
East Asia	Bugan	Li et al., 2007	32
East Asia	China Yunnan minorities	Yao et al., 2002	83
East Asia	Dai	Li et al., 2007	56
East Asia	Daur	Kong et al., 2003	45
East Asia	Han	Zhao et al., 2010	203
East Asia	Han Chinese	Jin et al., 2009	40
East Asia	Han Chinese	Yao et al., 2002	50
East Asia	Han Chinese	Zhang et al., 2003	51
East Asia	Han Chinese	Tajima et al., 2004a	60
East Asia	Han Chinese	Wen et al., 2004	61
East Asia	Han Chinese	Yao et al., 2003	76
East Asia	Han Chinese	Oota et al., 2002	82
East Asia	Han Chinese	Nishimaki et al., 1999	120
East Asia	Han Chinese	Tsai et al., 2001	155
East Asia	Han Chinese	Gan et al., 2008	197
East Asia	Han Chinese	Irwin et al., 2009	377
East Asia	Han Chinese	Niu et al., GenBank	403
East Asia	Japanese	Tajima et al., 2004	82
East Asia	Japanese	Oota et al., 2002	89
East Asia	Japanese	Seo et al., 1998	96
East Asia	Japanese	Tanaka et al., 2004	96
East Asia	Japanese	Horai et al., 1990	111
East Asia	Japanese	Mabuchi et al., 2006	124
East Asia	Japanese	Nagaia et al., 2003	133
East Asia	Japanese	Nishimaki et al., 1999	150
East Asia	Japanese	Budowle et al., 1999	162
East Asia	Japanese	Imaizumi et al., 2002	162
East Asia	Japanese	Maruyama et al., 2004	211
East Asia	Japanese	Asari et al., 2006	217
East Asia	Kazakhs	Yao et al., 2000	53

Table S5 (continued 5): Description and references for the comparative populations where haplotypes matching Rostov Scythian haplotypes.

Pool	Population	Reference	N
East Asia	Koreans	Kong et al., 2003	48
East Asia	Koreans	Jin et al., 2009	51
East Asia	Koreans	Horai et al., 1996	64
East Asia	Koreans	Derenko et al., 2007	103
East Asia	Koreans	Budowle et al., 1999	180
East Asia	Koreans	Lee et al., 1997	318
East Asia	Koreans	Lee et al., 2006	694
East Asia	Li	Peng et al., 2011	86
East Asia	Mongols	Kong et al., 2003b	48
East Asia	Mongols	Yao et al., 2004	49
East Asia	Mongols	Kolman et al., 1996	103
East Asia	Mongols	Cheng et al., 2005	201
East Asia	Oroqen	Kong et al., 2003b	44
East Asia	Taiwanese	Budowle et al., 1999	329
East Asia	Tibetans	Yao et al., 2002	41
East Asia	Tibetans	Qin et al., 2010	46
East Asia	Tibetans	Zhao et al., 2009	83
East Asia	Tibeto-Burmans	Wen et al., 2004a	40
East Asia	Tu	Yao et al., 2002	35
East Asia	Uighurs	Yao et al., 2000a	45
East Asia	Vietnamese	Irwin et al., 2007b	187
East Asia	Zhuang	Zhao et al., 2010a	132
Far East	Beringia	Shields et al., 1993	57
Far East	Nganasans	Volodko et al., 2008	40
Middle East	Afghanistani	Irwin et al., 2010	98
Balkans	Bosnians	Malyarchuk et al., 2003	144
Balkans	Bosnians	Harvey et al., GenBank	178
Balkans	Bulgarians	Calafell et al., 1996	30
Balkans	Bulgarians	Richards et al., 2000	111
Balkans	Croatians	Harvey et al., GenBank	63
Balkans	Croatians	Tolk et al., 2001	108
Balkans	Greeks	Bosch et al., 2006	25
Balkans	Greeks	Vernesi et al., 2001	48
Balkans	Greeks	Irwin et al., 2007	319
Balkans	Jews	Behar et al., 2008	71
Balkans	Macedonians	Zimmermann et al., 2007	200
Caucasus	Abazins	Nasidze et al., 2001	23

Table S5 (continued 6): Description and references for the comparative populations where haplotypes matching Rostov Scythian haplotypes.

Pool	Population	Reference	N
Caucasus	Adygs	Macaulay et al., 1999	50
Caucasus	Adygs	Lebedeva-Seryogin-Poltaraus et al., GenBank	107
Caucasus	Armenians	Metspalu et al., GenBank	192
Caucasus	Azeri	Nasidze et al., 2001	39
Caucasus	Azeri	Richards et al., 2000	48
Caucasus	Cherkessian	Nasidze et al., 2001	44
Caucasus	Georgian	Nasidze et al., 2001	57
Caucasus	Georgians	Quintana-Murci et al., 2004	20
Caucasus	Ingushian	Nasidze et al., 2001	35
Caucasus	Jews Mountain	Behar et al., 2008	74
Caucasus	Kabardinians	Nasidze et al., 2001	51
Caucasus	Kabardinians	Richards et al., 2000	101
Caucasus	Nogays	Bermisheva et al., 2004	206
Caucasus	Ossetians	Richards et al., 2000	106
Middle East	Arabs Gulf	Alshamali et al., 2008	249
Middle East	Arabs Iraq	Al-Zahery et al., 2003	52
Middle East	Arabs Iraq	Richards et al., 2000	116
Middle East	Arabs Saudi	Abu-Amro et al., 2007	109
Middle East	Gilaki	Nasidze et al., 2006	50
Middle East	Iranians	Nasidze et al., 2008	53
Middle East	Iranians	Derenko et al., 2007	82
Middle East	Iranians Central	Metspalu et al., 2004	36
Middle East	Iranians N	Metspalu et al., 2004	226
Middle East	Iranians S	Metspalu et al., 2004	138
Middle East	Jews Middle East	Behar et al., 2008	82
Middle East	Mazandarani	Nasidze et al., 2006	50
Near East	Arabs Bedouin	Behar et al., 2008	58
Near East	Arabs Palestinians	Behar et al., 2008	110
Near East	Druze	Shlush et al., 2008	311
Near East	Greeks	Irwin et al., 2007	91
Near East	Jews Middle East	Behar et al., 2008	12
Near East	Jews Middle East	Behar et al., 2008	135
Near East	Jordanians	Gonzalez et al., 2008	44
Near East	Kurds	Derenko et al., 2007	25
Near East	Turks	Comas et al., 1996	45
Near East	Turks	Quintana-Murci et al., 2004	50

Table S5 (continued 7): Description and references for the comparative populations where haplotypes matching Rostov Scythian haplotypes.

Pool	Population	Reference	N
Near East	Turks	Quintana-Murci et al., 2004	50
South Asia	Akhutota	Kumar et al., 2006	32
South Asia	Bihar, Muslims	Eaaswarkhanth et al., 2008	472
South Asia	India Assam	Cordaux et al., 2003	45
South Asia	India Assam	Cordaux et al., 2003	8
South Asia	India Chotanagpur tribes	Banerjee et al., 2005	80
South Asia	India Gujarat	Metspalu et al., 2004	53
South Asia	India Himachal	Metspalu et al., 2004	37
South Asia	India Kerala	Metspalu et al., 2004	55
South Asia	India Punjab	Kivisild et al., 1999	121
South Asia	India Punjab	Metspalu et al., 2004	109
South Asia	India Tripura	Cordaux et al., 2003	44
South Asia	India Uttar Pradesh	Kivisild et al., 1999	73
South Asia	India West Bengal	Metspalu et al., 2004	55
South Asia	India, unspecified	Budowle et al., 1999	19
South Asia	Maharashtra, India	Basu et al., 2003	30
South Asia	Pakistani Hazara	Quintana-Murci et al., 2004	23
South Asia	Pakistani Pathans	Cordaux et al., 2003	36
South Asia	Pakistani Pathans	Allah Rakha et al., 2010	230
South Asia	Pardhan, South Central India	Thanseem et al., 2006	193
South Asia	Punjab, India	Sharma et al., 2005	35
South Asia	Sri Lanka	Metspalu et al., 2004	50
South Asia	Tharu, India	Fornarino et al., 2009	173
South Asia	Uttar Pradesh, India	Basu et al., 2003	40
South East Asia	Bai	Yao et al., 2002	31
South East Asia	Boan	Wang et al., 2003	10
South East Asia	Chaoshanese south China	Wang et al., 2008	102
South East Asia	Chiang Mai	Zimmermann et al., 2009	190
South East Asia	China Yunnan minorities	Yao et al., 2002	31
South East Asia	Chong	Fucharoen et al., 2001	25
South East Asia	Hakka South China	Wang et al., 2008	170
South East Asia	Hmong-Mien	Wen et al., 2005	6
South East Asia	Indonesia	Tajima et al., 2004	54
South East Asia	Indonesia Ambon	Hill et al., 2007	43
South East Asia	Indonesia Borneo	Hill et al., 2007	68
South East Asia	Indonesia Sulawesi	Hill et al., 2007	38

Table S5 (continued 8): Description and references for the comparative populations where haplotypes matching Rostov Scythian haplotypes.

Pool	Population	Reference	N
South East Asia	Indonesia Sumatra	Hill et al., 2006	24
South East Asia	Laos	Bodner et al., 2011	214
South East Asia	Paiwan	Melton et al., 1998	7
South East Asia	Philippines	Tabbada et al., 2010	466
South East Asia	Thai	Kampuansai et al., 2006	496
South East Asia	Tu	Yao et al., 2002	35
South East Asia	Va	Yao et al., 2002	36
South East Asia	Vietnam	Li et al., 2007	43
South East Asia	Vietnamese	Coble et al., 2007	187

In Chapters One, Two and Three, I described the mitochondrial structures of ancient eastern European populations through time. The deep ancestry of the Saami, European genetic outliers was investigated but no potential could be identified in these ancient populations.

In Chapter Four, I investigate the genetic history of Sardinians, another example of genetic outliers in Europe. The geographical context of the Sardinian population is different from that of Saami of northern Europe in that they inhabit an isolated island in the middle of the Mediterranean Sea. Ancient DNA is used here in order to provide indications about the timescale of this isolation.

Chapter Four

Local Mitochondrial Continuity In central Sardinia: Ancient DNA Evidence From The Bronze Age

Abbreviations: ABC, Approximate Bayesian computation; ACAD, Australian Centre for Ancient DNA; A.D., Anno Domini; aDNA, ancient DNA; BayeSSC, Bayesian Serial SimCoal; Ct, threshold cycle; dNTP, deoxynucleoside triphosphate; Exo, exonuclease; F_{ST} , fixation index; HLA, Human Leukocyte Antigen; HVR-I, hypervariable region I; min, minute; mtDNA, mitochondrial DNA; qPCR, quantitative real-time Polymerase Chain Reaction; rCRS, revised Cambridge reference sequence; RPM, revolutions per minute; RSA, Rabbit Serum Albumin; SAP, shrimp alkaline phosphatase; SBE, single base extension; SNP, single nucleotide polymorphism; UV, ultraviolet; yBP, years Before Present.

ABSTRACT

Sardinians, inhabitants of the second-largest island in the Mediterranean Sea, have been extensively studied due to their status as genetic outliers among Europeans. The significant genetic segregation from mainland Europeans is accompanied by a clear diversification within Sardinia. These genetic features are possibly the result of the complex population history of Sardinians, founder events and genetic drift, which may have been accentuated through time by the mating isolation and the small population sizes of the Sardinian population. Population genetic processes make it challenging to reconstruct the population history of Sardinians using only modern genetic data. Therefore, we generated ancient mitochondrial hypervariable region I (HVR-I) sequences from prehistoric human remains in order to provide direct evidence of the past genetic diversity of Sardinians. Ancient genetic data were obtained from 16 Middle Bronze Age individuals from central Sardinia (3,200 - 3,400 years Before Present). The remains were collected from archaeological sites in conditions optimal for preventing contamination by modern DNA and for preserving ancient DNA after excavation. Comparison of the ancient mtDNA sequences with modern data sampled from four regions in Sardinia identified genetic similarities between ancient and modern central Sardinians. Coalescent simulation analyses could not reject local mitochondrial continuity between ancient and present-day central Sardinians. The genetic continuity between ancient and modern central Sardinians as shown by ancient DNA may explain the genetic differentiation of Sardinians from mainland Europeans. Long-term local genetic continuity in Sardinia indeed implies isolation from the demographic processes and events that have shaped the mtDNA gene pool of modern-day mainland Europeans.

INTRODUCTION

Sardinians, European genetic outliers

Sardinia is a large island of the Mediterranean Sea located 200 km west of the Italian Peninsula. Present-day Sardinians, similarly to Saami, Basques and Icelanders, were described as European genetic outliers, initially on the basis of classical markers (Cavalli-Sforza et al., 1993; Piazza, 1993). Although Sardinians were shown to belong to the European genetic cluster (e.g., from the study of autosomal loci: Rosenberg et al., 2002), their differentiation from Europeans was clearly confirmed by the study of various genetic markers: classical genetic markers (blood proteins; Memmì et al., 1998), the HLA system (Contu et al., 1992; Vona, 1997a; Grimaldi et al., 2001; Lampis et al., 2000), autosomal markers (e.g., Calò et al., 2003, Falchi et al., 2004; Pugliatti et al., 2006), mitochondrial DNA (mtDNA; Barbujani et al., 1995; Malaspina et al., 2000; Morelli et al., 2000; Richards et al., 2000; Fraumene et al., 2003; Fraumene et al., 2006; Falchi et al., 2006) and Y-chromosome polymorphisms (Semino et al., 2000; Scozzari et al., 2001; Ghiani & Vona, 2002; Quintana-Murci et al., 2003; Francalacci et al., 2003; Rootsi et al., 2004; Capelli et al., 2006).

Geographical isolation of Sardinia and initial settlement

The geographical isolation of Sardinia in the middle of the Mediterranean Sea is thought to have been the main factor involved in the genetic differentiation of Sardinians from mainland Europeans. Even during the Last Glacial Maximum (~20,000 years before present, yBP), when the sea level was at its lowest, Sardinia remained unconnected from the rest of the continent (Shackleton et al., 1984). This isolation probably explains why anatomically modern humans may have reached Sardinia later than any other large island in the Mediterranean Sea (Sondaar, 1998). Permanent human occupation of the island is thought to date back around 10,000 yBP but the origin and process of the initial peopling of Sardinia still remain unclear (Webster, 1996).

Archaeology and history of Sardinia

The archaeology of Sardinia is marked by rich Neolithic cultures (San Michele/Ozieri, megalithic cultures; 5,800 - 5,200 yBP; Calò et al., 2008). Neolithic communities are thought to have been scattered across Sardinia and to have engaged

limited contact with populations outside Sardinia (Webster et al., 1996). However archaeological studies revealed the existence of long-distance trade between Sardinia and mainland Europe during the Neolithic. Interactions around the Mediterranean Basin were shown by the retrieval, outside Sardinia, of artefacts made of obsidian stone, a volcanic glass used in arrowheads, spearheads and cutting utensils. Obsidian stones were shown to have been transported from Sardinia to mainland Europe and the precise geographical origin of some of the obsidian artefacts could be located in Monte Arci, Oristano province, western Sardinia (Tykot, 2002).

The Bronze Age archaeological record in Sardinia (3,800 – 2,238 yBP) is dominated by the Nuragic culture. The name of this culture was derived from the term ‘nuraghe’, which are stone towers characteristic of this period. During the Bronze Age, Sardinians underwent an increase in population density and intensified their contacts with populations of the Mediterranean area (Webster, 1996).

After the Bronze Age, the Sardinian history was marked by multiple invasions, raids and occupations by the Romans (1,545 - 2,238 yBP), the Phoenicians (2,100 yBP), the Vandals (1,545 yBP.), the Byzantines (1,467 yBP), Moorish and Berbers (500 - 1,300 yBP), the Pisans (700 – 800 yBP), the Genoans (600 – 700 yBP), the Spanish (200 - 700 yBP), the Austrians (200 – 3000 yBP), and the Savoyans (200 – 3000 yBP; Dyson & Rowland, 2007; Calò et al., 2008).

Genetic differentiation between modern-day Sardinians and Europeans

Invasions of Sardinia from various origins could have led to the homogenisation of the Sardinian gene pool with that of Europeans. However, these historical occupations have never represented real colonisations and genetic data confirmed that these events had a limited impact on the modern Sardinian gene pool (Moral et al., 1994).

The mtDNA haplogroups detected in present-day Sardinians fall within the European mtDNA diversity. However, significant differences in mtDNA haplogroup frequencies were shown to be responsible for genetic differentiation between Sardinians and Europeans (Morelli et al., 2000). At the haplotypic level, the outlier status of Sardinians when compared to Europeans was strengthened by the observation of distinct mtDNA haplotypes that are rare outside Sardinia (Morelli et al., 2000). On the paternal side, Y-chromosome haplotypes - notably haplogroup I2a1 (I-M26)

lineages - also showed distributions virtually restricted to Sardinia (Semino et al., 2000; Quintana-Murci et al., 2003; Rootsi et al., 2004).

The current debate about the genetic origins of Sardinians involves two opposing views: the first supports the modern Sardinian mtDNA gene pool having been mainly influenced by Upper Palaeolithic post-glacial recolonisation from the Franco Cantabrian refugium (north Spain/south France) around 15,000 yBP (Falchi et al., 2006), the second favours the demographic expansion from the Near East associated with the Neolithisation of Europe ~10,000 yBP (transition to agricultural lifestyle; Fraumene et al., 2003) as a main source of Sardinia's mtDNA diversity. The Sardinian-specific mtDNA lineage, U5b3a1a may be a signal of this Upper Palaeolithic expansion reaching Sardinia (Pala et al., 2009). This lineage appears to be of importance with regard to understanding the population history not only of Sardinians, but of Europeans as a whole. The mtDNA U5b3 clade could indeed represent a rare example of the role of the Italian Peninsula refugium in the post-glacial recolonisation of Europe. The contribution of the Italian Peninsula is however thought to be marginal compared to that of the Franco-Cantabrian refugium, as indicated by the low frequencies of the U5b3 sub-clade in Europe. In contrast, enrichment of this lineage in Sardinia might be the result of founder effect (Pala et al., 2009).

Mitochondrial differentiation among Sardinians

A significant level of mtDNA genetic heterogeneity was observed among regions within Sardinia (Morelli et al., 2000; Fraumene et al., 2003). The northern part of Sardinia, and in particular the region of Gallura (Olbia Tempio region; Figure 1), may have been subjected to high gene-flow from continental Italy, as reported in historical records. The Gallura population was indeed shown to present a more recent pattern of mtDNA variability, i.e., greater variability in mtDNA haplogroups and haplotypes, than in other parts of Sardinia (Morelli et al., 2000). In contrast, central Sardinians, and in particular inhabitants of the Ogliastra province, were shown to display a limited number of mtDNA lineages and to be one of the most homogeneous populations in Europe. This region was hypothesised to have retained ancient genetic traits due to its isolation from foreign occupations and from the north of Sardinia (Fraumene et al., 2003; Fraumene et al., 2006). The hypothesised genetic continuity in central Sardinia does not imply that the mtDNA structure of modern-day Ogliastra

should be considered as representative of the mtDNA diversity of ancient Sardinians. The gene pool of central Sardinians may indeed have been considerably reshaped by the demographic processes that have acted on the Sardinian population since prehistoric times, and that may be responsible for the genetic heterogeneity observed within modern-day Sardinians.

