Ancient DNA studies of human evolution

Christina Jane Adler

Australian Centre for Ancient DNA

School of Earth and Environmental Sciences

The University of Adelaide

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

January 2012

CONTENTS PAGE

CHAPTER ONE Introduction……. 1
1.1 The aims of the thesis…….. 3
1.2 Neolithic Revolution……... 5
1.2.1 Origins and reasons for the development of agriculture……. 5
1.2.2 The introduction of the Neolithic Package to Europe and the influence of the transition to agriculture on the population structure of modern Europeans……... 9
1.2.2.1 Support for the cultural diffusion hypothesis – Pleistocene settlement events ……... 13
1.2.2.2 Support for the demic diffusion hypothesis – Neolithic settlement events…… 17
1.2.3 Dietary changes associated with the introduction of agriculture……... 24
1.2.4 Health changes associated with the introduction of agriculture……. 27
1.3 Application of ancient DNA to study human evolutionary questions……. 30
1.3.1 Technical challenges of using ancient DNA…….. 31
1.3.2 Using ancient human DNA to study past demographic events…….. 34
1.3.3 Using ancient bacterial DNA to study the evolution of pathogens…… 36
1.3.3.1 Dental calculus……... 39
1.4 Thesis structure……..……. 46
References…….. 58

CHAPTER TWO Adler, C J, Haak, W, Donlon, D & Cooper, A 2011. Survival and recovery of DNA from ancient teeth and bones. *Journal of Archaeological Science*, 38, 956-964..…… 62
Supplementary material……... 72

CHAPTER THREE Ancient DNA reveals that population dynamics during the Late Neolithic and Early Bronze Age substantially influenced the population structure of modern Central Europeans……. 91
3.1 Introduction……..……... 92
3.2 Methods…… 96
3.2.1 Sample description and preparation……... 96
3.2.2 DNA extraction, PCR, real time PCR and sequencing…….. 97
3.2.3 Authentication criteria……... 98
3.2.4 Statistical analysis…….. 99
3.3 Results…….. 104
3.4 Discussion…… 108
3.5 Conclusion……... 111

CHAPTER FOUR. Adler, C J, Dobney, K D, Walker, A W, Haak, W, Bradshaw, C J A, Townsend, G, Kaidonis, J, Alt, K, Parkhill, J & Cooper, A. Ancient DNA from oral microbiota records the impact of dietary change on the evolution of human oral diseases. Prepared for submission to *Nature*……..….. **137**

Supplementary Material……... 155

CHAPTER FIVE Conclusion……... 189

5.1 Overall summary of the aims of the thesis……... 189

5.2 Summary of Aim 1……..... 190

5.2.1 Significance…… 190

5.2.2 Limitations of the research and future directions……... 191

5.3 Summary of Aim 2……..... 194

5.2.1 Significance…… 194

5.2.2 Limitations of the research and future directions……... 196

5.4 Summary of Aim 3……..... 199

5.2.1 Significance…… 200

5.2.2 Limitations of the research and future directions……... 202

5.5 Concluding remarks……. 205

References……. 207

Appendix A Jones, C. 2011. Researchers to drill for hobbit history. *Nature*, doi:10.1038/news.2011.702……... 214

Appendix B Haak, W, Balanovsky, O, Sanchez, J J, Koshel, S, Zaporozhchenko, V, Adler, C J et al., 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biol*, 8, e1000536…… 217

Abstract

The recovery of ancient DNA from preserved human remains provides a unique opportunity to directly study the impact of past events on prehistoric populations. The past event which is of interest to this thesis is the development of agriculture in Europe, also termed the 'Neolithic Revolution'. This event resulted in the transition from a hunter-gatherer to agriculturist lifestyle. The adoption of farming altered many aspects of human culture, including how societies are structured, where we obtain our food from and which diseases burden human populations. This thesis aimed to assess how the arrival of farming in Europe influenced the population structure and health of both prehistoric and modern human populations through the analysis of ancient genetic data. As human remains are a valuable resource, it was also of interest to evaluate the methods used to recover degraded genetic material from preserved remains.