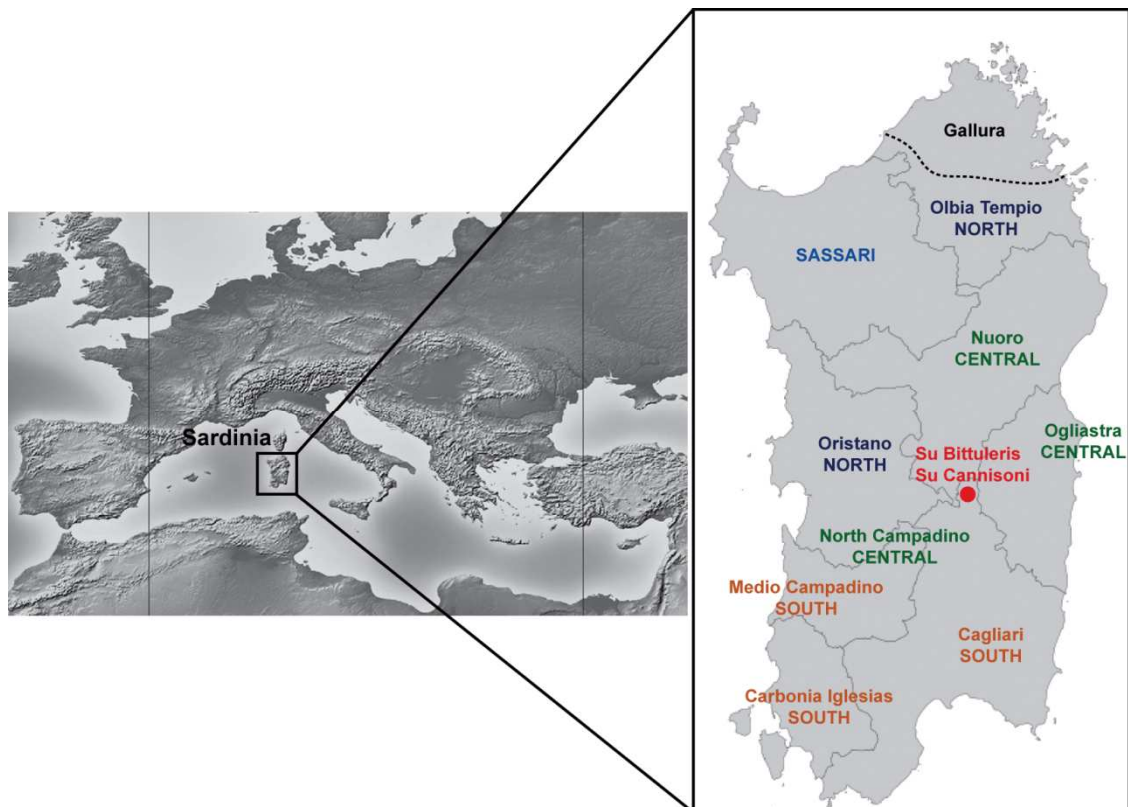


Figure 1. Map of Sardinia showing the location of Sardinia, the Gallura region, the archaeological sites sampled for ancient DNA (in red), and the modern populations used for comparison.

Impact of long-term demographic processes on the Sardinian gene pool

Several demographic factors, including founder effect and genetic drift, were suggested to have participated in the complex population history of Sardinians, their differentiation within Sardinia and from Europeans (Moral et al., 1994). In particular, genetic drift might have been accentuated by the geographical characteristics of the island, the history and cultural features of Sardinian communities that may have favoured isolation: habitat fragmentation, low population densities and migration rates

due to Sardinia's mountainous geography as well as linguistic/cultural barriers. As a consequence, significant endogamy, i.e., marriage within the group of origin, is thought to have driven loss of mtDNA lineages by genetic drift (Calò et al., 2008). Negligible matrimonial movement was supported by the study of matrimonial structures, especially in the mountainous areas of central Sardinia where a maximal rate of endogamy was calculated to be 90% between 1800 and 1980 Anno Domini (A.D.; 20 – 200 yBP; Vona, 1996). Consanguineous marriage was also shown to have been common during the last centuries (Moroni, 1972). Moreover, genetic drift in Sardinia has possibly been intensified by small population sizes, which are historically documented (Webster, 1996). The most important reduction in population size occurred 600 yBP because of food shortage, famine and as a result of a Black Plague epidemic that led to the Sardinian population size to be divided by two (Vona et al., 1997b). In the last 200 years, the Sardinian population has undergone a seven-fold census population size increase from 220,000 in 1677 A.D. (Vona et al., 1997b) to 1,632,000 in 2001 A.D. (source: Istituto Nazionale di Statistica, Italian National Institute of Statistics, 2001), due to demographic growth and immigration.

The reconstruction of population history based on modern biological data alone is considered problematic because founder effects and genetic drift affect biological markers. The study of ancient human remains can overcome this problem as it allows biological changes to be followed through time (as seen in Chapter One, Two and Three).

Biology of prehistoric Sardinians: cranial morphology and mtDNA

Biological continuity between five ancient Sardinian groups from the Late Neolithic to the Nuragic period was investigated using craniofacial morphometric data (D'Amore et al., 2010). Comparison with data from modern-day Sardinians and Italians revealed a genetic continuity between all Sardinian groups (prehistoric and modern-day) but a clear differentiation from present-day mainland Italians, as demonstrated by modern-day genetic markers (see references above). Craniometric similarities between modern-day Sardinians and Palaeolithic/Mesolithic individuals from mainland Europe (Northern Europe, Spain, Italy, and France) suggested a preservation of pre-Neolithic genetic traits in the modern-day population of Sardinia (D'Amore et al., 2010).

The craniometric results confirmed some of the conclusions drawn from investigation of the genetic relationship between prehistoric and modern-day Sardinians using ancient mtDNA. A population of 23 Bronze Age and Iron Age Nuragic individuals (2,700 – 3,430 yBP), collected from six sites across Sardinia, was characterised by low mtDNA diversity and by a lack of geographical and temporal genetic structure when compared to modern-day central Sardinians of the Ogliastra province (Caramelli et al., 2007). Preferential genetic continuity between ancient Nuragic individuals and modern-day central Sardinians compared to North Sardinians of Gallura was statistically tested using coalescent simulation analyses and Approximate Bayesian Computation (ABC; Ghirotto et al., 2010). Genetic discontinuity with North Sardinians might have been introduced by immigration from north Italy into north Sardinia, as described in historical records (Morelli et al., 2000), whereas genetic continuity in central Sardinia was probably a consequence of mating isolation, i.e. absence of genetic influx from continental Italy.

Here, we broaden the characterization of mtDNA diversity in prehistoric central Sardinia and further investigate maternal genetic continuity in this region since the Bronze Age. This was achieved by analysing mtDNA hypervariable region I (HVR-I) sequences of 16 Bronze Age human remains. I collected these samples in clean conditions from two cave sites in the Commune of Seulo, central Sardinia (Figure 1). Confirmation or further resolution of the mtDNA haplogroups determined by HVR-I sequencing was obtained by typing 22 coding region single nucleotide polymorphisms (SNPs) using SNaPshot technology as described in Haak et al., 2010. Further resolution within haplogroup H was obtained by typing coding region diagnostic SNPs for sub-haplogroups H1 and H3. The genetic structure of ancient central Sardinians was compared to the mtDNA makeup of modern-day Sardinians by haplogroup frequency comparison, and haplotype sharing analysis. Local genetic continuity between ancient and present-day central Sardinians was tested using coalescent simulation analysis with the program Bayesian Serial SimCoal (BayeSSC; Anderson et al., 2005) in an ABC framework (Beaumont et al., 2002).

MATERIAL AND METHODS

Archaeological context

We collected the samples during the excavation of two archaeological sites located in the Commune of Seulo, central Sardinia, Italy, coordinates: 39°52'N 9°14'E. The excavations were undertaken in August 2009 by Robin Skeates, Department of Archaeology, University of Durham, United Kingdom and Guiseppina Gradoli, COMET, Valorizzazione Risorse Territoriali, Sardinia, Italy. This project was funded by the British Academy, the Fondazione Banco di Sardegna, and the Prehistoric Society. The fieldwork was undertaken with the permission of the Direzione Generale per i Beni Archeologici (Roma), and in collaboration with the Soprintendenza per i Beni Archeologici della Sardegna and the Soprintendenza per i Beni Archeologici per le Provincie di Sassari e Nuoro. The excavation report for the 2009 campaign (Skeates, 2009) can be found on the University of Durham website: <http://www.dur.ac.uk/archaeology/research/projects/?mode=project&id=409>.

Two sites were sampled for aDNA: the rock cavity of Su Bittuleris and the open rock shelter of Su Cannisoni (Riparo sotto roccia Su Cannisoni). The Su Cannisoni site is part of the main band of rock shelter of the the Pissu is Ilippas hill and the Su Bittuleris is located just above the Su Cannisoni shelter.

The Su Bittuleris site is an 8 m-wide, 65 m-deep single chambered cave, which was known in the past as 'Su Omu'e is Ossu'. The soil of the site was composed of a mixture of recent dark brown organic soil, finer grey soil derived from the prehistoric mortuary deposit and yellow-brown silt derived from the erosion of the cave walls. The surface deposit was disturbed over a depth of around 0.22 m against the foundation of the dry-stone wall (recent). Rich archaeological deposit was collected from a 1 m-by-1 m grid square. The sampled cultural material contained a substantial quantity of human bones (large to highly fragmented), a few animal bones, fragments of ware, 16 obsidian artefacts (arrowheads), a bone pendant, and a shell bead. Human remains were collected by Jessica Beckett, osteologist, and stored in clean plastic bags. From these, a collection of 34 human teeth, identified as human and selected on the basis of their apparent good preservation, was sampled for aDNA analysis. Sampling was carried-out in conditions designed to limit contamination with exogenous modern DNA, i.e. wearing protective equipment (full-body suit, face mask

and gloves), sterilizing the collecting material with bleach, and without washing the samples.

The archaeological excavation of the Su Cannisoni rock shelter uncovered ware fragments, pieces of charcoal, a blade fragment of obsidian, as well as animal (sheep and goat) and human remains. These were found naturally cemented by flowstone at the bottom of an artificial pile of stones. Human remains consisted of a large set of disarticulated and fragmented human bones (including long bones and fragments of a sub-adult skull), two poorly preserved human skulls, and loose sub-adult teeth. Three of these loose teeth were collected for aDNA analysis. The rest of the human remains were excluded due to their apparent poor preservation. All artefacts and remains excavated at the Su Cannisoni site showed signs of severe weathering due to the exposed position of the rock-shelter.

Osteological analysis of the human remains suggested that they had been deposited as part of secondary burial practices. Movements of bones between the Su Bittuleris rock cavity and the Su Cannisoni rock shelter below it were demonstrated by the presence of vertebrae believed to belong to a same arthritic individual in both caves. Radiocarbon-dating of the Su Bittuleris and Su Cannisoni sites was carried out at the University of Oxford Radiocarbon Accelerator Unit. The following dates were obtained: $3,398 \pm 26$ uncalibrated radiocarbon yBP for an adult long-bone fragment from the Su Bittuleris site (OxA-22193) and $3,220 \pm 28$ uncalibrated radiocarbon yBP for a jaw bone sample from the Su Cannisoni shelter (OxA-22194).

Ancient DNA extraction

Genetic analyses of the ancient Sardinian samples were carried out at the Australian Centre for Ancient DNA (ACAD), University of Adelaide. In order to identify possible DNA contaminants arising from the ACAD, one tooth (individual 8354) was sent for independent replication to the Molecular Anthropology/Palaeogenetics Unit, University of Florence, Italy (David Caramelli, Alessandra Modi, Martina Lari).

The protocols used for sample decontamination, preparation, digestion and silica-based DNA isolation were identical as those described in the 'Material and Methods' section of Chapter Two.

Hypervariable-Region I sequencing and coding region GenoCore22 typing

The HVR-I sequencing (positions 15997 to 16409), sequencing of the SNPs at positions 3010 (haplogroup H1) and 6253 (haplogroup H3) and GenoCore22 typing of coding region SNPs were performed according to the same protocols as those described in Chapter One. Primers sequences can be found in Table S1, results of the GenoCore22 typing in Table S2.

Cloning

PCR products obtained for individual 8354 were cloned at the Florence Molecular Anthropology/Palaeogenetics Unit the TOPO-TA cloning kit (Invitrogen®) according to the manufacturer's protocol.

Comparative mtDNA dataset of modern-day Sardinians

The mtDNA data obtained for the ancient central Sardinian samples were compared to data from modern-day Sardinian populations. This modern-day dataset contained 766 HVR-I sequences sampled from four populations: north Sardinia (Olbia Tempio and Oristano provinces; N = 106), the Sassari province (N = 440), central Sardinia (Ogliastra, Nuoro and north Campadino provinces; N = 91) and south Sardinia (Cagliari, Carbonia Iglesias and Medio Campadino provinces; N = 129). These mtDNA HVR-I sequences were obtained by Antonio Torroni (Department of Genetics and Microbiology, University of Pavia, Italy) as part of a screening for rare variants amongst the mtDNA diversity in Sardinia (as described in Pala et al., 2009). This dataset was not published and was used here by courtesy of Antonio Torroni. The procedures involved in obtaining informed consents from Sardinian individuals sampled for DNA and in the analysis of their mtDNA were described in Pala et al., 2009.

Coalescent simulations

We tested whether genetic continuity between ancient and modern-day central Sardinia could be statistically rejected. This was achieved by simulating genealogies according to a pre-defined demographic model using the program BayeSSC. Distributions of population statistics were calculated for the simulated genealogies and compared to the statistics directly calculated from the haplotypic data obtained for ancient and present-day central Sardinians. Comparison between simulated and

observed population statistics allowed the fit of the tested model (genetic continuity) to be assessed.

Demographic models were simulated using the program BayeSSC. Sequence evolution was modeled using the following parameters: 25 years for the generation time, 7.5×10^{-6} substitutions per site per generation (Ho et al., 2008) for the mutation rate, 0.9841 for the transition/transversion ratio, 0.205 and 10 for the theta and kappa parameters of the gamma distribution of rates along the sequence.

The simulated model considered a single population of central Sardinians that could be sampled at two points in time: the present-day and the Bronze Age. The effective population size of the simulated central Sardinian population was allowed to evolve exponentially (expansion or contraction). The values of the corresponding growth rate were drawn from a uniform prior distribution, as were the values for the modern-day central Sardinian effective population size. In a first series of simulations, we explored present-day effective population sizes of central Sardinians between 500 and 200,000 (reported census population size of 322,648 inhabitants, source: Istituto Nazionale di Statistica, Italian National Institute of Statistics, 2001) and exponential growth rates between -0.025 and 0.025.

Simulated population measures were calculated for 100,000 simulated genealogies using BayeSSC (freely available at <http://www.stanford.edu/group/hadlylab/ssc/index.html>). Six of these simulated population measures (for each population: intra-population pairwise differences, nucleotide diversity and haplotype diversity) were selected and compared to the observed measures using a script written by Christian Anderson and available at: <http://www.stanford.edu/group/hadlylab/ssc/index.html>. Observed population measures were calculated using a program written by Christian Anderson and available on request.

The simulated and observed measures were compared using ABC (Beaumont et al., 2002, Ghirotto et al., 2010). The 1% of the simulations for which population parameters exhibited the smallest Euclidian distance with observed population measures was retained to construct posterior distributions. On the basis of these posterior distributions, additional simulations were carried out using narrowed priors for effective population sizes of central Sardinians between 500 and 100,000 and for exponential growths between -0.025 and 0. From the posterior distributions obtained from this second round of simulations, the population parameter values that optimized

the likelihood of a given model were estimated. These values were 11,000 for the modern-day effective population size and -0.012 for the growth rate (negative growth rate indicates expansion).

The estimated effective population size (11,000) and growth rate (-0.012) were then implemented in models in place of the previous priors and 10,000 genealogies were generated in BayeSSC. The observed F_{ST} value computed between ancient and modern-day central Sardinians was compared to the distribution of the F_{ST} values obtained in the 10,000 genealogies.

All the analyses of the BayeSSC simulation output were performed in R version 2.12 using a script written by Christian Anderson and available at: <http://www.stanford.edu/group/hadlylab/ssc/index.html>.

Phylogenetic network

Ancient central Sardinian HVR-I sequences were compared to sequences obtained from 23 Nuragic individuals from Sardinia (2,700 – 3,430 yBP) and previously published in Caramelli et al., 2007. A network representing the phylogenetic relationships between haplotypes was constructed using the program Network (Bandelt et al., 1995) before being re-drawn by hand.

RESULTS/DISCUSSION

Amplification success and authentication

Among the 34 Sardinian tooth samples subjected to DNA extraction, reliable HVR-I data could be obtained for 16 teeth out of 34 (47%) and these were included in the dataset for statistical analyses (Table 1). Twelve samples (35%) yielded no amplification product for the four HVR-I fragments targeted. Six samples (18%) yielded sporadic HVR-I sequences that could not be reproduced. The low amplification success rate observed might be explained by the relatively warm climatic conditions in central Sardinia, which are not optimal for DNA preservation. The relatively poor preservation of the samples could be assessed from macroscopic observation of the samples and comparison with a range of ancient specimens previously sampled for aDNA at the ACAD (Figure S1).

Table 1. Result overview for ancient mitochondrial DNA typing.

Sites and dates	Samples	HVR-I sequence (np 15,997-16,409) ^b 16,000+	Hg ^c (HVR-I)	Hg ^d (coding region)
Su Cannisoni shelter 3220 ± 28 rad yBP ^a	8354	rCRS ^b	H	H1
Su Bittuleris 3398 ± 26 rad yBP ^a	8358, 8418	126C-292T-294T	T2c	T
	8363, 8415, 8417	rCRS ^b	H	H1
	8368, 8421	298C	V	V
	8394, 8427	224C-311C	K	K
	8399	126C-148T-294T-296T	T2	T
	8405	126C-209C-292T-294T	T2c	T
	8406	126C-234T-294T-296T-304C	T2b	T
	8416	224C-311C-399G	K	K
	8420	189C-223T-278T	X	N
	8423	069T-126C	J	J

^a Dates are uncalibrated in radiocarbon years before present (rad yBP) ^b Variable nucleotide positions (np) when compared to the revised Cambridge Reference Sequence (rCRS, Andrews et al., 1999). Transitions are reported with lower-case letters, transversions with upper-case letters. ^c Haplogroup (Hg) assigned on the basis of Hypervariable Region I (HVR-I). ^d Hg determined by the coding-region GenoCoRe22 assay.

Strict precautions were taken to minimise the risk of contamination by modern DNA and to detect artefactual mutations arising from aDNA degradation. Seven criteria support the authenticity of the mtDNA data are presented here (Willerslev & Cooper et al., 2005; Gilbert et al., 2005).

(1) Samples were collected in exceptional conditions with regard to control of contamination by modern DNA and optimisation of sample preservation. Samples were collected from freshly excavated archaeological sites in virtually modern human ‘DNA-free’ conditions as described above (see Materials and Methods). Samples were stored in constant temperature/humidity conditions within the week of their collection from the sites.

(2) Pre-PCR DNA work was carried out at the ACAD, University of Adelaide, a purpose-built a positive air pressure laboratory dedicated to aDNA studies, which is physically isolated from any molecular biology laboratory amplifying DNA. Routine decontamination of the laboratory surfaces and instruments involves exposure to UV radiation and thorough cleaning using DNA oxidants such as bleach, as well as Decon (Decon labs) and Ethanol. In order to protect the lab environment from human DNA, researchers are required to wear protective clothes consisting of a whole body suit, a

facemask, a face shield, gum boots, and three pairs of surgical gloves that are changed on a regular basis.

(3) Contamination within the laboratory or in the reagents were monitored and controlled by blank controls (one extraction blank for every five ancient samples and two PCR/GenoCoRe22 blank controls for every 6 reactions). When comparing the haplotypes obtained from ancient samples with sequences belonging to the investigators, the only match was the haplogroup H revised Cambridge reference sequence (rCRS; Andrews et al., 1999). Extracts obtained from individuals displaying the rCRS haplotype (individuals 8354, 8363, 8415, 8417) were subjected to additional aDNA amplification and sequencing to confirm the rCRS haplotype. The whole mtDNA genome of individual 8354 was obtained as part of an independent study at the ACAD, allowing comparison with the complete mtDNA genome sequence of the operator harbouring the haplogroup H rCRS HVR-I sequence and the sequences did not match. This supports further the authenticity of the HVR-I haplotypes obtained from the Sardinian remains.

(4) Independent replication of sequencing using a second sample from a single individual could only be carried out for the individual found in the Su Cannisoni site. Three tooth samples of the same sub-adult individual (8354) were collected from this site. The rCRS haplotype of individual 8354 was replicated in an independent aDNA laboratory (Molecular Anthropology/Palaeogenetics Unit, University of Florence, Italy). The samples collected from the Su Bittuleris archaeological site were found as isolated loose teeth, i.e. not associated with any other part of the individuals' skeleton. As a consequence, it was not possible to collect two samples for a same individual in the objective to replicate genetic analyses independently.

(5) Cloning - performed at the University of Florence - was used to verify mutations in one ancient mtDNA haplotype (individual 8354). The sequences of the clones highlighted nucleotide positions modified by post-mortem damage, shown as inconsistent cytosine to thymine or guanine to adenosine base changes (Figure S2).

(6) Artefactual and hybrid sequences potentially arising from various exogenous DNA molecules, DNA degradation or jumping PCR were also tested through multiple replications of each individual HVR-I fragment. Sequences were obtained from at least two independent PCRs from independent extracts from two samples for each individual (i.e., a minimum of four independent PCRs). This strategy was chosen over cloning of single PCR products for most of the individuals examined,

since under low-template conditions, clone sequences from one single PCR can represent a biased distribution of sequences that were selectively amplified from a single highly degraded starting DNA template. We believe that a repetitive approach, based on multiple independent repetitions is a powerful alternative to cloning.

(7) The phylogenetic consistency of the haplotypes and matching haplogroup assignments of both HVR-I data and coding region SNPs, were indicative of the robustness of the mtDNA typing approach presented here.

Overall, considering the criteria and results above, we are confident about the control of contamination within the ACAD and about the authenticity of the genetic data obtained from single samples where replication was impossible.

Comparison with modern-day Sardinian haplogroups

Samples from the Su Bittuleris site were buried as loose teeth; consequently it was not possible to assign samples to particular individuals. It is possible that two samples within the examined sample set originated from the same individual in situations where osteological evidence could not disprove it: e.g., a sub-adult and an adult teeth or two upper-right adult canines cannot belong to the same individual. Moreover, the fact that some individuals could be maternally related might introduce a bias in the statistical analyses of the genetic data obtained for the Su Bittuleris site. Considering these biases, two approaches were used to calculate mtDNA haplogroup frequencies: 1) all haplotypes were taken into consideration (hypothesis that each sample represents a single individual), and 2) haplotypes observed multiple times in the same site (described as 'redundant haplotypes' in Table 2) were counted only once (hypothesis that samples yielding the same haplotype originate from same or related individuals).

The mtDNA haplogroups detected in the 16 ancient individuals of central Sardinia examined here (Table 2) and their frequencies (first number considers all haplotypes/second number excludes redundant haplotypes) were: haplogroups T (31.3/36.3%), H1 (25.0/18.2%), K (18.8/18.2%), V (12.5/9.1%), J (6.2/9.1%), and X (6.2/9.1%). All these haplogroups are common today throughout Europe (Table 2; Richards et al., 1996; Richards et al., 1998; Richards et al., 2000), and in present-day Sardinia (Figure 2; Morelli et al., 2000; Fraumene et al., 2003).

Ancient central Sardinians differed from the modern-day Sardinians (total) by their higher frequencies in haplogroups T, K, V and X, and lower frequencies of

haplogroups H and J (Table 2). Results obtained from analysis of haplogroup frequencies in ancient populations should however be considered with caution, as such analyses are sensitive to biases associated with the characteristics of ancient genetic datasets: small sample sizes, and limitations with the extent to which the sample set represents the ancient populations. The latter can play an important role when studying human populations. Discrepancies between an ancient population and the population represented in a burial can be caused by the selective burial of individuals on the basis of their social status, sex or genetic relationships. These aspects of ancient burials can be difficult to identify and characterise, even when the archaeological context of the burial is well described. As a consequence, biases in ancient haplogroup frequency data can be hard to quantify.

Table 2. Mitochondrial haplogroup frequencies in ancient Sardinians, modern Sardinians and modern Europeans.

	Ancient Central Sardinia	Ancient Central Sardinia	Modern Central Sardinia	Modern North Sardinia	Modern Sassari province	Modern South Sardinia	TOTAL modern Sardinia	Modern Europe
Source	This study		Antonio Torroni, unpublished data					Richards et al., 2000
Comment	All samples	Redundant haplotypes excluded						
N	16	23	91	106	440	129	766	1234
% Hg¹:								
T	31.3	36.3	6.6	13.2	5.9	13.2	8.2	7.0 – 9.0
H	25.0	18.2	39.6	33.0	50.2	46.5	45.9	44.5 - 48.2
K	18.8	18.2	6.6	4.7	3.9	3.9	4.3	5.0 - 6.6
V	12.5	9.1	2.2	6.6	5.5	1.6	4.6	3.9 - 5.4
J	6.2	9.1	22.0	6.6	11.1	12.4	12.0	8.3 - 10.4
X	6.2	9.1	3.3	1.9	2.0	1.6	2.1	1.1 – 2.0
HV	-	-	4.4	3.8	5.7	5.4	5.2	0.5 – 2.0
U5	-	-	7.7	15.1	8.0	9.3	9.2	8.1 - 10.3
W	-	-	1.1	1.9	1.8	0.8	1.6	1.5 - 2.5
Other	-	-	6.6	13.2	5.9	5.4	6.9	0.3 - 18.4

¹ Hg: haplogroup

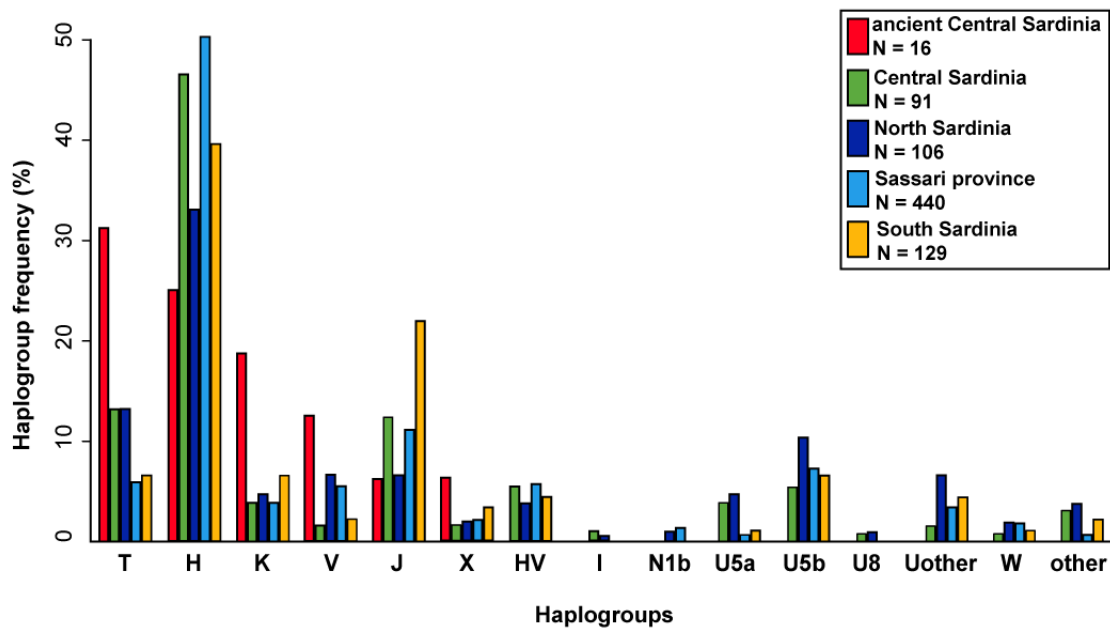


Figure 2. Comparison of mitochondrial haplogroup frequencies in ancient Central Sardinians (all haplotypes included) and modern Sardinians.

Comparison with modern-day Sardinian haplotypes

At the haplotypic level, exact matches could be found in modern-day Sardinians for seven out of the ten ancient central Sardinian haplotypes. Six out of the ten sequences obtained from the ancient central Sardinian sites appeared to be basal haplotypes (haplogroup H rCRS, T2c 16126C-16292T-16294T, V 16298C, K 16224C-16311C, X 16189C-16223T-16278T, J 16069T-16126C). These haplotypes are currently widely distributed in Europe and therefore are not very informative in terms of population geographical affinities. Accordingly, these basal haplotypes have exact matches in all modern-day Sardinian populations (Sassari, north, south and central Sardinia). The non-basal haplotype K 16224C-16311C-16399G (individual 8416) also had matches in all groups of present-day Sardinians. The wide distribution of matches for ancient central Sardinian basal sequences in modern-day Sardinian populations led to similarly high percentages of shared haplotypes for central Sardinia, Sassari, and south Sardinia (Figure 3). However, it should be noted that the larger sample size available for the Sassari region (N = 440 versus N = 91 to 129) may have increased the chance of detecting variants present in the ancient central Sardinian population. In addition, ancient central Sardinians displayed the highest percentage of shared haplotypes with modern-day central Sardinians, despite a smaller population

size for modern-day central Sardinians (N = 91). The genetic affinity of ancient central Sardinians for modern-day central Sardinians identified above on the basis of the analysis of haplogroup frequencies can consequently not be discarded. Of note, the lowest percentage of shared haplotypes was observed for present-day north Sardinians.

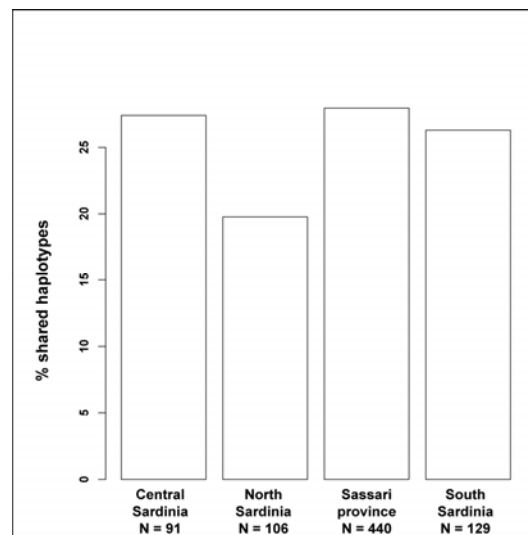


Figure 3. Distribution of haplotypes shared among ancient central Sardinians and present day Sardinians.

The remaining three haplotypes for which no match could be found in any of the modern-day Sardinian populations all belong to haplogroup T2. Private mutations were re-sequenced multiple times in order to verify that they did not represent artefactual mutations as a consequence of aDNA damage. The three haplotypes were subsequently searched in a comparative private database containing 7,789 modern-day Eurasian HVR-I sequences, including 3,558 haplotypes from Italy (Antonio Torroni and Anna Olivieri personal communication). These three haplotypes appeared to be unique or rare. Haplotype T2c 16126C-16209C-16292T-16294T (individual 8405) found no identical match in the modern comparative database. Haplotype 16126C-16148T-16294T-16296T (individual 8399) found an identical match in an Iranian individual and a one-mutation derivative with the additional 16201T mutation in an Italian individual from the Marche region, east Italy. Haplotype 16126C-16234T-16294T-16296T-16304C (individual 8406) found an identical match in a modern-day individual from Bulgaria. A one-mutation derivative displaying the additional 16292T

mutation was also found in two Iranian individuals. Although this haplotype contains the transition at position 16304, which is diagnostic of the T2b sub-haplogroup, the 16292T mutation could however also place this haplotype within the T2c sub-haplogroup. Additional information from the coding region would be necessary to determine whether the ancient central Sardinian and the modern-day Iranian haplotypes belong to the same T2 sub-clade.

These searches show that, despite their low occurrence, haplotypes that were not detected in the modern-day Sardinians were found to be part of the existing mtDNA diversity and are not likely to correspond to sequencing artefacts. The relative uniqueness of these haplotypes makes them, for now, markers of the Bronze Age central Sardinian gene pool.

The absence of overlap in informative haplotypes in ancient and central Sardinians (N = 91) does not provide direct evidence for a particular local genetic affinity between the two populations. These haplotypes being less frequent in ancient central Sardinians, they might have been more sensitive to random extinction through genetic drift, whose effect would have been accentuated by the past small population sizes previously suggested for Sardinians (Webster, 1996). The relatively small sample size available for modern-day central Sardinians also probably limits the detection of rare variants. However, the absence of these haplotypes in other more represented present-day Sardinians (N = 675) could also be the result of a relative isolation of central Sardinia through time that would have prevented the spread of these specific lineages throughout the island.

Overall, the results presented here, based on both haplogroup frequencies and haplotypic data, are in accordance with previous suggestions that central Sardinia has been genetically isolated from the rest of Sardinia and mainland Europe, whereas the northern part of Sardinia was subjected to a larger amount of gene flow from continental Europe (Morelli et al., 2000).

Comparison with other ancient Sardinian populations

Mitochondrial continuity between prehistoric Sardinians and modern-day central Sardinians was previously tested statistically (Ghirotto et al., 2010) using a database of sequences from 23 Bronze Age/Iron Age Nuragic individuals (2,700 – 3,430 yBP; Caramelli et al., 2007). However, the direct comparison of this Nuragic dataset with the ancient central Sardinian dataset presented here is problematic for two

reasons: 1) contrary to the ancient central Sardinians, the Nuragic individuals are not part of a locally defined population, and 2) the genetic data from the Nuragic individuals is not comparable to that obtained from ancient central Sardinians.

(1) The Nuragic dataset has been generated using samples collected from six different sites located all over Sardinia. As a consequence, the Nuragic dataset cannot be considered as representative of a geographically defined population and used for the investigation of local genetic continuity carried out in the present study. The Nuragic dataset could nevertheless be used in the study of local genetic continuity in Sardinia, if an absence of geographical (over the whole of Sardinia) and temporal (in the Bronze/Iron Age) structure could be demonstrated. Analyses of the Nuragic dataset suggested a genetic homogeneity in prehistoric times over the whole of Sardinia (Caramelli et al., 2007). However, the number of samples characterised for each site (one to eight) appears too limited to significantly support genetic homogeneity among the sites. Another argument previously proposed in favour of the geographical and temporal homogeneity in Nuragic individuals was the observation of the same rCRS haplotype in geographically dispersed sites. The H rCRS haplotype was indeed detected in four out of the six sites investigated, in the south, centre and north of Sardinia. However, the rCRS haplotype has previously been described as one of the most widely distributed haplotypes (Richards et al., 2000) and hence does not provide information about the genetic relationships between the populations that share this haplotype. In order to further support genetic homogeneity among prehistoric Sardinian populations a more substantial sampling in the different regions of Sardinia would be required. To conclude, the geographical dispersed samples combined with a lack of evidence for genetic homogeneity in prehistoric Sardinia precludes the integration of the Nuragic dataset (Caramelli et al., 2007) in the present examination of local genetic continuity in central Sardinia.

(2) The level of information with which the Nuragic mtDNA diversity was previously characterised is not comparable with that of the ancient central Sardinian dataset reported here, which relies on both HVR and coding region information. This approach allowed haplogroup assessment in cases where HVR-I was not informative enough. For example, the rCRS HVR-I haplotypes sequenced for individuals 8354, 8363, 8415 and 8417 could belong to several different haplogroups (e.g., haplogroups H and U). In this case, typing of the coding region SNPs permitted the resolution of these haplotypes within haplogroup H. Typing of coding region SNPs was also used

here as a quality control. Matching haplogroup assignments based on HVR-I and coding region data were considered as indicative of the authenticity of the sequences. Phylogenetic consistency, deduced from the comparison of HVR-I sequences and coding region haplogroup assignments can be used to detect amplification artefacts. For example, if HVR-I sequencing fails at detecting a mutation diagnostic of the haplogroup identified on the basis of coding region typing, HVR-I direct sequencing should be repeated in order to verify whether the absence of the particular mutation is authentic or whether it arose from damage or jumping-PCR.

In the case of the Nuragic dataset, only HVR-I sequences are available. Haplogroup assignments were proposed on the basis of the comparison of the Nuragic haplotypes with haplotypes obtained from the modern-day population of Ogliastro, central Sardinia (Fraumene et al., 2003). For the modern Ogliastro haplotypes, corresponding haplogroups had been assigned on the basis of Restriction Fragment Length Polymorphism analyses, but haplogroup assignment of Nuragic haplotypes were not verified experimentally with coding region typing. This approach was problematic, as exemplified by the haplotype 16129C (individuals FL04, ST-15 and ST54), which could either belong to haplogroups H or U.

The first consequence of the lack of information from the coding region in the Nuragic dataset is that haplogroup frequencies the Nuragic population and ancient central Sardinians cannot be directly compared. For example, the frequency of haplogroup H deduced from the Nuragic dataset is 91.3% versus 25.0%/18.2% (all/non-redundant haplotypes) in the ancient central Sardinian dataset. The frequency of haplogroup H in the Nuragic dataset might be over-estimated, as suggested by the phylogenetic network comparison of haplotypes obtained from the two prehistoric Sardinian populations available (Figure 4). The distribution of the Nuragic haplotype is clearly centred on the haplogroup H rCRS, whereas the ancient central Sardinian haplotypic population is more diverse. It could also be noted that some Nuragic haplogroup H haplotypes bear mutations that are diagnostic of other haplogroups detected in the ancient central Sardinian dataset. This is the case for haplotypes 16189C (individual AL07), 16223T (individuals ST08, PE25) and 16278T (individual SE60), whose mutations were found together in the ancient central Sardinian X 16189C-16223T-16278T haplotype (individual 8420). Similarly, the Nuragic haplotype 16126C (SE01, ST30) harbours the mutation diagnostic of the sister clades J/T. It is conceivable that in the absence of information from the coding region,

haplogroups were incorrectly assigned to some of the Nuragic haplotypes. This could only be resolved by an experimental haplogroup assignment based on coding region data for the Nuragic individuals. However, it can be proposed that the Nuragic and ancient central Sardinian populations were part of a common genetic background as closely related haplotypes were detected in these populations, especially within haplogroups H, J/T, T and V.

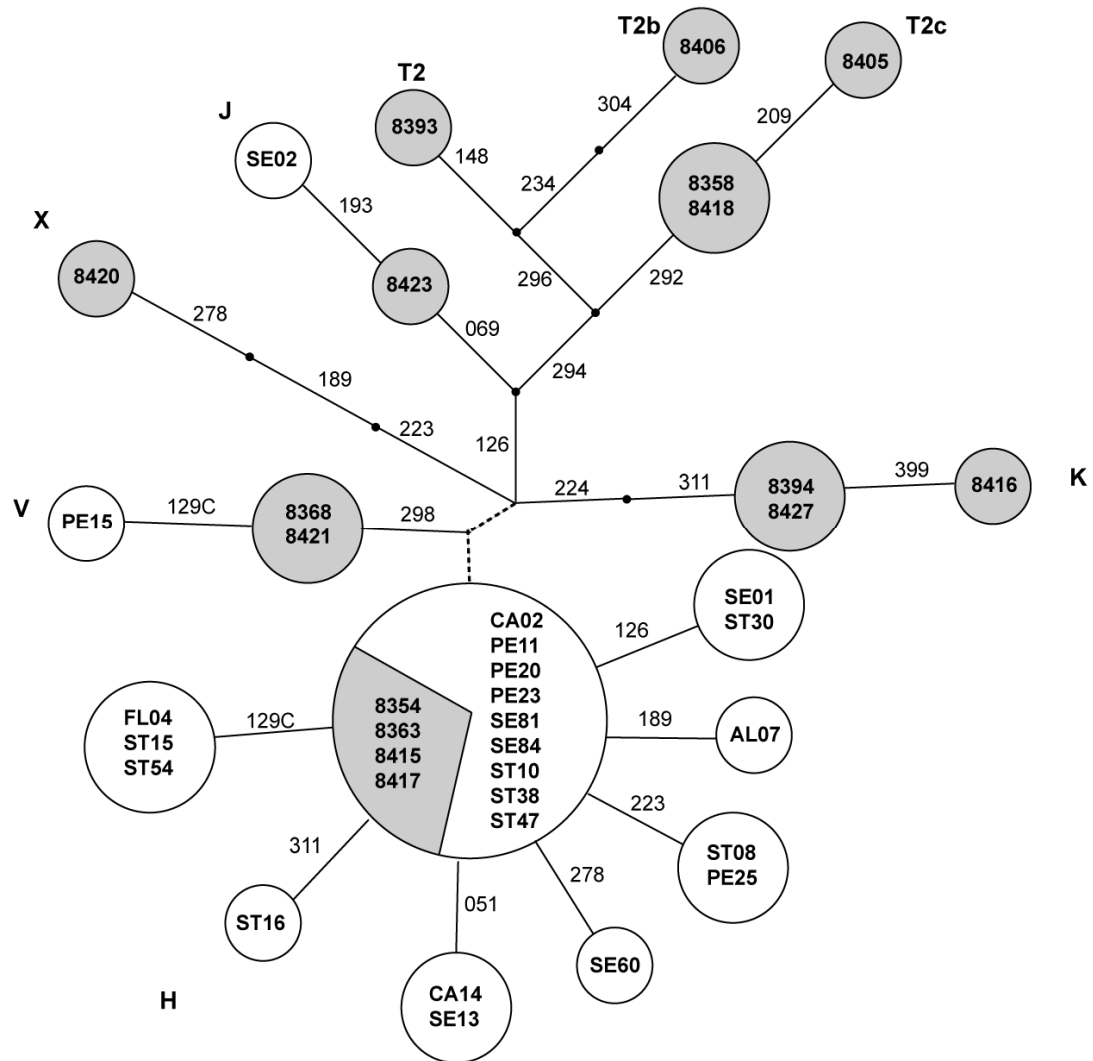


Figure 4. Network representation of hypervariable region I sequences of ancient central Sardinians (this study) and ancient Nuragic Sardinians (Caramelli et al., 2007). Mutations in HVR-I are reported according to the rCRS minus 16,000. The area of the circles is proportional to the number of samples displaying the particular sequence. Grey circles represent sequences obtained in this study. White circles represent sequences obtained in Caramelli et al., 2007. Capital letters indicate transversions. The network was drawn taking into account information from the coding region for the samples of this study (22 coding region SNPs). Haplogroup assessment for Nuragic haplotypes are shown as in Caramelli et al., 2007. Dotted lines represent information from the coding region in the absence of haplogroup-specific mutation in the hypervariable region I.