The affect of common sampling techniques on the recovery of DNA from skeletal remains was assessed by quantifying mitochondrial (mt) DNA in ancient human and bovid samples (n=42). The key finding is that drilling of bone and teeth under standard sampling conditions can reduce DNA amounts by over five orders of magnitude. Also, it was shown that current approaches do not sample the most DNA rich tissues (the cementum layers of tooth roots) and have failed to show where DNA survives accurately and how it degrades over time. The results promote a standardised approach to the genetic analysis of human remains.

To investigate whether a single, continuous population has existed in Central Europe since the Early Neolithic until today, a dataset of ancient mtDNA from Early Neolithic and Late Neolithic/Early Bronze Age cultures from Central Europe was compiled (n=108). Population continuity from the introduction of agriculture in the Early Neolithic until today was found to be an unlikely model of demographic history, as indicated by the results from population genetic and coalescent analyses of the ancient mtDNA, which was compared to extant populations from the same region. Internal population changes in Europe between the Early Neolithic and Late Neolithic/Early Bronze Age appear to have contributed substantially to the population structure of extant Central Europeans, as all the Late Neolithic/Early Bronze Age cultures examined (Bell Beakers, Corded Ware or Unetice) were found to be more likely ancestors of modern Central Europeans than either of the Early Neolithic cultures (LBK and Rössen) investigated. Haplogroup distributions suggest that Palaeolithic mtDNA haplogroups which were infrequent in the Early Neolithic, such as haplogroups H and U, became more frequent during the Late Neolithic. These findings alter the traditional view of which past settlement events in Europe have contributed to the current genetic structure of the continent and demonstrate that past events which are not associated with major climatic or economic

changes in human history, such as internal migrations in Europe during the Late Neolithic/Early Bronze Age, can also substantially alter population structure.

Ancient pathogen DNA was used to explore how farming affected the health of prehistoric humans. This required the identification of a genetically preserved source of human-associated microbiota. Dental calculus samples recovered from human remains, which dated between the Neolithic to Medieval period, were found to contain ancient oral microbial DNA. Preserved microbial DNA was extracted from calculus samples (n=28) and used to create PCR amplicon libraries of the 16S rDNA gene. Phylogenetic analyses revealed that the diversity of oral microbiota in early agriculturists was significantly higher than that found in modern Europeans. During the early stages of agriculture there was a dominance of bacteria involved in periodontal disease. This suggests periodontal disease was an early consequence of an agricultural diet, which was high in carbohydrates and low in processed sugar and grain. In contrast, modern Europeans have a restricted suite of bacteria, mainly associated with tooth decay. The current dominance of decay associated taxa may be because of the introduction of refined sugar and grain to the diet which occurred during the Industrial Revolution (150-200 years ago). The results highlight the impact of changes in human ecology to the co-evolved mutualism between humans and their microbiota.

By using three disparate approaches to research, this thesis has provided original insights into the lives of prehistoric humans. The results have demonstrated that accompanying the early stages of agriculture, there were changes in population structure and the presence of bacteria associated with oral diseases. The application of ancient DNA techniques to examine dental calculus, enabled for the first time, the direct analysis of the past state of human associated microbiota, which will have multiple implications for the study of prehistoric human lifestyle. This thesis also contributed more generally to the field of ancient DNA by highlighting that commonly used sampling techniques were detrimental to the recovery of DNA from skeletal remains. By combining three different approaches to ancient DNA research, this thesis has provided original insights into the lives of prehistoric humans, which advances the fields of archaeology, anthropology, modern and ancient population genetics, forensics and medical research.

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 Christina Jane Adler 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.

The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holder(s) of those works. 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.

The published works in my thesis include:

Adler, C J, Haak, W, Donlon, D & Cooper, A 2011. Survival and recovery of DNA from ancient teeth and bones. *Journal of Archaeological Science,* 38**,** 956-964.