Test for mtDNA continuity between ancient and modern-day central Sardinians

We tested the hypothesis of mtDNA continuity between the ancient central Sardinian population presented here and modern-day central Sardinians (Figure 5). This was achieved by performing coalescent simulation analyses using BayeSSC. ABC permitted to estimate, population parameters (present-day central Sardinian effective population size and exponential growth rate), which minimised the Euclidian distance between observed and simulated molecular diversity indices (Table 3). An effective population size of 11,000 was estimated for modern-day central Sardinia, a value compatible with previous estimates for the effective population size of the Ogliastra region, central Sardinia (8,947 with a 95% credible interval of 2,645 to 65,724; Ghirotto et al., 2010). Of note, a negative growth rate of -0.012 was estimated from these simulations suggesting that the population of central Sardinia has expanded since the Bronze Age (negative growth rates indicate expansions). Reduction in the population size of Sardinians followed by a recent growth, have been historically recorded. The growth rate estimated in the present coalescent simulation analyses represents a growth rate that has been ‘averaged’ over the past 3,000 years (since the Bronze Age), period over which the central Sardinian population seems to have overall increased in size.

In these simulations, the F_{ST} calculated between ancient and modern-day central Sardinian populations fitted within the distribution of F_{ST} values obtained from the simulations under the model of genetic continuity (Figure 6). As a consequence, the model of local genetic continuity between ancient and modern-day central Sardinians cannot be rejected.

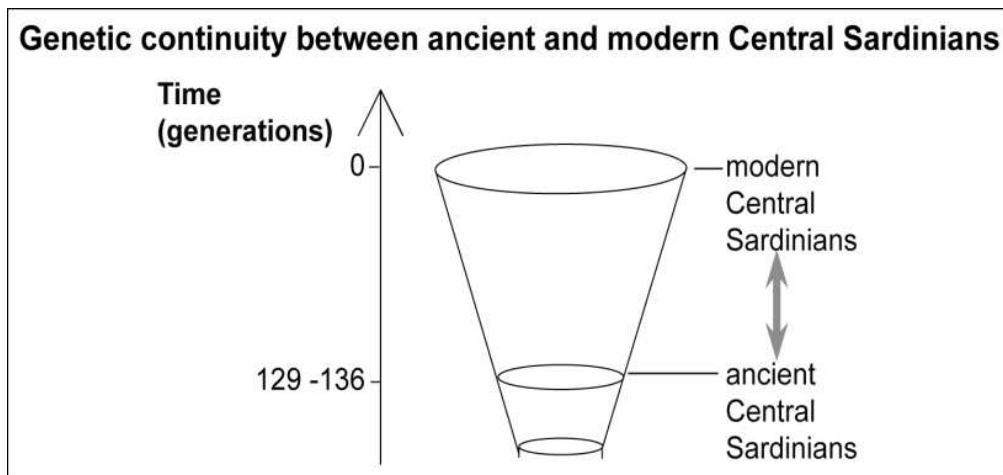


Figure 5. Graphical representation of the demographic model of genetic continuity between ancient and modern central Sardinians simulated in Bayesian Serial SimCoal.

Table 3. Molecular diversity indices and population differentiation statistics used in coalescent simulation analyses.

	Ancient Central Sardinians	Modern Central Sardinians
N	16	91
Intra-population		
Pairwise differences	3.52500	4.82564
Haplotype diversity	0.86700	0.94795
Nucleotide diversity	0.00856	0.01171
Inter-population		
FST	0.01582	

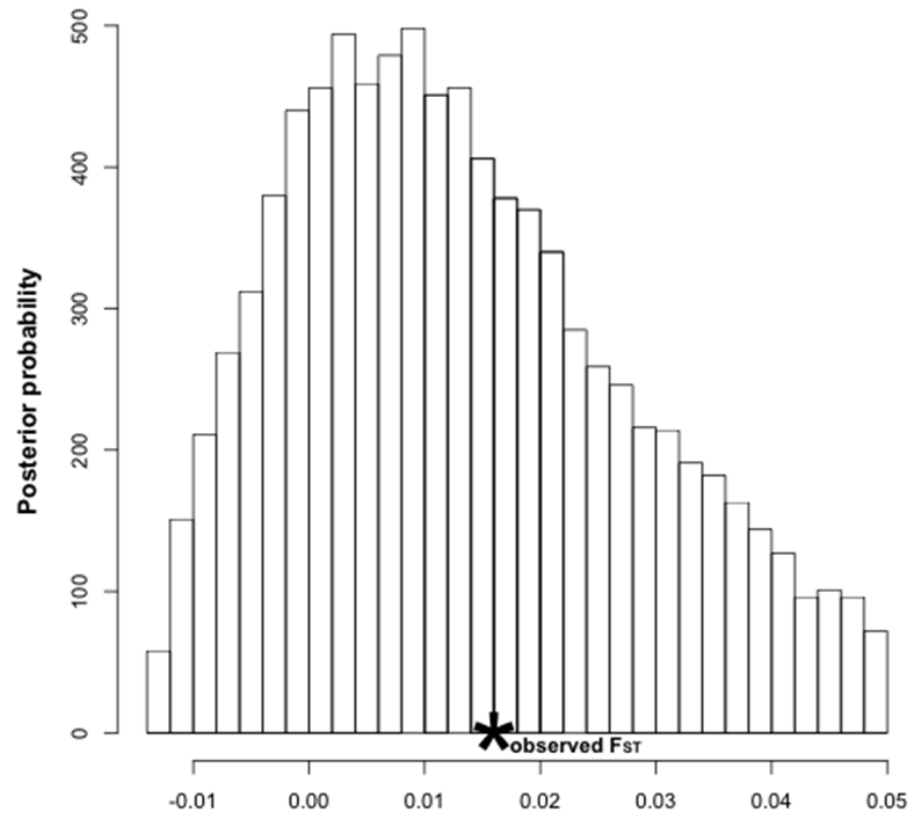


Figure 6. Comparison between the fixation index (F_{ST}) observed between ancient and modern central Sardinians (represented by a star) and the posterior distribution of F_{ST} values obtained through coalescent simulations in Bayesian Serial SimCoal.

CONCLUSION

The mtDNA structure of the Bronze Age population (N=16) of central Sardinia described here fits within the present-day European, and in particular the present-day Sardinian, diversity. Haplogroup frequency and haplotype-based analyses revealed genetic similarities between ancient and present-day Sardinians and in particular present-day central Sardinians. By giving a direct snapshot into the mtDNA variability of central Sardinians in the Bronze Age, aDNA provided direct evidence for local genetic continuity over the past 3,000 years. This result implies that a limited genetic impact on the central Sardinian mtDNA gene pool of the numerous invasions that Sardinia has undergone since the Bronze Age. Genetic continuity between ancient and present-day Sardinians also suggests that central Sardinians have been isolated from the population events (migrations) that have occurred in mainland Europe, thus providing a possible explanation for the outlier status of Sardinians.

We identified local population continuity in central Sardinia since the Bronze Age despite a certain number of methodological limitations associated with the characteristics of the archaeological site (unknown number of independent individuals, absence of second sample for independent replication, poor preservation of the samples). Collection of the samples in conditions minimising the risk of contamination by modern DNA and optimising the post-excavation preservation of aDNA was crucial in obtaining the robust ancient mtDNA dataset presented here.

LIST OF SUPPLEMENTARY MATERIALS

Figure S1: Pictures of selected samples from the Su Cannisoni (A: individual 8354) and Su Bittuleris sites (B-D: individuals 8363, 8417, 8418).

Figure S2: Clone sequences for sample 8354B.

Table S1: Sequences of primers and probes used for sequencing and typing in the mitochondrial hypervariable region I and coding region.

Table S2: Results of SNP typing in the mtDNA coding region using the GenoCoRe22 SNaPshot assay. SNPs typed on the L-strand are reported in capital letters in the reference rCRS profile, whereas SNPs typed on the H-strand are reported in small letters. Missing data signifies allelic dropout or fluorescence signal below the background threshold (100 relative fluorescent units, rfu). ‘g/a’ indicates the presence of a mixed signal for the position interrogated. A mixed signal was repeatedly obtained at position 12612 (haplogroup J) with the detection of an additional G base. However, the rest of the profile never could phylogenetically support the presence of the G base at this particular position. Despite of the allelic dropout consistently observed at position 10034 (haplogroup I), the remaining of the profile and their HVR-I sequence was always consistent.

ACKNOWLEDGEMENTS

We warmly thank Robin Skeates, Department of Archaeology, University of Durham, United Kingdom, and Giuseppina Gradoli, COMET, Valorizzazione Risorse Territoriali, Sardinia, Italy, for their invitation to the 2009 Sardinian excavation campaign. We also thank them for organising the logistic aspects of the excavation, granting us access to the samples and facilitating administrative authorisations and permits. We also acknowledge their scientific contribution as far as the archaeological and geological contexts of the sites are concerned. We also thank Jessica Beckett for her help on the sites and for identification of the teeth. Special thanks to David Caramelli, Alessandra Modi and Martina Lari, Molecular Anthropology/Palaeogenetics Unit, University of Florence, Italy, for performing the independent replication and the cloning. We are particularly grateful to Antonio Torroni, Anna Olivieri, and Maria Pala, Department of Genetics and Microbiology, University of Pavia, Italy, for sharing the unpublished modern-day Sardinian dataset, performing preliminary analyses, and for their very helpful suggestions. We thank Jessica Metcalf for linguistic revision and helpful comments.

REFERENCES

1. Andrews, R., Kubacka, I., Chinnery, P., Lightowlers, R., Turnbull, D., Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23, 147.
2. Bandelt, H.J., Forster, P., Sykes, B.C., Richards, M.B. (1995). Mitochondrial portraits of human populations using median networks. *Genetics* 141, 743-753.

3. Barbujani, G., Bertorelle G., Capitani G., Scozzari R. (1995). Geographical structuring in the mtDNA of Italians. *Proc Natl Acad Sci* 92, 9171-9175.
4. Beaumont, M.A., Zhang, W., Balding, D.J. (2002). Approximate Bayesian computation in population genetics. *Genetics* 162, 2025-2035.
5. Calò, C.M., Varesi, L., Memmi, M., Moral, P., Vona, G. (2003). A pentanucleotide repeat polymorphism (TTTTA) in the apolipoprotein (a) gene--its distribution and its association with the risk of cardiovascular disease. *Coll Antrop* 27, 105-115.
6. Calò, C.M., Melis, A., Vona, G., Piras, I.S. (2008). Sardinian Population (Italy): a Genetic Review. *Int J Mod Anthropol* 1, 1-121.
7. Capelli, C., Redhead, N., Romano, V., Calì, F., Lefranc, G., Delague, V., Megarbane, A., Felice, A. E., Pascali, V. L., Neophytou, P. I., Poulli, Z., Novelletto, A., Malaspina, P., Terrenato, L., Berebbi, A., Fellous, M., Thomas, M. G., Goldstein, D. B. (2006). Population structure in the Mediterranean basin: a Y chromosome perspective. *Ann Hum Genet* 70, 207-225.
8. Caramelli, D., Vernesi, C., Sanna, S., Sampietro, L., Lari, M., Castri, L., Vona, G., Floris, R., Francalacci, P., Tykot, R., Casoli, A., Bertranpetit, J., Lalueza-Fox, C., Bertorelle, G., Barbujani, G. (2007). Genetic variation in prehistoric Sardinia. *Hum Genet* 122(3-4), 327-336.
9. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton Univ. Press, Princeton NJ.
10. Contu L., Arras M., Carcassi C., La Nasa G., Mulargia M. (1992). HLA structure of the Sardinian population: a haplotype study of 551 families. *Tissue Antigens* 40, 165-174.
11. D'Amore, G., Di Marco, S., Floris, G., Pacciani, E., Sanna, E. (2010). Craniofacial morphometric variation and the biological history of the peopling of Sardinia. *Homo* 61, 385-412.
12. Dyson, S.L., Rowland, R.J. (2007). *Archaeology and History in Sardinia from the Stone Age to the Middle ages: Shepherds, Sailors and Conquerors*. Eds UPenn Museum of Archaeology.
13. Falchi A., Giovannoni L., Calò C.M., Piras I.S., Moral P., Paoli G., Vona G., Varesi L. (2006). Genetic history of some western Mediterranean human isolates through mtDNA HVR1 polymorphisms. *J Hum Genet* 1, 9-14.
14. Francalacci, P., Morelli, L., Underhill, P. A., Lillie, A. S., Passarino, G., Useli, A., Madeddu, R., Paoli, G., Tofanelli, S., Calò, C. M., Ghiani, M. E., Varesi, L., Memmi, M., Vona, G., Lin, A. A., Oefner, P., Cavalli-Sforza, L. L. (2003). Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. *Am J Phys Anthropol* 121, 270-279.
15. Fraumene, C., Petretto, E., Angius, A., Pirastu, M. (2003). Striking differentiation of sub-populations within a genetically homogeneous isolate (Ogliastra) in Sardinia as revealed by mtDNA analysis. *Hum Genet*, 1-10.
16. Fraumene, C., Belle, E.M., Castri, L., Sanna, S., Mancosu, G., Cosso, M., Marras, F., Barbujani, G., Pirastu, M., Angius, A. (2006). High resolution analysis and phylogenetic network construction using complete mtDNA sequences in sardinian genetic isolates. *Mol Biol Evol* 23, 2101-2111.
17. Ghiani, M.E., Vona, G. (2002). Y-chromosome-specific microsatellite variation in a population sample from Sardinia (Italy). *Coll Antropol* 26, 387-401.

18. Ghirotto, S., Mona, S., Benazzo, A., Paparazzo, F., Caramelli, D., Barbujani, G. (2010). Inferring genealogical processes from patterns of Bronze-Age and modern DNA variation in Sardinia. *Mol Biol Evol* 27, 875-886.
19. Gilbert, M.T., Bandelt, H.J., Hofreiter, M., Barnes, I. (2005). Assessing ancient DNA studies. *Trends Ecol Evol* 20, 541-544.
20. Grimaldi, M.-C., Crouau-Roy, B., Amoros, J.-P., Cambon-Thomsen, A., Carcassi, C., Orru, S., Viader, C. Contu, L. (2001) West Mediterranean islands (Corsica, Balearic islands, Sardinia) and the Basque population: contribution of HLA class I molecular markers to their evolutionary history. *Tissue Antigens* 58, 281-292.
21. Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., Der Sarkissian, C.S., Brandt, G., Schwarz, C., Nicklisch, N., Dresely, V., Fritsch, B., Balanovska, E., Vilems, R., Meller, H., Alt, K.W., Cooper, A., Genographic consortium. (2010). Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. *PLoS Biol* 8, e1000536.
22. Lampis R., Morelli L., De Virgilis S., Congia M., Cucca F. (2000). The distribution of HLA class II haplotypes reveals that the Sardinian population is genetically differentiated from the other Caucasian populations. *Tissue Antigens* 56, 515-521.
23. Malaspina P., Cruciani F., Santolamazza P., Torroni A., Pancrazio A., Akar N., Bakalli, V., Brdicka, R., Jaruzelska, J., Kozlov, A., Malyarchuk, B., Mehdi, S. Q., Michalodimitrakis, E., Varesi, L., Memmi, M. M., Vona, G., Vilems, R., Parik, J., Romano, V., Stefan, M., Stenico, M., Terrenato, L., Novelletto, A., Scozzari, R. (2000). Patterns of male-specific inter-population divergence in Europe, West Asia and North Africa. *Ann Hum Genet* 64, 395-412.
24. Memmi M., Moral P., Calò C.M., Autuori L., Mameli G.E., Succa V., Varesi L., Vona G. (1998). Genetic structure of southwestern Corsica (France). *Am J Hum Biol* 10, 567-577.
25. Moral, P., Marogna, G., Salis, M., Succa, V., Vona, G. (1994). Genetic data on Alghero population (Sardinia): contrast between biological and cultural evidence. *Am J Phys Anthropol* 93, 441-453.
26. Morelli L., Grosso M.G., Vona G., Varesi L., Torroni A., Francalacci P. (2000). Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia. *Hum Biol* 72, 585-595.
27. Pala, M., Achilli, A., Olivieri, A., Kashani, B. H., Perego, U. A., Sanna, D., Metspalu, E., Tambets, K., Tamm, E., Accetturo, M., Carossa, V., Lancioni, H., Panara, F., Zimmermann, B., Huber, G., Al-Zahery, N., Brisighelli, F., Woodward, S. R., Francalacci, P., Parson, W., Salas, A., Behar, D. M., Vilems, R., Semino, O., Bandelt, H. J., Torroni, A. (2009). Mitochondrial haplogroup U5b3: a distant echo of the epipaleolithic in Italy and the legacy of the early Sardinians. *Am J Hum Genet* 84, 814-821.
28. Piazza A. (1993). Who are the Europeans? *Science* 260:1767-1769.
29. Pugliatti, M., Rosati, G., Carton, H., Riise, T., Drulovic, J., Vécsei, L., Milanov, I. (2006). The epidemiology of multiple sclerosis in Europe. *Eur J Neurol* 13, 700-722.
30. Quintana-Murci, L., Veitia, R., Fellous, M., Semino, O., Poloni, E.S. (2003). Genetic structure of Mediterranean populations revealed by Y-chromosome haplotype analysis. *Am J Phys Anthropol* 121, 157-171.
31. Richards, M., Côrte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H., Sykes, B. (1996). Paleolithic

- and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59, 185-203.
32. Richards, M., Macaulay, V., Bandelt, H., Sykes, B. (1998). Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 62, 241-260.
 33. 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., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Renzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A., Bandelt, H.-J. (2000). Tracing European founder lineages in the near Eastern mtDNA pool. *Am J Hum Genet*, 1251-1276.
 34. Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., Help, H., Bermisheva, M., Kutuev, I., Bara, L., Perici, M., Balanovsky, O., Pshenichnov, A., Dion, D., Grobei, M., Zhivotovsky, L. A., Battaglia, V., Achilli, A., Al-Zahery, N., Parik, J., King, R., Cinnioli, C., Khusnutdinova, E., Rudan, P., Balanovska, E., Scheffrahn, W., Simonescu, M., Brehm, A., Goncalves, R., Rosa, A., Moisan, J. P., Chaventre, A., Ferak, V., Füredi, S., Oefner, P. J., Shen, P., Beckman, L., Mikerezi, I., Terzi, R., Primorac, D., Cambon-Thomsen, A., Krumina, A., Torroni, A., Underhill, P. A., Santachiara-Benerecetti, A. S., Villems, R., Semino, O. (2004). Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet* 75, 128-137.
 35. Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., Feldman, M.W. (2002). Genetic structure of human populations. *Science* 298, 2381-2385.
 36. Shackleton, J.C., van Andel, T.H., Runnels, C.N. (1984). Coastal paleogeography of the central and western Mediterranean during the last 125,000 years and its archaeological implications. *J Field Archaeol* 11, 307-314.
 37. Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., Richards, M. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84, 740-759.
 38. Soares P., Achilli A., Semino O., Davies W., Macaulay V., Bandelt H.J., Torroni A., Richards M.B. (2010). The archaeogenetics of Europe. *Curr Biol* 20, R174-R183.
 39. Sondaar, P.Y. (1998). Palaeolithic Sardinians: Paleontological evidence and methods. In *Sardinian and Aegean chronology*, M.S. Balmuth and R.H. Tykot, eds. Oxford: Oxbow Books, pp. 45-51.
 40. Scozzari, R., Cruciani, F., Pangrazio, A., Santolamazza, P., Vona, G., Moral, P., Latini, V., Varesi, L., Memmi, M. M., Romano, V., De Leo, G., Gennarelli, M., Jaruzelska, J., Villems, R., Parik, J., Macaulay, V., Torroni, A. (2001). Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum Immunol* 62, 871-884.
 41. Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A. S., Cavalli-Sforza, L. L., Underhill, P. A. (2000). The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290, 1155-1159.

42. Tykot, R.H. (2002). Chemical fingerprinting and source tracing of obsidian: the central Mediterranean trade in black gold. *Acc Chem Res* 35, 618-627.
43. Vona G., Calb C.M., Lucia G., Mameli G.E., Succa V., Esteban E., Moral P. (1996). Genetics, Geography and culture: the population of S. Pietro Island (Sardinia, Italy). *Am J Phys Anthropol* 100(4), 461-471.
44. Vona G., Bitti P.P., Succa V., Mameli G.E., Salis M., Secchi G., Calò C.M. (1997a). HLA phenotype and haplotype frequencies in Sardinia (Italy). *Coll Anthropol* 21, 461-475.
45. Vona G. (1997b). The peopling of Sardinia (Italy): history and effects. *Int J Anthropol* 12(1), 71-87.
46. Webster G.S. (1996). A prehistory of Sardinia 2300–500 BC. SheYeld Academic Press, SheYeld.
47. Willerslev, E., Cooper, A. (2005). Ancient DNA. *Proc Biol Sci* 272, 3-16.

SUPPLEMENTARY MATERIALS



Figure S1: Pictures of selected samples from the Su Cannisoni (A: individual 8354) and Su Bittuleris sites (B-D: individuals 8363, 8417, 8418).

1. FCS
2. 8,355-48 clone1
3. 8,355-48 clone2
4. 8,355-48 clone3
5. 8,355-48 clone4
6. 8,355-48 clone5
7. 8,355-48 clone6
8. 8,355-48 clone7
9. 8,355-48 clone8
10. 8,355-48 clone9
11. 8,355-48 clone10
12. 8,355-48 clone11
13. 8,355-48 clone12
14. 8,355-48 clone13
15. 8,355-48 clone14
16. 8,355-48 clone15
17. 8,355-48 clone16
18. 8,355-48 clone17
19. 8,355-48 clone18
20. 8,355-48 clone19

Figure S2: Clone sequences for sample 8354B.

Table S2: Results of SNP typing in the mtDNA coding region using the GenoCoRe22 SNaPshot assay. SNPs typed on the L-strand are reported in capital letters in the reference rCRS profile, whereas SNPs typed on the H-strand are reported in small letters. Missing data signifies allelic dropout or fluorescence signal below the background threshold (100 relative fluorescent units, rfu). ‘g/a’ indicates the presence of a mixed signal for the position interrogated. A mixed signal was repeatedly obtained at position 12612 (haplogroup J) with the detection of an additional G base. However, the rest of the profile never could phylogenetically support the presence of the G base at this particular position. Despite of the allelic dropout consistently observed at position 10034 (haplogroup I), the remaining of the profile and their HVR-I sequence was always consistent.

	13928_R9	3594_L3'4	10550_K	11467_U	4248_A	8994_W	13263_C	13368_T	11719_preHV_R0	8280delB	4580_V	6371_X	10238_N1	7028_H	10034_I	5178_D	14766_HV	12705_R	10873_N	12612_J	2758_L2'6	10400_M	Hg ¹
rCRS	c	G	t	A	T	G	t	G	c	A	c	g	T	g	T	g	G	g	a	a	G	g	H
Sample																							
8354	C	G		A	T	G	T	G	C	A	C	G	T	G		G	G	G	A	A	G	G	H
8358	C	G		A	T	G	T	A	T	A	C	G	T	A		G	A	G	A	A	G	G	T
8418	C	G		A	T	G	T	A	T	A	C	G	T	A		G	A	G	A	A	G	G	T
8363	C	G		A	T	G	T	G	C	A	C	G	T	G		G	G	G	A	g/a	G	G	H
8415	C	G		A	T	G	T	G	C	A	C	G	T	G		G	G	G	A	g/a	G	G	H
8417	C	G		A	T	G	T	G	C	A	C	G	T	G		G	G	G	A	g/a	G	G	H
8368	C	G		A	T	G	T	G	C	A	T	G	T	A		G	G	G	A	g/a	G	G	V
8421	C	G		A	T	G	T	G	C	A	T	G	T	g/a		G	G	G	A	g/a	G	G	V
8394	C	G	C	G	T	G	T	G	T	A	C	G	T	A		G	A	G	A	g/a	G	G	K
8427	C	G	C	G	T	G	T	G	T	A	C	G	T	A		G	A	G	A	g/a	G	G	K
8399	C	G	T	A	T	G	T	A	T	A	C	G	T	A		G	A	G	A	g/a	G	G	T
8405	C	G		A	T	G	T	A	T	A	C	G	T	A	T	G	A	G	A	g/a	G	G	T
8406	C	G		A	T	G	T	A	T	A	C	G	T	A		G	A	G	A	g/a	G	G	T
8416	C	G		G	T	G	T	G	T	A	C	G	T	A		G	A	G	A	g/a	G	G	K
8420	C	G		A	T	G	T	G	T	A	C	g/a	T	a/g	T	G	A	A	G		G	G	N
8423	C	G		A	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	G	G	G	J

¹ Hg: haplogroup

General discussion

Conclusion

Abbreviations: ACAD, Australian Centre for Ancient DNA; aDNA, ancient DNA; F_{ST} , fixation index; HVR-I, hypervariable region I; Linienbandkeramik culture, LBK; mtDNA, mitochondrial DNA; PCA, Principal component Analysis; STR, Small Tandem Repeat; yBP, years Before Present.

GENERAL DISCUSSION

This study investigated temporal patterns of genetic diversity by typing mitochondrial DNA (mtDNA) from ancient human remains in order to reconstruct past human history (e.g., migration, local genetic continuity). Ancient DNA (aDNA) was sampled in time from populations from three geographical regions:

1. north east Europe (Chapter One and Chapter Two) with populations of the Mesolithic (~7,500 years Before Present, yBP), the Bronze Age (3,500 yBP) and the 18th century Anno Domini (A.D.; 200 yBP);
2. the Black Sea region (Chapter Three) as represented by horse-riding Scythian nomads of the Rostov-on-Don area (southwest Russia; 2,200 – 2,600 yBP);
3. central Sardinia (Chapter Four) in the Bronze Age (3,200 – 3,400 yBP).

This study faced the typical difficulties involved when working with ancient human DNA, such as variable preservation of DNA, DNA damage causing artificial mutations, risk of contamination by modern DNA, and sampling biases. Several controls and authentication criteria were followed in order to validate the ancient DNA sequences presented in this work. Ancient DNA data was compared to large mtDNA databases of modern populations of Eurasia. This allowed changes or persistence in the distribution of human mtDNA diversity to be detected through time. Despite evident methodological challenges, this work exemplifies the power of ancient DNA to extend the knowledge of human population history by confirming/refute hypothesis drawn from the current distribution of the human genetic diversity and/or archaeological and historical research. This work unravelled: cryptic large-scale population events such as pre-Bronze Age migrations from Siberia to northeast Europe (Chapter One and Chapter Two), the mixed origins of an archaeologically defined population, such as the Scythians of the Black Sea (Chapter Three), as well as local genetic continuity, such as that observed in central Sardinia since the Bronze Age (Chapter Four). On the basis of the work presented in this study, I will discuss the methodological and analytical aspects involved in working with ancient human DNA, as well as how the present study will contribute to the knowledge of human genetic history.

METHODOLOGY OF ANCIENT HUMAN DNA STUDY

Ancient DNA amplification success rates

DNA amplification success rates - i.e. the proportion of individuals yielding reliable ancient DNA sequences - are indicative of the preservation of DNA in ancient samples (the better preserved the samples, the more successful the sequencing of reliable genetic data). DNA amplification success rates were found to vary significantly among the ancient human populations investigated here (Table 1) due to various parameters. The age of the remains is an important factor that impacts on DNA preservation in ancient specimens. Other parameters include environmental: temperature, humidity, exposure to air, acidity, physicochemical properties of the soil, and the extent of variation of these conditions. The preservation of DNA in ancient human remains can also be influenced by human funerary traditions that include a range of burial practices: natural burial, chemical treatment of corpses (embalmmment, mummification), successive inhumation/exhumation, excarnation (removal of the flesh from corpses). Finally, the preservation of DNA in ancient remains may be impacted by the handling of the samples during and after excavation: cleaning of the samples, storage duration and conditions, and treatments for conservation purposes (e.g., in ethanol, formalin, wax, resin).

In this study, the lowest amplification success rate (24%) was obtained for the oldest sites, Uznyi Oleni Ostrov and Popovo, dated ~7,500 yBP (Chapter One), suggesting a poorer preservation of DNA. The post-excavation treatment of the tooth samples collected from the Uznyi Oleni Ostrov site in the 1950s involved cleaning of the teeth, conservation treatments (coating of the teeth with wax or resin) and intensive handling. It is possible that the significant time from excavation and the extensive treatment of the teeth affected the amplification success rate observed for the samples of Uznyi Oleni Ostrov. In contrast, the youngest site examined in this study, the Saami graveyard of Chalmny-Varre (18th century A.D. or 200 yBP; Chapter One), yielded the highest amplification success rate of 100% (Table 1). It is apparent from the observed amplification success rates, that the age of the samples was not the only factor influencing DNA preservation in the ancient human samples examined. While both sites of Bolshoy Oleni Ostrov (Kola Peninsula, north west Russia; Chapter One) and Su Bittuleris (central Sardinia, Italy; Chapter Four) were dated around 3,100 – 3,500 yBP, a discrepancy in their amplification success rates was observed: 100% for

Bolshoy Oleni Ostrov and 47% for Su Bittuleris. Discrepancies in the climatic conditions of the sites, in the time from excavation, or in the post-excavation treatment of the samples could be at the origin of the discrepancies observed in the amplification success rates between the Bolshoy Oleni Ostrov and the Su Bittuleris sites. The post-excavation conditions have been monitored to optimise DNA preservation of the central Sardinian remains (i.e., no brushing/washing of the samples, storage in cold conditions, short storage time), as a consequence, the poorer DNA preservation observed for the Sardinian samples might be due to environmental conditions of the burial, such as the Mediterranean climate (high and variable temperatures when compared to those of north east Europe). The amplification success rate for the Scythian samples (94%) is high considering the relative antiquity of the remains (2,200 – 2,800 yBP).

Table 1: Amplification success rates for the ancient human populations investigated.

	Ancient North East Europeans			Scythians	Ancient Sardinians
Sites	Uznyi Oleni Ostrov/ Popovo	Bolshoy Oleni Ostrov	Chalmny-Varre	various	Su Bittuleris/ Su Cannisoni
Location	Karelia/ Archangelsk region, North West Russia	Kola Peninsula, North West Russia	Kola Peninsula, North West Russia	Rostov-on-Don, South West Russia	Central Sardinia
Age	7,500 yBP	3,500 yBP	200 – 300 yBP	2,300 – 2,800 yBP	3,200 – 3,400 yBP
Amplification success rate	24% (11/46)	100% (23/23)	100% (42/42)	94% (16/17)	47% (16/34)

yBP, years Before Present

Authenticity of ancient mtDNA data

In all studies presented here, strict authenticity criteria and consistency of the genetic data allowed us to propose that contamination by exogenous DNA has had a minimal impact on the results presented in this study. Strict authenticity criteria were followed in this study. Below, we discuss the use of cloning and the contextual

interpretation of sequences in the light of population data for authenticating aDNA data.

Cloning was claimed to be an absolute necessity to authenticate ancient genetic data (Goloubinoff et al., 1993; Handt et al., 1994; Handt et al., 1996; Kolman & Tuross, 2000; Hofreiter et al., 2001; Cooper & Poinar 2000, Gilbert et al., 2003). Non-authentic mutations originating from contamination, aDNA damage or jumping-PCR, were believed to be identifiable by their rarer occurrence among the population of clone sequences and excluded from the consensus or 'correct' sequence. However, cloning was not used extensively in this study for the reasons described below. In extreme cases of low-template conditions, clone sequences can represent one or very few highly degraded starting DNA templates that took over during early cycles of the PCR. As a consequence, damaged sequences may represent the overwhelming majority of the clone sequenced and be misleadingly interpreted as authentic mutations. In addition, the lack of mutagenicity in the cloning process is questionable (Pavlov et al., 2004). Finally, sequences arising from random contamination or stochastic events occurring during the amplification reaction can be excluded by replicating direct-sequencing of PCR products independently and multiple times. Comparative experiments showed that repeated direct sequencing was as reliable, if not more reliable, than the analysis of 20 clones from a single PCR product (Pruvost et al., 2008). As an alternative to cloning, I chose a repetitive approach that involved the replication of the sequencing of each variable position at least twice for each sample. This eventually led to a four-fold coverage of the mutations for individuals from which two samples were analysed from independent samples. When cloning was performed, it confirmed the haplotypes determined by direct-sequencing of the PCR products in all cases. I consider that this repetitive approach, based on multiple independent repetitions is a powerful alternative to cloning in order to detect artificial mutations and provides confidence about the authenticity of aDNA sequences.

The consistency of the genetic data was also assessed with regard to the geographical, archaeological, anthropological, historical, evolutionary or genetic context of the five examined archaeological sites:

(1) Uznyi Oleni Ostrov, Mesolithic site in Karelia, north west Russia (~7,500 yBP; Chapter One and Chapter Two): the high frequencies of haplogroup U4 and U5a detected in the population of Uznyi Oleni Ostrov was consistent with regard to previously described ancient hunter-gatherer populations of prehistoric Europe

(Bramanti et al., 2009; Malmström et al., 2009; Krause et al., 2010). The lack of a positive match in the modern database for the newly identified C1f complete mitochondrial genome implies that haplogroup C1f carriers are very rare today, suggesting that recent contamination of the C1 Uznyi Oleni Ostrov samples with modern DNA is very unlikely.

(2) Bolshoy Oleni Ostrov, Bronze Age site in the Kola Peninsula, north west Russia (~3,500 yBP; Chapter One): the mtDNA structure emerging from the analysis of this site was not expected from the genetic makeup of present-day populations in this area. However, the proposed migration from Siberia to northeast Europe before the Bronze Age was supported by the observation of mtDNA lineages of Siberian origin in the present-day neighbouring populations of Karelia, and the Volga-Ural region.

(3) Chalmny-Varre, Saami graveyard in the Kola Peninsula, northwest Russia (18th century A.D., 200 yBP Chapter One): the consistency of the sequences obtained from this site was made evident by the retrieval of aDNA harbouring a very specific mutation pattern (the ‘Saami motif’) and a haplogroup distribution similar to that observed in modern-day Saami (high frequencies of haplogroups U5b and V).

(4) Scythian sites of the Rostov-on Don area, Black Sea region, south west Russia (2,200 – 2,600 yBP; Chapter Three): the genetic data obtained from the Scythian specimens was consistent with the population history, archaeology and previously described genetic diversity of culturally related Iron Age nomadic populations of the Eurasian Steppe. Similarly to these populations, a mixed genetic structure made up of ‘western’ and ‘eastern’ Eurasian lineages could be detected in the Scythians. This pattern was observed for several populations that have been genetically characterised by other studies carried out in various laboratories (Lalueza-Fox et al., 2004; Keyser et al., 2009; Ricaut et al, 2004a; Ricaut et al, 2004b; Chikisheva et al., 2007; Pilipenko et al., 2010). As a consequence, it seems very unlikely that this pattern emerged independently in different laboratories as a result of contamination.

(5) Su Cannisoni and Su Bittuleris, Bronze Age sites of central Sardinia, Italy (3,200 – 3,400 yBP; Chapter Four): the pattern of mtDNA variability was consistent with genetic continuity between Bronze Age and modern central Sardinia. This could have been the result of contamination of the samples by Sardinian archaeologists during the excavations. However, contamination could be rejected, as no overlap

could be observed between the sequences obtained from the excavation staff and the ancient remains.

Another argument for the low levels of intra-laboratory contamination and cross-contamination among ancient samples analysed in this work is the lack of significant overlap of haplotypes among archaeological sites. The populations studied here are geographically and temporally distinct, and accordingly, display distinct mtDNA structures.

Replication of ancient genetic data

Replication of aDNA analyses by an independent laboratory is one of the most decisive criteria to prove that ancient genetic data has not arisen from laboratory-specific contamination or artefacts. However, independent replication cannot be achieved for all samples and, should be undertaken carefully. Independent replication of ancient genetic data can be limited by the availability of samples. For example, in the study of the Bronze Age population of Su Bittuleris, central Sardinia, burial practices consisting in re-burying selected disarticulated body parts of the remains in a secondary burial. This did not permit two samples being extracted per individual, which is necessary for independent replication. Such burial practices are not uncommon in prehistoric communities. As a consequence, the exclusion of such sites from aDNA analyses on the ground that genetic analyses cannot be replicated independently would prevent access to important information about past human genetic diversity. For these reasons, alternative protocols should be established to increase the confidence in the quality of the aDNA data generated from these samples. Although not ideal, DNA extraction from two independent parts of the same sample may help ruling out laboratory-specific contamination during extraction, especially if experiments are carried out weeks apart using fresh reagents. Similarly, extensive replication of the sequencing may help ruling out contamination in DNA amplification reagents/materials. In such cases, it appears crucial to follow as many authenticity criteria as possible.

The principle of independent replication is based on the recognition that the independent laboratory performing the replication follows strict and appropriate procedures in order to avoid contaminations. This can however be difficult to assess and to question. Problematic situations can arise when different sequences are

obtained through independent replication, as it was the case in this study for one Scythian individual (RD-3; Chapter Three). The cause of the mismatch between independently generated sequences from a same individual can be difficult to determine and could be due to contamination of one of the samples, error in the anthropological identification/sample mix-up or laboratory contamination. It is sometimes possible to identify which conflicting sequence most likely represents a contaminant in the context of the analysed population (e.g., European lineages in Asian, African or American samples or vice versa). Unfortunately, it was not the case in the Scythian study presented here.

When ancient genetic data can be independently checked, failure to replicate even a single aDNA sequence can cast doubt on the authenticity of the whole dataset obtained from the ancient population. A possible solution would be to replicate the analyses from one additional sample in one of the laboratories, or ideally from two additional samples in both laboratories. However, access to additional archaeological samples can be limited, which is why independent replication should be performed with extreme care early on in the process. This would avoid requesting additional samples from museum collections and performing multiple additional analyses. In an ideal scenario, samples used for independent replication should be sent directly to the independent laboratory from the excavation site or the museum collection.

Contextual definition of the archaeological sites sampled for ancient DNA

In the same way in which information about the history, ethnicity, and cultural/linguistic/social background of modern populations is an essential requirement of human population genetics studies, solid archaeological/anthropological contextual information is pivotal for the interpretation of ancient population genetic data. Contextual information can be drawn from the study of morphology, archaeological artefacts, funerary practices or the architecture and organisation of funerary monuments and areas. Defining the context of the human remains is crucial for the formulation of research questions. Hypotheses about the genetic history of populations with regard to their genetic affinity to other ancient or modern populations can then be formulated on the basis of cultural or anthropological information.