Jones, C. 2011. Researchers to drill for hobbit history. *Nature*, doi:10.1038/news.2011.702.

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 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biology*, 8, e1000536.

……..

Christina Jane Adler

9th January 2012

Acknowledgements

First and foremost I would like to thank the people who supervised my PhD, Professor Alan Cooper, Dr Wolfgang Haak and Dr Denise Donlon. Their guidance, assistance and encouragement during my PhD will always be appreciated. Also, I would like to thank the Australian Research Council and the School of Earth and Environmental Sciences at The University of Adelaide for supporting my candidature.

I would like to recognise my fellow lab members at the Australian Centre for Ancient DNA for their support and wide range of knowledge, which was freely shared. In particular, I would like to thank those who gave their time to proof read my thesis, including Nicolas Rawlence and Elen Shute. I would also like to thank my fellow officemates, including Fran MacGilvary, Patricia Cross and Pancho Gonzalez Pinilla, who have provided great support and friendship. In particular, I would like to thank Kym Abrams for her friendship and constant supply of coffee. As well as members of my School, I have also been greatly assisted by Professor Grant Townsend and Dr John Kaidonis from the Dental School at The University of Adelaide. I would also like to extend a big thank you to Professor Keith Dobney from The University of Aberdeen, whose support has been essential to the research carried out in this PhD.

My final thank you is to my family and friends. In particular, I would like to thank my Mum and Dad for their continued support and care during my PhD, and officially recognise that my Dad read nearly every word of my thesis and provided helpful feedback. I am especially thankful to my understanding fiancé, Adam Skinner, who has generously given his time to read my thesis, and provide careful and considered feedback.

CHAPTER ONE

Introduction

For the majority of human existence (circa 200, 000 years (McDougall et al., 2005)), hunted and gathered wild foods have formed the basis of the human diet (Milton, 1999). In the last 10, 000, years humans have drastically altered both their diet and how they obtain food by modifying wild plants and animals to create domesticate populations (Braidwood et al., 1961). The domestication of plants and animals by hunter-gatherers and the adoption of an agricultural lifestyle was one of the most profound events in human history, representing a 'totally new concept of how to make a living' (Gronenborn, 1999).

The focus of this thesis is the development of agriculture in Europe, also termed the 'Neolithic Revolution' (Childe, 1925). In the Near East, the Neolithic period began approximately 10, 000 years ago (7000 Before Christ (BC)) (Pinhasi et al., 2005) with the domestication of wild grasses (Harlan and Zohary, 1966) and, subsequently, the development of animal husbandry (Reed, 1959). Agricultural practices later spread from the Near East across Europe, replacing the traditional lifestyle of the indigenous hunter-gatherers (Gronenborn, 1999). This spread occurred at a rate of approximately 1 km per year according to regression analysis of radiocarbon dates (Ammerman and Cavalli-Sforza, 1971).

The transition from a hunter-gatherer to agriculturist lifestyle in Europe associated with population movement (Price et al., 2004, Haak et al., 2010), cultural differentiation (Gronenborn, 1999), drastically altered diet (Tauber, 1981, Richards et al., 2003b), the spread of previously unseen infectious diseases (Wolfe et al., 2007) and the emergence of the Indo-European language, which gave rise to the majority of modern European languages (Gimbutas, 1963, Renfrew, 1987). Agriculture changed many aspects of human culture. In particular, a sedentary farming lifestyle initiated social stratification (Salamini et al., 2002), which led to the centralisation of power and the formation of cities and states (Weisdorf, 2005). The development of farming was also a precursor to the industrialisation of food production (Galor and Moav, 2002). Industrialisation enabled people to dedicate time to activities other than growing food, so that society could support a sector of individuals who were 'non-food specialists' (Weisdorf, 2004), including scientists, economists and lawyers. The creation of this sector of individuals contributed to the accumulation of wealth in societies and the establishment of what is referred to as 'civilisation'.