Anthropological assessment of the sex and age of individuals, as well as archaeological identification of social structures within burial grounds is needed to

rule out possible representational bias that may be introduced in the analysis of the ancient genetic data. When comparing an ancient genetic dataset to other ancient or modern datasets, the fact that ancient individuals can be closely related, or can represent only a portion of the population under study (most of the time a social elite) can lead to the identification of artificial genetic affinities. For example, it is possible that the Scythian individuals (Chapter Three) only correspond to a minority of individuals of higher social status within Scythians. This, however, could not be assessed due to a lack of archaeological contextual information for the Scythian sites of the Rostov-on-Don area.

Analysis of the structure of burial grounds or monuments can lead to the identification of multiple groups within one funerary place. This is an important point, as individuals should not be considered part of a single population in aDNA analyses if they in fact represent anthropologically and genetically distinct entities. The same logic applies in cases where the funerary place has been in use for an extended period of time and its population is made up of temporally heterogeneous individuals. For example, the mtDNA dataset obtained for hunter-gatherers of central/eastern Europe is representative of a similar situation, where studied individuals were dated from 4,250 to 15,400 yBP (Bramanti et al., 2009). However, the specific genetic traits shared among these hunter-gatherers, which consisted in a high occurrence of haplogroup U during this long period of time, justify the pooling of these individuals into a single population for the analyses performed here.

Anthropological and archaeological contextual information is not always easily accessible even if essential to the analysis and interpretation of ancient genetic data. In this study, there were gaps in the contextual information available from articles written in English regarding the age, sex and organisation of the individuals within the burial sites. The lack of anthropological information for the samples obtained from central Sardinia (Chapter Four) was due to the specificities of the funerary practices of the Bronze Age central Sardinian population.

ANALYSIS OF ANCIENT GENETIC DATA

The temporal heterogeneity that characterises ancient genetic datasets adds a level of complexity to the analysis of genetic data. The main problems introduced by aDNA are small population sizes, sequence errors as a result of aDNA damage (e.g.,

Axelsson et al., 2008; Rambaut et al., 2009), and the time discrepancy between compared populations, as described below. Their impact on the interpretation of genetic analyses is variable according to the type of data analysed: haplogroup frequencies (by Principal Component Analysis, Multidimensional Scaling, genetic distance maps, fixation index, F_{ST} , calculation) or DNA sequences (by phylogenetic trees and networks, haplotype sharing analysis, F_{ST} calculation, coalescent simulations).

Analyses based on haplogroup frequencies

Analyses of haplogroup frequencies offered here an excellent means to visualise mtDNA data and gauge population relationships. They served as a basis to formulate hypotheses about genetic affinities between populations prior to haplotype-based analyses. Analyses of haplogroup frequencies are only moderately affected by artificial mutations, damage-induced mutations only impact haplogroup frequencies in cases where they alter haplogroup-defining nucleotide positions. In this work in particular, this bias was controlled as haplogroup assessment was constrained by genetic information obtained from both the hypervariable region I (HVR-I) and the coding-region.

However, analyses of haplogroup frequencies can be very sensitive to small sample sizes that increase the randomness of the identified genetic affinities. Small sample sizes can lead to the failure at detecting relevant haplogroups with regard to the genetic history of the population under study.

In the case of the maternally-inherited mtDNA, another possible bias may be un-identified maternal relationships between individuals of the studied populations. Kinship can lead to the overestimation of frequencies for shared mtDNA haplotypes, which are not representative of those of the total population. For example, in the population of Uznyi Oleni Ostrov/Popovo (Chapter One), three individuals were found to belong to haplogroup C1 that is represented by a single C1f haplotype. It cannot be rejected that the C1 carriers were maternally related. In this case, the frequency of haplogroup C1 should not be 3/11 (27%) but 1/9 (11%). Molecular identification of kinship is possible through the use of commercial forensic identification kits based on the analysis of nuclear Small Tandem Repeats (STR; e.g., Applied Biosystems AmpFISTR® STR genotyping kit range, used for example in Keyser-Tracqui et al., 2003). Attempts were made to amplify nuclear DNA from extracts obtained from all

the populations studied here using the molecular techniques currently available at the ACAD, but remained unsuccessful. In addition, little anthropological/archaeological information was available about kinship within the populations studied. As a consequence, the archaeological identification of possible maternal relationships within the examined populations was not possible. The impact of maternal relationships among individuals sampled for aDNA were discussed in the case of the populations of Uznyi Oleni Ostrov (Chapter One) and Su Bittuleris/Su Cannisoni (Chapter Four). They were thought to be negligible for the populations of Bolshoy Oleni Ostrov (Chapter One) where relatively high diversity was observed, and Scythians (Chapter Three) from which geographically dispersed samples were obtained.

Contrary to biases relative to the survival of DNA among individuals of the populations under study, biases introduced by selective burials are not random. Selective burial and their associated biases are a particularity of ancient human DNA studies. In past societies, individuals were selectively buried on the basis of various criteria, one of which being the social status that can be inheritable. Selective burial is evident in the funerary monuments used by nomadic populations of the Central Asian Steppes. These ‘kurgans’ are characterised by rich funerary artefacts and material evidence for a superior social status was associated to the remains. In the study of Scythians (Chapter Three), however, this effect was possibly limited by the sampling of multiple different sites in the Rostov-on-Don area, from which only one or two individuals were sampled. This collection strategy reduces the risk of sampling members of the same family. The significant diversity observed in the data does not support selective maternal sampling.

The reduction of sequence data to haplogroup frequency data can sometimes lead to misleading results, i.e. artificial population affinities. The genetic affinity observed between Uznyi Oleni Ostrov and modern Western Siberians on the basis of the shared pattern of high frequencies in haplogroups U4, U5a and C is a clear example of the reduced power of resolution and should be considered with caution (Chapter One). Detailed analysis on the haplotype level showed that the high frequency in haplogroup C was possibly overestimated and phylogenetic analysis of the C1 haplotype could not prove any link with western Siberian populations. As a result, it seems important to further analyse potential affinities based on haplogroup frequencies via haplotype-based analyses.

Analyses based on haplotypic data

Haplotype-based analyses can prove very powerful at identifying genetic affinities between populations. The identification of a particular haplotype in an ancient population that exhibits a very restricted distribution in a specific modern population makes the genetic link identified between the two populations difficult to question. This was illustrated by the retrieval of the haplogroup U5b1b ‘Saami motif’ in the Chalmny-Varre graveyard that genetically confirmed the affiliation of the graveyard within the Saami culture. However, analyses involving haplotypic data are more sensitive to errors due to DNA damage (e.g., deamination of cytosines; Stiller et al., 2006; Gilbert et al., 2007) or high DNA fragmentation leading to jumping-PCRs (Gilbert et al., 2003). An artificial mutation altering an informative nucleotide position could be misleadingly interpreted. However, this possible source of errors has been eliminated here by following a range of stringent authenticity criteria.

Time discrepancy between compared populations complicates the analysis of sequence data because differences observed among populations could be the result of population events, such as migrations, but could also be due to the sole effect of genetic drift. Time discrepancy between populations can only be taken into consideration in coalescent simulation-based analyses. The likelihood of tested scenarios given the genetic data could be statistically compared for the ancient populations of northeast Europe (Chapter One) under a variety of demographic models using the Bayesian Serial SimCoal program (Anderson et al., 2005) and an Approximate Bayesian Computation approach (Beaumont et al., 2002). These methods are ‘model-based’, i.e., they only compare a selection of predefined demographic models, and as such do not reconstruct the ‘real’ demographic scenario from the data. However, they allow the distribution of population parameters to be estimated and demographic scenarios to be rejected and/or compared.

Even if haplotypic data proved in this work to provide more resolution for the identification of population affinities than haplogroup frequency-based data, HVR-I sequences appeared to lack resolution in some instances. As expected when examining past populations, a significant number of basal haplotypes were detected in the ancient populations investigated: 33% in Uznyi Oleni Ostrov/Popovo (Chapter One), 40% in Bolshoy Oleni Ostrov (Chapter One), 0% in Chalmny-Varre (Chapter One), 21% in Scythians (Chapter Three) and 50% in central Sardinians (Chapter Four). These haplotypes display wide geographical, and also temporal, distributions that make

particular genetic affinities difficult to identify. An increase of resolving power through analysis of complete mitochondrial genomes could provide additional information about the geographical and/or temporal relationships among populations. Here, the sequencing of the haplogroup C1f mitochondrial genome of an individual from the Uznyi Oleni Ostrov Mesolithic graveyard allowed the resolution of the phylogenetic position of this haplotype within the haplogroup C1 phylogenetic tree. Hypotheses about the origins of the C1 lineages in Europe could consequently be ruled out and more precise models for the population history of Uznyi Oleni Ostrov were proposed. Currently, the number of published ancient complete mitochondrial genomes of anatomically modern humans (*Homo sapiens*) is still small compared to that of HVR-I sequences. These genomes were obtained from isolated individuals: a 3,400 - 4,500 year-old Palaeo-Eskimo from Greenland (Gilbert et al., 2008), a 5,100 - 5,350 year old mummified corpse from the Alps (Ermini et al., 2008) and 30,000 year old remains from Kostenki 14, west Russia (Krause et al., 2010). The reconstruction of human history is currently still limited by the lack of published human mitochondrial genome sequences from well-defined modern populations (8,731 complete mtDNA genome entries in the PhyloTree database; van Oven & Kayser, 2009; versus more than 200,000 HVR sequences in GenBank). Eventually, the study of early human migrations and population origins will surely benefit from advances allowing the study of modern and ancient individuals with the resolution of whole mitochondrial genomes at the scale of populations. An approach that would integrate temporal sampling in a population genetics framework would promise to be very powerful at revealing unexpected aspects of the ancestry of human populations, such as genetic continuities or migrations. The study of ancient whole mitochondrial genomes is also promising for the investigation of the human mtDNA mutation rate. Temporally sampled and well-dated complete mitochondrial genomes from ancient specimens would help circumvent the limitations of the calibration methods currently in use for the dating of more recent, i.e., Holocene phylogenetic events, i.e., divergence, within the human mitochondrial tree (Ho & Gilbert, 2010).

CONTRIBUTION OF ANCIENT DNA TO THE RECONSTRUCTION OF HUMAN POPULATION HISTORY

As the authentication of ancient human DNA is a challenging task, the number of genetic studies of ancient human populations has been low compared to studies on animals, plants and other organisms. In recent years, however, the number of studies describing the genetic diversity and in particular mtDNA variation in ancient human populations has significantly increased. This enabled direct comparison of ancient mtDNA data obtained in this study from northeast Europeans (Chapter One and Chapter Two), Iron Age Scythians (Chapter Three) and Bronze Age Sardinians (Chapter Four) to genetic information available from other ancient and modern-day human Eurasian populations (Figure 1 and Table 2). Principal Component Analysis (PCA) of mtDNA haplogroup frequencies was performed in order to identify genetic affinities of human populations in time and space (Figure 2; see Chapter Three for Material and Methods, as well as the description/references of the populations used for comparison). For comparison, I selected ancient populations located over a wide geographical range, including Sardinia, Spain, central Europe, eastern Europe, Scandinavia, western/southern/central/eastern Siberia, Mongolia, the Tarim basin in China, and Kazakhstan (Figure 1 and Table 2). These populations varied in age ranging from the Palaeolithic/Mesolithic (4,250 – 30,000 yBP) to the Neolithic (5,000 – 7,500 yBP), Bronze Age (2,700 – 3,980 yBP), the Iron Age (1,500 – 2,800 yBP) and the 18th century (200 yBP). The ancient mtDNA data provided important temporal aspects for the reconstruction the genetic history of Eurasian populations such as:

- 1) the genetic diversity of European Palaeolithic/Mesolithic populations and the absence of genetic continuity with Neolithic and present-day populations of Europe,
- 2) the influence of eastern Eurasia lineages on the mtDNA gene pool of eastern Europe,
- 3) the genetics of European genetic outliers (Saami and Sardinians).

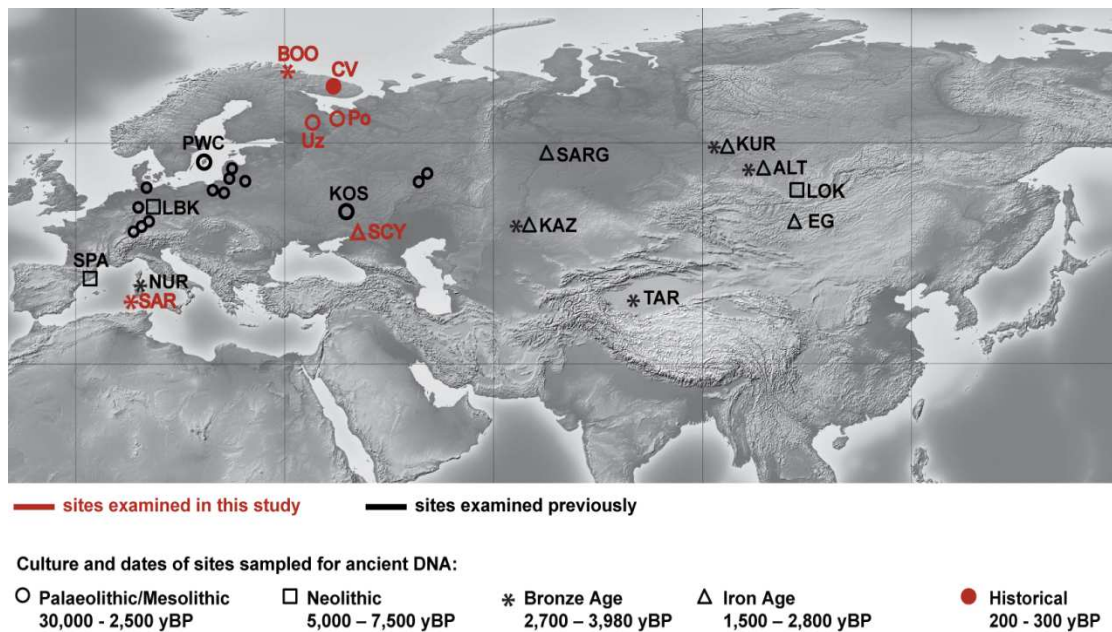


Figure 1: Map showing the location of sites sampled for ancient DNA in this study (red) and in other studies (black). See Table 2 for population description and references

Table 2. Description of ancient populations.

Description	Abbreviation	Reference	N	Date range (yBP ^a)
PALAEOLITHIC / MESOLITHIC				
Uznyi Oleni Ostrov / Popovo	UzPo	This study	11	7,000 - 7,500
Central and east European hunter-gatherers	HG	Bramanti et al., 2009; Krause et al., 2010	22	4,250 - 30,000
Scandinavian Pitted-Ware Culture individuals	PWC	Malmström et al., 2009	19	4,500 - 5,300
NEOLITHIC				
Linearbandkeramik, Derenburg, Germany	LBK	Haak et al., 2010	47	7,000 - 7,500
Camí de Can Grau, Spain	SPA	Sampietro et al., 2007	11	5,000 - 5,500
Kitoi Neolithic of Lake Baikal	LOK	Mooder et al., 2005	30	6,130 - 7,140
BRONZE AGE				
Bronze Age individuals of Kazakhstan	KAZ-BA	Lalueza-Fox et al., 2004	13	2,700 - 3,400
Bronze Age Kurgan individuals of South Siberia	KUR-BA	Keyser et al., 2009	11	2,800 - 3,800
Bronze Age individuals of the Altai	ALT-BA	Chikisheva et al., 2007	3	3,500 - 4,000
Bronze-Age individuals of the Tarim Basin	TAR	Li et al., 2010	20	3,980
Nuragic Sardinians	NUR	Caramelli et al., 2007	23	2,700 - 3,430
Central Sardinians	SAR	This study	16	3,200 - 3,400
IRON AGE				
Iron Age individuals of Kazakhstan	KAZ-IA	Lalueza-Fox et al., 2004	12	2,100 - 2,800
Iron Age Kurgan individuals of South Siberia	KUR-IA	Keyser et al., 2009	15	1,600 - 2,800
Pazyryk culture individuals of the Altai	ALT-IA	Ricaut et al, 2004a; Ricaut et al, 2004b; Chikisheva et al., 2007; Pilipenko et al., 2010	11	2,300 - 2,500
Sargat individuals of South West Siberia	SAR	Bennett & Kaestle., 2010	5	1,500 - 2,500
Xiongnu individuals of the EgyinGol Valley, Mongolia	EG	Keyser-Tracqui et al., 2003	46	2,200 - 2,300
Scythians of the Black Sea area	SCY	This study	15	2,200 - 2,600
OTHER				
Chalmny-Varre Saami	CV	This study	42	200 - 300

^ayears Before Present

Genetic diversity of European Palaeolithic/Mesolithic populations and absence of genetic continuity with Neolithic and present-day populations of Europe

In Europe, Upper Palaeolithic (~10,000 - 45,000 yBP) and Mesolithic (~6,000 – 12,000 yBP) human populations consisted of small groups of mobile foragers who mainly relied on hunting, fishing, and plant gathering as food sources. Ancient DNA was previously recovered from Palaeolithic/Mesolithic populations of central/eastern Europe and Scandinavia (4,250 - 30,000 yBP; HG and PWC, respectively, on the PCA biplot Figure 2, references in Table 2). These populations were characterised by high frequencies and diversity in haplogroup U. On the PCA biplot, the elevated frequency of haplogroup U in the populations of the central/eastern Europe (HG; 72.7%) and Scandinavia (PWC; 53.7%) causes the two populations to group together at the periphery of present-day North East Europeans. Ancient mtDNA data from the Mesolithic sites of Uznyi Oleni Ostrov and Popovo (north west Russia, ~7,500 yBP, Chapter One) expanded knowledge of mtDNA diversity in European Mesolithic populations to the North East. This study is the third independent study clearly showing that high frequencies of haplogroup U (41.6% in Uznyi Oleni Ostrov/Popovo) were a common feature of eastern European hunter-gatherers. The haplogroup U component of Upper Palaeolithic/Mesolithic European populations was composed of the sub-clades U4 (in PWC, HG, UzPo), U5a (in PWC, HG, UzPo), U5b (in HG) and U2 (in UzPo and the ~30,000 yBP individual of Kostenki, in the Don River Valley in western Russia; KOS; Krause et al., 2010).

Mitochondrial data obtained from Palaeolithic/Mesolithic populations of Europe allows the direct verification of hypotheses formulated on the basis of modern mtDNA data. Previous analyses calculated coalescent age estimates for haplogroups found in Europe (Richards et al, 1996; Richards et al., 1998; Richards et al, 2000; Soares et al., 2009; Soares et al., 2010; Malyarchuk et al., 2010). Coalescent ages differ among studies, but haplogroup U consistently appear as the oldest haplogroup in Europe, with coalescent age estimates between 53,600 and 57,200 yBP being estimated by Soares et al., 2009. These dates suggest that haplogroup U was found in Europe before the advent of the Neolithic (~6,000 - 9,000 yBP), which was clearly demonstrated in the three studies of Palaeolithic/Mesolithic populations of Europe (Bramanti et al., 2009; Malmström et al., 2009; Chapter One). The presence of haplogroups U4 and U5, and to a lesser extent U2, appeared as real genetic signatures of central/eastern European hunter-gatherer, which is in accordance with these clades

being dated prior to the Neolithic on the basis of modern-day mtDNA data (coalescent ages being estimated around 25,300 – 47,200 for haplogroup U5, 11,000 – 31,200 yBP for haplogroup U4 and 40,300 – 67,200 for haplogroup U2; Soares et al., 2009). Similar genetic makeups dominated by haplogroup U were previously observed in Bronze Age nomadic populations of central Eurasia: Kurgans (KUR-BA; Andronovo and Karasuk cultures) of central Siberia, and Altaians (ALT-BA) of southern Siberia. Accordingly, these populations (KUR-BA and ALT-BA) occupy a position close to Palaeolithic/Mesolithic European foraging populations (HG and PWC) on the PCA biplot (Figure 2). This suggests either that the genetic substratum of Mesolithic Europe extended further East than today in the Bronze Age, or a recent migration from Europe into nomadic populations of central/southern Siberia (KUR-BA and ALT-BA).