As many of modern society's characteristics can be attributed to the development of agriculture, there is a long history of research examining the Neolithic Revolution, with archaeological studies dating back over a century (Breasted, 1906-7). Investigations of the Neolithic Revolution have been multidisciplinary, with studies conducted in a variety of research fields, including anthropology, archaeology, economics, genetics and linguistics. Within these fields of research, there has been debate concerning the reasons why agriculture arose, how the transition to farming occurred in Europe and hence how this transition influenced the settlement history of Europe. In addition, investigations have focused on how farming-related lifestyle changes, such as alterations in diet and social organisation, have affected the health of modern populations.

Although the Neolithic Revolution has played a crucial role in European history, the influence of the Neolithic period on the population structure and disease state of modern Europeans remains incompletely understood. Population studies of modern human DNA have presented inconsistent scenarios to describe how farming spread from the Near East to Europe and hence how the transition to agriculture has influenced the settlement history of Europe (Ammerman and Cavalli-Sforza, 1984, Sokal et al., 1991, Semino et al., 1996, Torroni et al., 1998, Richards et al., 2000a, Chikhi et al., 2002, Dupanloup et al., 2004). While ancient human DNA studies of early European farmers have revealed that the introduction of agriculture, at least in Central Europe, was associated with the influx of migrants with genetic similarities to modern Near Eastern populations (Bramanti et al., 2009, Haak et al., 2010), the contribution of the first farmers and later Neolithic and Early Bronze Age cultures to the genetic makeup of contemporary Central Europeans is still unresolved.

The introduction of agriculture to Europe had a major impact on health, with agriculturist communities displaying an increased prevalence of infectious diseases, nutritional disorders and stress makers, compared to hunter-gatherers (Larsen, 1995, Aufderheide et al., 1998, Richards, 2002). Human associated pathogens are thought to have increased in abundance in response to the transition of hunter-gatherer bands into large complex societies associated with domesticated animals and accompanying zoonoses (Taylor et al., 2001, Wolfe et al., 2007). Disease has played a major role in human history, with global pandemics having periodically swept through populations around the world (Zhang et al., 2007). Perhaps, one of the best known examples is the Black Death in the 14th Century, which is thought to have killed 30-60% of the population in both China and Europe, and caused a series of religious, social and economic upheavals with major impacts on the course of human history (Zhang et al., 2007). Despite the critical role of disease in human history, there is limited knowledge of

how diseases associated with farming originated and evolved, because prehistoric human remains are generally restricted to skeletal hard parts (bones and teeth), which preserve little or no direct evidence of the nature or past diversity of disease (Wood et al., 1992). Palaeopathological lesions are rarely diagnostic to a specific pathogen, because diseases often present in a variety of forms (Wood et al., 1992, Miller et al., 1996). As a result, there is limited knowledge about how present-day diseases that originated with the development of agriculture have changed over time.

1.1 The aims of the thesis

In this PhD, ancient DNA is used to investigate the Neolithic Period and the impact of the introduction of agriculture on the genetic makeup and health of current European populations. The Neolithic Revolution is investigated using both ancient human DNA and human-associated microbial DNA recovered from European remains. Ancient DNA methods were used to track genetic changes across time and space. In comparison, the analysis of modern DNA can only be used to infer the genetic signature of past events and is often confounded because the genetic composition of contemporary populations reflects the entire history of a group, which may include multiple occurrences of admixture between populations, as well bottlenecks (sharp reduction in effective population size).

It is critical that the smallest possible amount of destructive sampling is involved in the recovery of genetic material as human remains are an extremely valuable resource for multiple fields, such as morphological examination (Hillson, 1979, Hanihara, 2008) and isotopic analysis (Bentley, 2006), in addition to possessing cultural significance (Donlon, 1994). To minimise the amount of destructive sampling, areas of the skeleton should be targeted which contain the greatest amount of DNA. However, surprisingly, the distribution of DNA in preserved remains and the impact sampling techniques on the recovery of DNA had not been determined. A significant component of the research presented in this thesis involves the assessment of the impact of DNA sampling techniques on the amount of DNA recovered from preserved hard remains, which include the skeleton and teeth.

This thesis brings together three different areas of ancient DNA research to investigate the impact of the Neolithic Revolution on prehistoric and modern populations; 1) the technical assessment of current DNA sampling methods for skeletal remains, 2) the analysis of past populations genetics, and 3) the study of changes in the composition and diversity of pathogens during biological and cultural transitions. These three disparate approaches provide

complementary perspectives on the influence of the introduction of agriculture to past and present European populations. The specific aims are detailed below:

Aim 1: To quantify the amount of DNA in skeletal remains and assess the impact of commonly used skeletal sampling techniques on the recovery of DNA**.** To quantify the amount of DNA in skeletal remains, the relative concentrations of DNA within different tooth tissues will be assessed. To determine the impact of sampling methods on extracted DNA, common techniques such as drilling and pulverising will be assessed. In addition, the role of environmentally-influenced degradation rates on DNA survival in specimens will be examined. All these factors will be assessed to enable the reliable study of DNA from human skeletal remains.

Aim 2: To determine if a single, continuous population has existed in Central Europe since the introduction of agriculture in the Early Neolithic until today. To examine the likelihood of population continuity though out the Neolithic period in Central Europe, ancient human mitochondrial (mt) DNA will be recovered from Central European remains that span the Early Neolithic (5500 – 4250 BC) to the Late Neolithic and Early Bronze Age period (2700 – 2050 BC). Population genetic analyses will be performed to compare the Neolithic and Bronze Age cultures to each other and to modern Central European populations, and the results will be used to infer the likelihood of population continuity from the Early Neolithic until today. The population genetic analyses of the temporally-spaced ancient DNA dataset will be used to examine whether the Early or Late Neolithic and Bronze Age cultures are more likely to be the ancestors of modern Central Europeans.

Aim 3: To demonstrate that archaeological dental calculus (calcified plaque on teeth) preserves genetic material from ancient oral microbes. To demonstrate that ancient dental calculus deposits contain oral microbial DNA, the taxonomic and phylogenetic relationship between microbial sequences obtained from ancient dental calculus, modern oral samples (plaque and saliva) and sources of potential contamination, including laboratory reagents and environmental materials (soil, sediment and water), will be described. To investigate whether the introduction of an agricultural diet has influenced the composition of the human oral microbiota, the bacterial diversity of ancient dental calculus samples from early European farming groups will be compared to the oral microbial composition of modern European populations. In addition, the composition of oral microbiota in dental calculus from several prehistoric cultural groups will be compared. The results will provide the first direct genetic evidence of

how lifestyle changes associated with the introduction of agriculture have influenced the health of past populations.

1.2 Neolithic Revolution

The development of agriculture in the 'Fertile Crescent' of the Near East (Breasted, 1906-7, Braidwood et al., 1961), which encompasses the modern-day countries of Israel, Lebanon, Iraq and Jordan (Gopher et al., 2002), largely halted the lifestyle of hunting and gathering food, and established the social, economic and biological characteristics of modern societies. For example, the social structure of humans was altered with the shift from a hunter-gatherer lifestyle, typified by seasonal, nomadic movement, driven by the need to find food (Testart et al., 1982, Gamble, 1999), to an agricultural lifestyle, which due to the presence of year round harvesting and permanent settlements in early farming communities (Savard et al., 2006) has been suggested to be sedentary (Savard et al., 2006). A sedentary lifestyle resulted in large numbers of individuals living permanently in close proximity, which gave rise to the development of complex social interactions, organised by households (Bar-Yosef and Belfer-Cohen, 1992). The organisation of society by households has been seen as the precursor for the social stratification of modern populations (Bar-Yosef and Belfer-Cohen, 1989, Salamini et al., 2002). A sedentary, agricultural lifestyle also provided a predictable supply of food, resulting in the growth of populations in concentrated areas (Shennan and Edinborough, 2007). The development of now widespread diseases associated with the development of agriculture (Cohen and Armelagos, 1984, Aufderheide et al., 1998), such as leprosy (Roberts et al., 2002) and tuberculosis (Formicola et al., 1987) is thought to have occurred due to the increased size and density of populations (Shennan and Edinborough, 2007, Wolfe et al., 2007). In addition, the spread of infectious diseases in agricultural populations is thought to have been associated with their close living arrangements with domesticated animals (Larsen, 1995, Diamond, 1997). Furthermore, the adoption of a cereal based diet is thought to have resulted in the development among agricultural populations of oral diseases (Marsh, 2010) and nutritional disorders (Larsen, 1995).