Despite an elevated frequency of haplogroup U (41.6 %), the Mesolithic population of Uznyi Oleni Ostrov/Popovo (UzPo) does not cluster with the other two previously described Palaeolithic/Mesolithic populations of Europe (HG and PWC) on the PCA biplot (Figure 2). This is due to the presence of haplogroup C1 at significant frequencies in the population of Uznyi Oleni Ostrov. The sequencing of the complete mitochondrial genome of a C1 carrier from the Uznyi Oleni Ostrov allowed the identification of the new sub-haplogroup C1f. The absence of matching or closely related haplotype in modern human populations left the question of the origin of C1f partially unresolved. However, it can be proposed that this lineage represents a genetic influence from the East, as most of the haplogroup C diversity is found today in eastern Eurasia where it is thought to have originated. The observation of the C1f lineage only in Mesolithic Uznyi Oleni Ostrov suggested low frequencies of C1f in the Mesolithic and/or a relative mating isolation of Mesolithic foragers as well as important impact of demographic processes after the Mesolithic such as population extinction and/or replacement. The detection of the C1f lineage in Mesolithic Uznyi Oleni Ostrov is a good example of the power of aDNA to detect past genetic diversity that had remained unnoticed from the study of present-day human mtDNA diversity because of extinction or under-sampling of whole mitochondrial genomes in modern-day populations.

When compared to mtDNA data from modern-day Europeans, the high frequencies of haplogroup U observed in Palaeolithic/Mesolithic populations of Europe associated with the low frequencies, if not absence, of other main haplogroups found in Europeans today (haplogroups H, HV, I, J, K, T, V, W, X) suggest an

important genetic discontinuity between Palaeolithic/Mesolithic populations and present-day Europeans. Numerous demographic events can be proposed to be at the origin of this genetic discontinuity. Among them, the Neolithic transition has been the most significantly studied (e.g., see review in Richards, 2003).

When comparing mtDNA data obtained from Neolithic populations (5,000 – 7,500 yBP) and that from Palaeolithic/Mesolithic populations, the PCA biplot (Figure 2) suggests a genetic discontinuity between Mesolithic and Neolithic populations. Unlike in the Palaeolithic/Mesolithic, Neolithic populations were characterised by low frequencies of haplogroup U (0% in the Linienbandkeramik culture, LBK, population of Derenburg, Germany; 9.1% in the SPA population of Camí de Can Grau', Spain; see Table 2 for all references to ancient population studies). The mtDNA data obtained for the two Neolithic populations of Germany (Derenburg, LBK) and Spain (SPA) showed the heterogeneous and complex nature of the Neolithic transition in Europe (Sampietro et al., 2007) on the PCA biplot (Figure 2). The mtDNA data for the LBK population was clearly found to be a discontinuity with both Mesolithic and modern European populations, in particular as a result of the high frequency and diversity in haplogroup N1a (Haak et al., 2005; Bramanti et al. 2009, Haak et al., 2010). Haplogroup N1a was not detected in the neighbouring populations of hunter-gatherers sampled for aDNA (Bramanti et al., 2009) and is rare today with a distribution centred in the Near East. This pattern of geographical and temporal variability provided evidence for migration(s) associated with the spread of the Neolithic from the Near East into central Europe (Boyle & Renfrew, 2000; Whittle, 1996). Despite the fact that a clear genetic discontinuity between Mesolithic, Neolithic and present-day has been identified in central/eastern/north eastern/Europe, the introduction of new lineages did not erase the Mesolithic genetic substratum as U lineages still persist in the European gene pool and constitute a significant part of it, especially U5 and U4 in north/north eastern Europe. It is also possible that the LBK population represents a genetic isolate compared to other Neolithic populations of central Europe. Further sampling of Neolithic populations in this area could help assess the level to which the LBK population of Derenburg is representative of the mtDNA diversity in Neolithic populations of central Europe. In contrast with the dominance of haplogroup N1 observed in the Neolithic population of central Europe, diverse haplogroups were detected in the Spanish Neolithic population: H, I, J, T, U4, W (Sampietro et al., 2007), which was found to be comparable to present-day populations of Europe. Data

describing the pre-Neolithic and early Neolithic mtDNA diversity in southern/western Europe is currently missing to determine whether the pattern of mtDNA variability observed in both the Spanish Neolithic and the present-day was already present in the Mesolithic and Early Neolithic gene pool of southern/western Europe (i.e., genetic continuity between the Mesolithic, the Early and Late Neolithic and the present day in western/southern Europe) or is the result of a migration associated with the Neolithic.

The complexity and geographical heterogeneity of the Neolithic transition in Europe has been proposed on the basis of archaeological (e.g., Price et al., 2000; Bocquet-Apel et al. 2009) and anthropological data (e.g., Pinhasi et al., 2009). Previous studies suggested a rapid diffusion of the Neolithic around the Mediterranean Basin and a slower diffusion along the Danube River into central Europe (Whittle & Cummings, 2007; Price, 2000). In regions found at the periphery of the main route of agricultural spread, such as northeast Europe, the spread of agriculture was hindered by ecological and climatic conditions not optimal for agriculture (Zvelebil, 1986; Zvelebil, 1996). There, Mesolithic (and associated lineages) persisted for a longer period of time and the Neolithic transition is thought to have been more gradual and to have involved more genetic continuity. For these reasons, the Neolithic transition in Europe has been defined as a ‘mosaic’ process (von Cramon-Taubadel et al., 2011), which is supported by aDNA data.

The reconstruction of the genetic history of Europeans has centred on the relative contribution of the Palaeolithic versus the Neolithic genetic components in modern Europeans (e.g., reviewed in Richards, 2003). However, ancient DNA studies suggest that the full reconstruction of the genetic history of Europeans is not limited to the resolution of this question. Late Neolithic (e.g., associated with the Bell Beaker culture) and post-Neolithic events (e.g., associated with the Bronze Age and the Iron Ages, respectively; ~5,000 – 3,000 yBP and ~3,000 – 1,400 yBP), or the Roman Empire, 27 Before Christ – 476 Anno Domini), as well as population processes (population expansions, reduction in population sizes, migrations and population extinction), appear of great relevance to the investigation of the genetic history of Europeans.

Mitochondrial influence of eastern Eurasia in eastern Europe

The study of the populations of Bronze Age Bolshoy Oleni Ostrov (Chapter One) and Iron Age Scythians (Chapter Three), as well as previously described ancient

nomadic populations of central Eurasia, shows that the genetic influences between western and eastern Eurasia have been dynamic in time and space. In Eastern Europe, we detected genetic influences from Siberia and central Asia both in the north (Kola Peninsula, Chapter one and the South, Chapter Three). In the course of human population history, eastern Europe seems to have been the recipient of multiple genetic influxes from Siberia and Central Asia, as the gene pool of the Bolshoy Oleni Ostrov population and of the Scythians are genetically different (Figure 2). The advantage provided by aDNA here was that dates and archaeological cultures could be associated with these influxes. These migrations appear to have moderately impacted the gene pool of modern East Europeans probably because of the dilution or mating isolation of new coming lineages.

Origins of European genetic outliers: Saami and Sardinians

The ancient mtDNA data presented here also provided elements for the reconstruction of the genetic history of European genetic outliers: the Saami of northern Europe (Chapter One) and Sardinians (Chapter Four).

The Saami differ from the rest of European populations due to their lifestyle (semi nomadic with an economy based on reindeer herding and fishing), as well as their anthropological and genetic characteristics. The origins and the timing of arrival of the populations ancestral to present-day Saami in northern Europe remain mysterious. Possible ancestors of the Saami could not be identified in the ancient populations examined here. The fact that these ancestors seem difficult to identify, from both the ancient and modern record, suggests that their genetic impact has been limited on the present-day gene pool of Europeans. Founder events and successive population bottlenecks have been proposed before to explain the significant genetic differentiation of present-day Saami from Europeans (Tambets et al., 2004).

In contrast to the Saami, genetic continuity between prehistoric and present-day populations could be identified for the other European genetic outliers examined here, the Sardinians. Genetic continuity in central Sardinia indicates that the population has been relatively isolated from any external genetic input that could have led to homogenisation with the rest of the European population since the Bronze Age. Genetic differentiation of Sardinians is also probably caused by a strong endogamy, i.e., marriage within the group of origin. The population of Bronze Age central Sardinians (SAR) appears genetically close to the Neolithic population from Spain

(SPA) on the PCA biplot (Figure 2). The mtDNA gene pools of both ancient populations (SAR and SPA) share common genetic features characteristic of the Mediterranean area, in particular relatively high frequencies of mtDNA haplogroups H, J and T. Of note, the previously described population of Bronze/Iron Age Nuragic Sardinians (NUR) probably has its marginal position compared to other ancient and present-day Mediterranean populations on the biplot (Figure 2, references for populations in Table 1). This could be explained by the significantly higher frequencies of haplogroup H that the Nuragic population displays and that might be artificial (see Discussion Chapter Four). It can be proposed that Sardinian and Spanish Neolithic individuals arose from a similar genetic background. Then, as ancestors of present-day Sardinians reached the centre of the island they became isolated. It is possible that the genetic isolation of central Sardinia dates back to the initial settlement of this region.

The power of ancient DNA

In this study, aDNA allowed hypotheses drawn from the study of present-day human mtDNA diversity to be directly verified (e.g., the antiquity of haplogroup U in Europe, Chapter One). This study also exemplifies the power of aDNA to uncover lost genetic diversity (e.g., the C1f lineage in Chapter Two) and past cryptic migrations (e.g., in north east Europe Bolshoy Oleni Ostrov, Chapter One) that could not be detected from the analysis of the present-day distribution of human genetic diversity. In these instances, aDNA reveals the dynamic and plastic nature of human population processes. Ancient DNA also enabled the detection of human population processes in the past that differed in terms of their geographical scales, as well as the influence exerted by geography and/or lifestyle. In central Sardinia, aDNA demonstrated a local maternal genetic continuity since the Bronze Age that probably is the result of a strong endogamy and genetic isolation (Chapter Four). Conversely, the genetic analysis of Iron Age Scythians of the Black Sea area revealed genetic influences from various geographical sources over the whole of central Eurasia. Mixed maternal origins are expected from populations like Scythians, which were characterised by a nomadic lifestyle. Geography also played a role in the long-distance spread of mitochondrial lineages in the past, which was favoured in open environments such as the central Eurasian Steppe.

The genetic comparison of populations in a defined geographical area (north east Europe in Chapter One and Sardinia in Chapter Four), sampled through time, allows temporal changes in the mtDNA composition of past populations to be followed locally and dynamic population processes to be detected (continuity, migrations). In order to detect the source of these migrations, a geographically wide sampling of the genetic diversity through time (Iron Age Eurasia; Chapter Three) is also important. In addition, broad geographical sampling provides temporally defined snapshots of the distribution of the mtDNA variability that could reveal helpful to detect past genetic affinities between archaeologically similar populations (Iron Age Eurasia; Chapter Three).

With the future accumulation of genetic information from populations over larger temporal and geographical scales, human genetic diversity will be mapped with a better resolution on both scales. This will enable the detection of long-term local genetic continuity, or human migrations and their origins. This study shows that, in a well defined archaeological context, aDNA, when compared with modern genetic data, has the potential to reconstruct, pieces by pieces, the genetic history of humans.

CONCLUSION

The study presented here describes the mtDNA structure of three groups of ancient Eurasian populations broadly sampled in time (200 - 7,500 yBP) and space (north east Europe, the Black Sea area and central Sardinia). They represent a range of prehistoric (Mesolithic, Bronze Age, Iron Age) and historical (Saami) cultures. The ancient human genetic data from 107 individuals was compared to a large comparative database of ancient and modern populations in order to investigate mitochondrial continuity and/or discontinuity through time. Ancient DNA provided valuable genetic evidence for the timed reconstruction of human population history at the scale of Eurasia.

Significance and contribution to knowledge

The results presented in this study are significant to the scientific fields investigating human past and present genetic diversity, i.e., aDNA and human genetics, but also to disciplines studying human evolution, past migrations and cultures, such as anthropology, archaeology, and history.

This work contributes significantly, in terms of the number of ancient mtDNA sequences generated (107), to the geographical and temporal mapping of human mtDNA diversity, which is still scarce as far as prehistoric Europe is concerned. As a consequence, this study provides valuable insights into prehistoric European populations. Recent advances in the field of ancient DNA has made the study of ancient human population a promising subject of research (Rasmussen et al., 2010). As a consequence, the data presented in this work will certainly constitute reliable datasets that can be used for comparison in future studies as more aDNA data will accumulate.

This work also presents novel molecular biology techniques, currently under development at the ACAD. In Chapter Two, a DNA library specifically enriched for ancient human mtDNA was constructed using these techniques and allowing a complete ancient mitochondrial genome to be sequenced. The sequencing of this complete mtDNA genome revealed a portion of the human mtDNA diversity that had not been detected before from the sampling of modern-day populations. This work is

significant as only three other ancient complete mitochondrial genome sequences have been published to date (Gilbert et al., 2008; Ermini et al., 2008; Krause et al., 2010).

Finally, the application of a range of statistical analyses to compare newly acquired ancient genetic data to previously published ancient and modern data is not common in aDNA studies. In this respect, the integrative analytical approach used here contributes the originality of this work. It allowed the direct reconstruction of human migrations, and genetic continuities.

Problems encountered

The main problems encountered in this study were methodological difficulties associated with the limitations of aDNA work. One problem was the poor preservation of samples from the Uznyi Oleni Ostrov site (Chapter One) that led to low amplification success rates and small population sizes sub-optimal for some statistical analyses of the genetic data. Another problem that also arose from the relatively poor preservation of the samples was that nuclear DNA could not be retrieved using with the molecular techniques currently in use at ACAD, thus rendering the analyses of Y-chromosome and autosomal markers impossible. As a consequence, additional genetic aspects such as patterns of paternal diversity or genetic adaptation could not be investigated. Lastly, the replication of one Scythian sample in an independent laboratory (Chapter Three) proved problematic but could eventually be solved by repeating the analysis.

Considering the difficulties and costs associated with the genetic and contextual characterisation of ancient samples, sample sets should be chosen with care prior to aDNA analyses. A balance should be found between the a priori anthropological/archaeological/genetic significance of the samples (i.e., arbitrarily assessed on the basis of their age, place of origin, archaeological/anthropological context), and the risk of facing low amplification success rates, or poor contextual characterisation.

Future direction

On the basis of the results presented in this study, I think that future work in the field of ancient human DNA should focus on three main directions: 1) broadening the sampling of ancient human remains for DNA analysis in space and time, 2)

sequencing ancient human mitochondrial genomes, 3) investigation of the diversity of nuclear markers, such as Y-chromosome and autosomal markers.

1. Future work should focus on broadening the sampling of ancient genetic data at the temporal and geographical scales. The current sampling of ancient human mtDNA is still sparse in space and time. Future aDNA work should aim to apply a population approach, in which as many ancient individuals as possible should be sampled to allow statistically significant analyses of mtDNA data. The sampling of ancient populations providing additional reliable mtDNA data could prove essential to the investigation of a range of questions regarding the genetic history of humans. On the basis of the current work, I suggest that future sampling of ancient populations in Eurasia should focus on the following geographical areas and time periods:

Mitochondrial data from Palaeolithic and Mesolithic populations of western and southern Europe is needed to complement the characterisation of the mtDNA diversity during the Palaeolithic/Mesolithic, which is currently restricted to central, eastern and north eastern Europe. This data could provide direct evidence to verify or falsify hypotheses drawn from the study of mtDNA in modern-day populations, in particular with regard to the role of the Holocene recolonisation from southern glacial refugia (~10,000 yBP; Torroni et al., 2001; Achilli et al., 2004; Pereira et al., 2005) and its contribution to the modern-day European gene pool. Palaeolithic and Mesolithic populations of western and southern Europe could also yield interesting mtDNA data for the investigation of the genetic continuity between hunter-gatherers and Neolithic early farmers. Finally, additional mtDNA data from Neolithic populations in Europe would also be necessary to better characterise the genetic heterogeneity that has already been observed from the Neolithic mtDNA currently available (Haak et al., Sampietro et al., 2007; Haak et al., 2010; Lacan et al., 2011).

Post-Mesolithic European populations should also be sampled for aDNA in order to identify the migrations involved in the spread of the mtDNA that are common in present-day populations around Europe but were found at low frequencies in Mesolithic populations of central, eastern, northern and north eastern Europe (haplogroups H, HV, I, J, K, T, W).

In order to solve the mysteries of the origins of the Saami, ancient populations of northern Scandinavia (Norway, Sweden, and Finland) should be sampled for ancient mtDNA. This could potentially allow the ancestral population of the Saami to

be identified and the origins of the ‘Saami motif’ (haplogroup U5b1b1) to be localised.

The availability of mtDNA from populations from the Black Sea area and dated to the Bronze Age would allow the hypothesis of the local origins of the Scythians to be directly tested.

The previously proposed genetic homogeneity in Nuragic Sardinia (Bronze age/Iron Age, 1,800 – 3,800 yBP; Caramelli et al., 2007) could be tested with the availability of additional mtDNA sampled from populations all over Sardinia for this time period.

2. Future research should focus on broadening the knowledge of past human mtDNA diversity by sequencing complete mitochondrial genomes from ancient human remains. This will be achieved provided that aDNA enrichment techniques currently under development (Stiller et al., 2009; Briggs et al., 2009; Burbano et al., 2010; Maricic, et al., 2010) are applied. Ancient mitochondrial genome sequences provide more resolution to human mitochondrial phylogeography and to the identification of population affinities. This is particularly true in cases where HVR-I sequences are uninformative basal haplotypes, which can be widespread temporally and spatially. Sequencing ancient human mitochondrial genomes may also reveal previously unknown (extinct or under-sampled) mtDNA lineages or clades within the human mtDNA tree that would contribute to a better understanding of human evolutionary history. Finally, the availability of dated sequences of mitochondrial genomes could be of crucial importance for the investigation of the human mtDNA mutation rate. Currently, dating of divergence events based on available estimates of the human mtDNA mutation rate is highly debated (e.g., see review by Endicott et al., 2009). The use of inter-species calibration points, e.g., divergence date between humans and chimpanzees as estimated from the dating of fossils (used for example in Mishmar et al., 2003), was proposed to be inappropriate to the dating of intra-species events (e.g., haplogroup coalescent ages; Endicott & Ho, 2008; Ho & Endicott, 2008) because of the rate heterogeneity that can exist among species (e.g., Douzery et al., 2003). A solution to this problem is internal calibration using dated aDNA sequences (Ho et al., 2005). The analysis of temporally sampled complete mtDNA sequences appears then as the most direct way to investigate and correct the existing biases in the estimation of the human mtDNA mutation rate and divergence dates (Ho & Gilbert, 2010).

3. In the field of ancient human DNA, efforts should be maintained to develop typing of nuclear markers in ancient human remains. Less copies of nuclear DNA are present per cell than copies of mtDNA; therefore, post-mortem degradation of DNA makes ancient nuclear DNA more difficult to recover than ancient mtDNA. As a consequence, the reliable retrieval of nuclear DNA from ancient human remains has previously necessitated optimal environmental conditions for DNA preservation in ancient samples. However, the answer to questions regarding the evolution of nuclear markers in human populations only requires access to limited nuclear information, most of the time Single Nucleotide Polymorphisms (SNPs). For example, lactase persistence in humans, i.e., the ability to digest milk, was investigated through typing of two SNPs (Burger et al., 2007). In this context, the impact of post-mortem DNA fragmentation becomes less of a problem in that appropriate methods for enrichment of short fragments of aDNA are developed and applied (Stiller et al., 2009; Briggs et al., 2009; Burbano et al., 2010; Maricic, et al., 2010). These methods may allow ancient nuclear data to be recovered from less optimally preserved samples that would not have yielded aDNA without a preliminary aDNA enrichment step. Enrichment methods could be applied to the typing of SNPs in the paternally-inherited non recombining part of the Y-chromosome and autosomal chromosomes.