1.2.1 Origins and reasons for the development of agriculture

Agriculture was adopted by hunter-gatherer populations multiple times, independently in a number of locations (Gronenborn, 1999). Different plants were domesticated in isolation in a number of environments; wheat and barley in the Near East (Harlan and Zohary, 1966), rice and millet in China (Londo et al., 2006, Zong et al., 2007) and maize in Mesoamerica, which

encompasses modern day central Mexico, Belize, Guatemala, El Salvador, Honduras, Nicaragua, and Costa Rica (Flannery, 1973, Pope et al., 2001). A domesticate plant or animal is distinct from the wild progenitor as it has been selected by humans, potentially based on a certain advantageous feature, such as successful reproduction, and has then been used to create a distinct domesticate population (Hillman and Davies, 1990). According to this definition of domestication, there is only inconclusive archaeological and botanical evidence to support the proposed independent development of plant domestication in Africa, the Amazon and the highlands of Papua New Guinea (Smith, 1998, Diamond, 2002). For example, the strongest evidence that plants were manipulated in Papua New Guinea prior to the introduction of yams from South East Asia approximately 3000 years ago (Sauer, 1952), is the recovery of 7000 year old banana phytoliths and taro starch grains from Kuk swamp, that is located in the highlands (Denham et al., 2003). However, these plants appear to have been cultivated, not domesticated, as indicated by the morphological similarity between the recovered phytoliths and starch grains to wild varieties of banana and taro, respectively (Denham et al., 2003, Neumann, 2003). In comparison to the questionable development of agriculture in Papua New Guinea, there is little doubt from archaeological evidence that farming originated in the Near East at the onset of the Neolithic (Childe, 1925, Bar-Yosef and Goren, 1973).

The development of agriculture in the Near East was associated with both the domestication of plants and animals, and the appearance of an archaeological assemblage referred to as the 'Neolithic Package'. This archaeological package consisted of permanent settlements, typically in the form of long houses, pottery and inhumations with grave goods (Childe, 1925, Braidwood et al., 1961). The spread of the Neolithic Package was gradual, taking over five centuries to travel from the Near East to Europe (Gronenborn, 1999, Kuijt and Goring-Morris, 2002). The Natufian culture (11,100 – 7,700 BC) was the earliest group to display some of the traits of the Neolithic Package in the Near East (Kuijt and Goring-Morris, 2002). The traits displayed by the Natufians which were characteristic of the Neolithic Package included a sedentary lifestyle, as they lived in pit houses (Stekelis and Yizraeli, 1963, Bar-Yosef and Goren, 1973), and the use of grave goods, such as necklaces and decorations, made from marine shells, bones and teeth (Bar-Yosef and Belfer-Cohen, 1992). However, the Natufians were essentially sedentary hunter-gatherers, with no indication they domesticated plants (Bar-Yosef and Goren, 1973). The lack of evidence for plant domestication may be due to the poor preservation of botanical remains at Natufian archaeological sites, which are typically found in exposed locations (Bar-Yosef, 1998). Additionally, plant remains may not have been recovered from Natufian sites, as the majority of excavations were conducted prior to the

1960's, before the routine use of plant collection techniques such as dry-sieving and flotation of soils (Willcox, 1996, Bar-Yosef, 1998). However, the few dated-grains recovered from Natufian sites have turned out to be recent (Hedges et al., 1992). This suggests that while Natufians had begun to display aspects of the Neolithic Package, this group were yet to domesticate plants.

In the Near East, the Natufian culture was succeeded during the early Holocene (approximately 10, 000 BC) by the Pre-Pottery Neolithic culture or stage (9,750 – 6,450 BC) (Kenyon, 1956, Kuijt and Goring-Morris, 2002). During the Pre-Pottery stage, plants such as wheat and barley were cultivated (Hillman and Davies, 1990), and animals were domesticated (Bar-Yosef and Belfer-Cohen, 1992). Debate has focused on whether wild plants were extensively collected and cultivated, or domesticated during the Pre-Pottery stage (Hillman and Davies, 1990, Willcox, 1996). Either way, there was an unprecedented use of plants in the Near East during the Pre-Pottery stage as indicated by the large amounts of plant material recovered from early Neolithic sites compared to elsewhere in the world at that time in history (Willcox, 1996, Smith, 2001, Kuijt and Goring-Morris, 2002). Plants recovered from Pre-Pottery Neolithic sites include cereals, such as wheat, barley and rye, pulses (e.g. lentils and peas) and wild almond (Willcox, 1996, Smith, 2001, Kuijt and Goring-Morris, 2002).

Towards the end of the Pre-Pottery Neolithic stage in the Near East, animal husbandry developed in the Levant (modern day Jordan, Palestine, Israel, Syria, the Sinai peninsula of Egypt and southeast Turkey) (Bar-Yosef and Belfer-Cohen, 1992). Domesticated animals included goat, pig, cattle and, possibly, sheep (Kohler-Rollefson et al., 1988). The domesticated animals were morphologically distinguishable from their wild ancestors, displaying reduced sexual dimorphism, smaller brains, bodies and horns, and altered horn shape (Davis, 1981, Gepts and Papa, 2001). The full blown presentation of the Neolithic Package, which included domesticated plants and animals, permanent settlements, inhumations with grave goods, and pottery, was observed in the Pottery Neolithic stage (6300–5850 BC) (Kuijt and Goring-Morris, 2002). Many of the archaeological features of the Pottery stage were very similar to the Pre-Pottery stage, with plant and animal domestication, sedentism and inhumations observed, with the addition of ceramics (Akkermans et al., 2006, Asouti, 2006).

The reasons for the development of agriculture during the Pre-Pottery Neolithic stage in the Near East are thought to be related to both environmental pressures and social causes. The potential environmental causes for the development of agriculture include: the declining availability of herding animals caused by either over-hunting (Smith, 1975, Smith, 1998) or

environmental changes associated with climate variation during the late Pleistocene epoch (120, 000 to 10, 000 years before present) (Richerson et al., 2001, Barnes et al., 2002, Guthrie, 2003, Shapiro et al., 2004); the development of a more stable, wetter climate during the Holocene (Kutzbach, 1981, Davis et al., 2003) better suited to plant domestication (Childe, 1935); or, the pressure of an increasing population size that required a more reliable food supply (Cohen, 1977, North and Thomas, 1977). Counter-arguments have been made against each of these proposed environmental hypotheses which attempt to explain the reasons for the development of farming. The declining number of herding animals during the late Pleistocene was probably not the driver of agricultural development, as animal populations appear to have been already in decline (Richerson et al., 2001, Barnes et al., 2002, Guthrie, 2003, Shapiro et al., 2004). For example, analysis of ancient DNA from bison remains has revealed that 15, 000 years before the introduction of farming, the genetic diversity of bison had severely declined, indicating a reduction in population size (Shapiro et al., 2004). This indicates that climate changes before the introduction of agriculture had probably impacted the number of other large herding animals across the world. Also, even though the number of herding animals may have declined before the onset of agriculture, there were still plentiful amounts of food during the Holocene in the Levant region, including wild cereals, legumes and fruits (Bar-Yosef, 1998). Climate change has been suggested to be too gradual to have caused the abrupt behavioural change which is associated with the development of agriculture (Perrot, 1962). Of the proposed environmental hypotheses, there has been least support for the idea that an increase in the population size of hunter-gatherers forced the development of agriculture out of the necessity to provide sufficient food (Weisdorf, 2004). As a result, it is nearly impossible to determine if the population expansion observed in terminal hunter-gatherer groups, such as the Natufians and the early Pre-Pottery Neolithic cultures of the Near East (Bar-Yosef, 1998), was the driver or product of agricultural practices.