The study of Y-chromosome data allows the paternal history of ancient human populations to be reconstructed, thus providing the complementary history to the maternal one from mtDNA data. Demographic and social structures of human societies (matriarchal or patriarchal), marriage patterns (matrilocal or patriloc), as well as male-driven expansions, can have contrasting impacts on the distribution of uniparentally-inherited markers such as mtDNA and the Y-chromosome. This is why the reconstruction of human population history should be based on the comparison of mtDNA and Y-chromosome data. It would be interesting to determine whether the Y-chromosome variability in the ancient populations examined in this study confirms the genetic affinities identified on the basis of mtDNA data. Y-chromosome data could then be used to confirm the genetic affinity of the populations of Mesolithic Uznyi Oleni Ostrov/Popovo and Bronze Age Bolshoy Oleni Ostrov in northeast Europe with present-day western and central Siberian populations (Chapter One). The retrieval of lineages belonging to Y-chromosome haplogroup N1 (M231) in ancient populations of northeast Europe would be a strong evidence for the genetic affinity with present-day populations of Siberia, where this haplogroup is found at its highest frequencies

(Derenko et al., 2007). Moreover, Y-chromosome data retrieved from historical Saami of the Chalmny-Varre graveyard of north east Europe (Chapter One) would indicate whether the genetic stability observed on the maternal side of the Saami for the last 200 years can also be observed on the paternal side. Genetic continuity between ancient and modern central Sardinians could be tested in the same way that it has been investigated for mtDNA data (Chapter Four). These analyses could reveal patterns of strong endogamy – in the case where Y-chromosome data also indicate genetic continuity since the Bronze Age – or matrilocality – in the case where Y-chromosome data are not consistent with local genetic continuity in central Sardinia. In addition, in the study of the origins and population history of Iron Age Scythians (Chapter Three), Y-chromosome data could put into perspective the mixed genetic makeup of this ancient population of nomads. Genetic information about the paternal genetic structure of Scythians could provide interesting insights about possible male driven population processes as observed in present-day populations of Central Asia or Neolithic individuals of southern France (Lacan et al., 2011).

Autosomal nuclear markers could be targeted in ancient individuals to reveal interesting information about population structure, population history, and human evolution. Genetic affinities identified on the basis of autosomal genetic data are less likely to be impacted by biases introduced by sex-specific population processes. SNPs situated on sex and autosomal chromosomes could provide valuable information about the sex of the specimens (e.g., amelogenin gene; e.g., Keyser et al., 2003), their phenotype (e.g., pigmentation genes, Lalueza-Fox et al., 2007), their genetic affinity with modern-day populations (e.g., ancestry informative markers; Keyser et al., 2009), as well as the evolution of metabolic functions (e.g., SNPs associated with lactase persistence, Burger et al., 2007) and diseases. Genetic affinity between ancient and modern populations could also be assessed from the study of a large number of nuclear SNPs (e.g., Jakobsson et al., 2008) that can be typed using hybridisation microarrays (e.g., Affimetrix genome-wide human SNP Array 6.0 targets 906,600 SNPs in the human genome). Finally, the (nearly) total amount of genetic information contained in ancient specimens, i.e., complete genomes, can potentially be retrieved using high through-put sequencing techniques ('next generation sequencing'), which have previously been applied to ancient human DNA (Rasmussen et al., 2010). While the application of these methods to the sequencing of complete human genomes is still limited today, rapid developments in high through-put techniques and in the analysis of

the data they generate could lead to the routine sequencing of ancient complete genomes in the coming decades.

With the development of high through-put sequencing techniques, we expect the amount of available genetic data from modern-day and ancient individuals to increase significantly in the near future. For this reason, statistical analysis methods taking into account the main features of ancient genetic data - sequencing errors, small sample sizes and diachronic sampling - should be developed and applied. The generation of a large amount of genetic data from ancient human remains broadly sampled over time and space would provide high resolution in the reconstruction of human genetic history.

REFERENCES

1. Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V., Moral, P., Dugoujon, J.M., Roostalu, U., Loogväli, E.L., Kivisild, T., Bandelt, H.-J., Richards, M., Villems, R., Santachiara-Benerecetti, A.S., Semino, O., Torroni, A. (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75(5), 910-918.
2. Anderson, C., Ramakrishnan, U., Chan, Y., Hadly, E. (2005). Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics* 21, 1733-1734.
3. Axelsson, E., Willerslev, E., Gilbert, M.T., Nielsen, R. (2008). The effect of ancient DNA damage on inferences of demographic histories. *Mol Biol Evol* 25(10), 2181-2187.
4. Beaumont, M.A., Zhang, W., Balding, D.J. (2002). Approximate Bayesian computation in population genetics. *Genetics* 162, 2025-2035.
5. Bennett, C.C., Kaestle, F.A. (2010). Investigation of Ancient DNA from Western Siberia and the Sargat Culture. *Hum Biol* 82 (2).
6. Bocquet-Appel, J.P., Naji, S., Vander Linden, M., Kozłowski, J.K. (2009). Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of 14C dates. *J Archaeol Sci* 36, 807-820.
7. Boyle, K., Renfrew, C. (2000) *Archaeogenetics: DNA and the population prehistory of Europe*. Cambridge: McDonald Institute for Archaeological Research. 342
8. Bramanti, B., Thomas, M., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M., Jankauskas, R., Kind, C., Lueth, F., Terberger, T., Hiller, J., Matsumara, S., Forster, P., Burger, J. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326, 137-140.
9. Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., Pääbo, S. (2009). Primer extension capture: target sequence retrieval from heavily degraded sources. *Science*. 325(5938), 318-21.
10. Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., Good, J. M., Maricic, T., Johnson, P. L., Xuan, Z., Rooks, M., Bhattacharjee, A., Brizuela, L., Albert, F. W., de la Rasilla, M., Fortea, J., Rosas, A., Lachmann, M., Hannon, G. J., Pääbo, S. (2010) Targeted investigation of the Neandertal genome by array-based sequence capture. *Science*. 328(5979), 723-5.
11. Burger, J., Kirchner, M., Bramanti, B., Haak, W., Thomas, M. G. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc Natl Acad Sci U S A* 104(10), 3736-41.
12. Caramelli, D., Vernesi, C., Sanna, S., Sampietro, L., Lari, M., Castri, L., Vona, G., Floris, R., Francalacci, P., Tykot, R., Casoli, A., Bertranpetit, J., Lalueza-Fox, C., Bertorelle, G., Barbujani, G. (2007). Genetic variation in prehistoric Sardinia. *Hum Genet* 122(3-4) 327-336.
13. Cooper, A., Poinar, H.N. (2000). Ancient DNA: do it right or not at all. *Science* 289(5482), 1139.
14. Chikisheva, T.A., Gubina, M.A., Kulikov, I.V., Karafet, T.M., Voevoda, M.I., Romaschenko A.G. (2007). A paleogenetic study of the prehistoric populations of the Altai. *Anthropology* 1(32), 130-142.

15. Derenko, M., Malyarchuk, B., Denisova, G., Wozniak, M., Grzybowski, T., Dambueva, I., Zakharov, I. (2007). Y-chromosome haplogroup N dispersals from south Siberia to Europe. *J Hum Genet* 52(9), 763-70.
16. Douzery, E. J., Delsuc, F., et al. (2003). Local molecular clocks in three nuclear genes: divergence times for rodents and other mammals and incompatibility among fossil calibrations. *J Mol Evol* 57 Suppl 1, S201-213.
17. Endicott, P., Ho, S.Y., Metspalu, M., Stringer, C. (2009). Evaluating the mitochondrial timescale of human evolution. *Trends Ecol Evol* 24(9), 515-9.
18. Ermini, L., Olivieri, C., Rizzi, E., Corti, G., Bonnal, R., Soares, P., Luciani, S., Marota, I., De Bellis, G., Richards, M.B., Rollo, F. (2008). Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr Biol* 18, 1687-1693.
19. Gilbert, M. T., Hansen, A.J., Willerslev, E., Rudbeck, L., Barnes, I., Lynnerup, N., Cooper, A. (2003). Characterization of genetic miscoding lesions caused by postmortem damage. *Am J Hum Genet* 72(1), 48-61.
20. Gilbert, M. T., Binladen, J., et al. (2007). Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Res* 35(1), 1-10.
21. Gilbert, M.T., Kivisild, T., Grønnow, B., Andersen, P.K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Götherström, A., Campos, P.F., Rasmussen, M., Metspalu M.; Higham, T.F., Schwenniger, J.L., Nathan, R., De Hoog, C.J., Koch, A., Møller, L.N., Andreasen, C., Medgaard, M., Villems, R., Bendixen, C., Willerslev, E. (2008). Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* 320, 1787-1789.
22. Gilbert, M. T., Binladen, J., Miller, W., Wiuf, C., Willerslev, E., Poinar, H., Carlson, J. E., Leebens-Mack, J. H., Schuster, S. C. (2007). Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Res.* 35(1):1-10.
23. Goloubinoff, P., Pääbo, S., Wilson, A.C. (1993). Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proc. Natl. Acad. Sci. U. S. A.* 90, 1997–2001.
24. Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K.W., Burger, J. (2005). Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310, 1016-1018.
25. Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., Der Sarkissian, C.S., Brandt, G., Schwarz, C., Nicklisch, N., Dresely, V., Fritsch, B., Balanovska, E., Villems, R., Meller, H., Alt, K.W., Cooper, A., Genographic consortium. (2010). Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. *PLoS Biol* 8, e1000536.
26. Handt, O., Höss, M., Krings, M., Pääbo, S. (1994). Ancient DNA: methodological challenges. *Experientia* 50 (6), 524-529.
27. Handt, O., Krings, M., Ward, R.H., Pääbo, S. (1996). The retrieval of ancient human DNA sequences. *Am J Hum Genet* 59(2), 368-376.
28. Ho, S., Phillips, M., Cooper, A., Drummond, A. (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol* 22, 1561-1568.
29. Ho, S., Endicott, P. (2008). The crucial role of calibration in molecular date estimates for the peopling of the Americas. *Am J Hum Genet* 83, 142-146; author reply 146-147.

30. Ho, S. Y., Gilbert, M. T. (2010). Ancient mitogenomics. *Mitochondrion*. 10(1), 1-11.
31. Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., Pääbo, S. (2001). Ancient DNA. *Nat Rev Genet* 2(5), 353-359.
32. Jakobsson, M., Scholz, S. W., Scheet, P., Gibbs, J. R., VanLiere, J. M., Fung, H. C., Szpiech, Z. A., Degnan, J. H., Wang, K., Guerreiro, R., Bras, J. M., Schymick, J. C., Hernandez, D. G., Traynor, B. J., Simon-Sanchez, J., Matarin, M., Britton, A., van de Leemput, J., Rafferty, I., Bucan, M., Cann, H. M., Hardy, J. A., Rosenberg, N. A., Singleton, A. B. (2008). Genotype, haplotype and copy-number variation in worldwide human populations., *Nature* 451(7181), 998-1003.
33. Keyser, C., Bouakaze, C., Crubézy, E., Nikolaev, V., 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.
34. 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.
35. Kolman, C. J., Tuross, N. (2000). Ancient DNA analysis of human populations. *Am J PhysAnthropol* 111(1), 5-23.
36. Krause, J., Briggs, A., 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.
37. Lacan, M., Keyser, C., Ricaut, F.X., Brucato, N., Duranthon, F., Guilaine, J., Crubézy, E., Ludes, B. (2011). Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *ProcNatlAcadSci U S A* 108(24), 9788-9791.
38. Lalueza-Fox, C., Sampietro, M., Gilbert, M., Castri, L., Facchini, F., Pettener, D., Bertranpetit, J. (2004). Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient Central Asians. *Proc Biol Sci* 271, 941-947.
39. Lalueza-Fox, C., Römpler, H., Caramelli, D., Stäubert, C., Catalano, G., Hughes, D., Rohland, N., Pilli, E., Longo, L., Condemi, S., de la Rasilla, M., Fortea, J., Rosas, A., Stoneking, M., Schöneberg, T., Bertrandoetit, J., Hofreiter, M. (2007). A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318, 1453-1455.
40. Li, C., Li, H., Cui, Y., Xie, C., Cai, D., Li, W., Mair, V.H., Zhang, Q., Abuduresule, I., Jin, L., Zhu, H., Zhou, H. (2010). Evidence that west-east admixed populations lived in the Tarim basin as early as the early Bronze Age. *BMC Biol* 8(15).
41. Malmström, H., Gilbert, M., Thomas, M., Brandström, M., Storå, J., Molnar, P., Andersen, P., Bendixen, C., Holmlund, G., Götherström, A., Willerslev, E. (2009). Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. *CurrBiol* 19, 1758-1762.
42. Malyarchuk, B., Derenko, M., 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.
43. Maricic T, Whitten M, Pääbo S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5(11), e14004.
44. Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., Sukernik, R.I., Olckers, A., Wallace, D.C. (2003). Natural selection shaped regional mtDNA variation in

- humans. *Proc Natl Acad Sci U S A* 100, 171-176.
45. Moolder, K.P., Weber, A.W., Bamforth, F.J., Lieveise, A.R., Schurr, T.G., Bazaliiski V.I., Savel'ev, N.A. (2005). Matrilineal affinities and prehistoric Siberian mortuary practices: a case study from Neolithic Lake Baikal. *J. Archaeol. Sci.* 32(4), 619-634.
 46. Pavlov, A., Pavlova, N.V., Kozyavkin, A., Slesarev, A.I. (2004). Recent developments in the optimization of thermostable DNA polymerases for efficient applications. *Trends Biotechnol.* 22, 253–260.
 47. Pereira, L., Richards, M., Goios, A., Alonso, A., Albarrán, C., Garcia, O., Behar, D.M., Gölge, M., Hatina, J., Al-Gazali, L., Bradley, D.G., Macaulay, V., Amorim, A. (2005). High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res* 15(1), 19-24.
 48. Pilipenko, A.S., Romaschenko, A.G., Molodin, V.I., Parzinger, H., Kobzev, V.F. (2010). Mitochondrial DNA studies of the Pazyryk people (4th to 3rd centuries BC) from northwestern Mongolia. *Archaeological and Anthropological Sciences* 2 (4), 231-236.
 49. Pinhasi, R., von Cramon-Taubadel, N. (2009). Craniometric data supports demic diffusion model for the spread of agriculture into Europe. *PLoS One.* 4(8), e6747.
 50. Price, T.D. (2000). *Europe's first farmers*. Cambridge: Cambridge University Press. 395 p.
 51. Pruvost, M., Schwarz, R., BessaCorreia, V., Champlot, S., Grange, T., Geigl, E-M. (2008). DNA diagenesis and palaeogenetic analysis: critical assessment and methodological progress. *PaleoPaleoPaleo* 266, 211-219.
 52. Rambaut, A., Ho, S. Y., Drummond, A.J., Shapiro, B. (2009). Accommodating the effect of ancient DNA damage on inferences of demographic histories. *Mol Biol Evol* 26(2), 245-248.
 53. Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J., S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., Bertalan, M., Nielsen, K., Gilbert, M., T., Wang, Y., Raghavan, M., Campos, P. F., Kamp, H., M., Wilson, A. S., Gledhill, A., Tridico, S., Bunce, M., Lorenzen, E. D., Binladen, J., Guo, X., Zhao, J., Zhang, X., Zhang, H., Li, Z., Chen, M., Orlando, L., Kristiansen, K., Bak, M., Tommerup, N., Bendixen, C., Pierre, T. L., Grønnow, B., Meldgaard, M., Andreassen, C., Fedorova, S. A., Osipova, L. P., Higham, T. F., Ramsey, C. B., Hansen, T. V., Nielsen, F. C., Crawford, M. H., Brunak, S., Sicheritz-Pontén, T., Villems, R., Nielsen, R., Krogh, A., Wang, J., Willerslev, E. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463(7282), 757-62.
 54. Ricaut, F.X., Keyser-Tracqui, C., Bourgeois, J., Crubézy, E., Ludes, B. (2004a). Genetic analysis of a Scytho-Siberian skeleton and its implications for ancient Central Asian migrations. *Hum Biol* 76, 109-125.
 55. Ricaut, F.X., Keyser-Tracqui, C., Cammaert, L., Crubézy, E., Ludes, B. (2004b). Genetic analysis and ethnic affinities from two Scytho-Siberian skeletons. *Am J Phys Anthropol* 123, 351-360.
 56. Richards, M., Côrte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H., Sykes, B. (1996). Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59, 185-203.
 57. Richards, M., Macaulay, V., Bandelt, H., Sykes, B. (1998). Phylogeography of mitochondrial DNA in Western Europe. *Ann Hum Genet* 62, 241-260.

58. 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., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Renzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A., Bandelt, H-J. (2000). Tracing European founder lineages in the near Eastern mtDNA pool. *Am J Hum Genet*, 1251-1276.
59. Richards, M. (2003). The neolithic invasion of Europe. *Annu Rev Anthropol* 32, 135–162.
60. Sampietro, M., Lao, O., Caramelli, D., Lari, M., Pou, R., Marti, M., Bertrandpetit, J., Lalueza-Fox, C. (2007). Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc Biol Sci* 274(1622), 2161-2167.
61. Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A., Salas, A., Oppenheimer, S., Macaulay, V., Richards, M. (2009). Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *American Journal of Human Genetics*, 740-759.
62. Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H-J., Torroni, A., Richards, M.B. (2010). The archaeogenetics of Europe. *Curr. Biol.* 20(4): R174-R183.
63. Stiller, M., Green, R.E., Ronan, M., Simons, J.F., Du, L., He, W., Egholm, M., Rothberg, J.M., Keates, S.G., Ovodov, N.D., Antipina, E.E., Baryshnikov, G.F., Kuzmin, Y.V., Vasilevski, A.A., Wuenschell, G.E., Termini, J., Hofreiter, M., Jaenicke-Despres, V., Pääbo, S. (2006). Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proc Natl Acad Sci U S A* 103, 13578-13584.
64. Stiller M, Knapp M, Stenzel U, Hofreiter M, Meyer M. (2009). Direct multiplex sequencing (DMPS)--a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. *Genome Res* 19(10), 1843-8.
65. Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogväli, E.L., Tolk, H.V., Reidla, M., Metspalu, E., Pliss, L., Balanovsky, O., Pshenichnov, A., Balanovska, E., Gubina, M., Zhadanov, S., Osipova, L., Damba, L., Voevoda, M., Kutuev, I., Bermisheva, M., Khusnutdinova, E., Gusar, V., Grechanina, E., Parik, J., Pennarun, E., Richard, C., Chaventre, A., Moisan, J.P., Barác, L., Perici, M., Rudan, P., Terzi, R., Mikerezi, I., Krumina, A., Baumanis, V., Koziel, S., Rickards, O., De Stefano, G.F., Anagnou, N., Pappa, K.I., Michalodimitrakis, E., Ferák, V., Füredi, S., Komel, R., Beckman, L., Villems, R. (2004). The Western and Eastern roots of the Saami--the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *Am J Hum Genet* 74, 661-682.
66. Torroni, A., Bandelt, H.-J., Macaulay, V., Richards, M., Cruciani, F., Rengo, C., Martinez-Cabrera, V., Villems, R., Kivisild, T., Metspalu, E., Parik, J., Tolk, H.V., Tambets, K., Forster, P., Karger, B., Francalacci, P., Rudan, P., Janicijevic, B., Rickards, O., Savontaus, M.L., Huoponen, K., Laitinen, V., Koivumäki, S., Sykes, B., Hickey, E., Novelletto, A., Moral, P., Sellitto, D., Coppa, A., Al-Zaheri, N., Santachiara-Benerecetti, A.S., Semino, O., Scozzari, R. (2001). A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 69(4): 844-852.

67. van Oven, M., Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30, E386-394. www.phylotree.org/
68. von Cramon-Taubadel, N., Pinhasi, R. (2011). Craniometric data support a mosaic model of demic and cultural Neolithic diffusion to outlying regions of Europe. *Proc Biol Sci*.
69. Whittle, A.W.R. (1996). *Europe in the Neolithic: the creation of new worlds*. Cambridge: Cambridge University Press. 443 p.
70. Whittle, A.W.R., Cummings, V. (2007) *Going over: the mesolithic-neolithic transition in North-West Europe*. Oxford: Oxford University Press. 632 p.
71. Zvelebil, M. (1986). Mesolithic societies and the transition to farming: problems of time, scale and organization. In *Hunters in transition: Mesolithic societies of temperate Eurasia and their transition to farming* (eds M. Zvelebil, R. Dennell & L. Domanka), pp. 167–188. Cambridge, UK: Cambridge University Press.
72. Zvelebil, M. (1996). The agricultural frontier and the transition to farming in the circum-Baltic area. In *The origins and spread of agriculture and pastoralism in Eurasia* ed. D. R. Harris