Social factors have also been proposed to explain the development of agriculture (Hayden, 1990, Weisdorf, 2004). Farming has been described as a product of the innate human desire to ‘civilise’ (Hayden, 1990, Weisdorf, 2004). The civilising argument proposes that agricultural development was driven by the aspiration of humans to improve their social standing through the production of delicacies (Hayden, 1990). Also, agriculture has been seen as a consequence of humans’ urge to develop diverse societies which include individuals whose occupation is not dedicated to generating food (Weisdorf, 2004). However, both these social factors appear to be products of agriculture rather than drivers, with the production of delicacies and the non-food sectors of society only substantially developing during the Industrial Revolution

(Galor and Moav, 2002). Presently, no single hypothesis that has been proposed to explain the origins of agriculture has received the majority of support from the research community.

The fact that no single explanation for the development of farming has been agreed upon by researchers, indicates that the stimulus for agricultural development may have been multifaceted. For example, Richerson and colleagues (2001) have proposed that both climate change and humans innate ability to domesticate plants and animals, together could have been the stimulus for the development of agriculture. The domestication of plants and animals has been proposed to be an innate human ability, as agriculture has developed independently, at various times in history and in varying environmental contexts (Darwin, 1868). Hence, humans could always develop agriculture, however were limited in doing so until the climate became suitable for plant domestication (Richerson et al., 2001). Richerson and colleagues (2001) hypothesis would explain why the adoption of agriculture in the Near East and other areas, such as China and Mesoamerica, only occurred during the Holocene, when climates became suitable for plant domestication.

1.2.2 The introduction of the Neolithic Package to Europe and the influence of the transition to agriculture on the population structure of modern Europeans

Agricultural practices from the Near East spread to Europe during the Neolithic period (5700-2050 BC), and the adoption of farming resulted in the abolition of the indigenous hunter-gatherer lifestyle (Price, 1987, Price, 1991, Gronenborn, 1999, Price et al., 2001, Soares et al., 2010). The Neolithic package spread from the Levant into the Carpathian Basin of East-Central Europe (by approximately 6 000 BC), which covers the modern-day countries of Hungary and Slovakia, and parts of Serbia, Croatia, Romania, Slovenia, Austria and the Ukraine (Price et al., 2001). In the Carpathian Basin, the adoption of agricultural practices from the Levant has been determined by the recovery of Near Eastern plant (wheat and barley) and animal (cattle, goat, pig and sheep) domesticates, found in association with the early Neolithic, Starcevo-Körös-Cris culture complex (Bogucki, 1996, Gronenborn, 1999). Following the establishment of the Neolithic Package in the Carpathian Basin, farming practices spread to Europe by two main routes, which can be traced by the varied pottery styles found along these routes; the Danubian rivers route into Central, and eventually Northern Europe, by the Linear Pottery Culture (*Linearbandkeramik,* LBK) (5700-5400 BC), which is distinguished by the presence of incised-line pottery (Gronenborn, 1999, Price et al., 2001), and the Mediterranean coastal route westward to Spain by the Cardial Ware culture, typified by the appearance of impressed pottery (Price, 1987, Zilhao, 2001). As agricultural practices spread from the Carpathian Basin to Central-Northern and Western Europe, there

was variation across the continent in the mode by which the hunter-gatherer lifestyle was displaced.

The archaeological record suggests that while the adoption of the Neolithic Package in Europe resulted in the near-complete replacement of the hunter-gatherer lifestyle in Central Europe, hunter-gatherer elements were only partially displacement in Northern and Western Europe (Price, 1987, Price, 1991, Gronenborn, 1999, Price et al., 2001, Soares et al., 2010). The complete replacement of the hunter-gatherer lifestyle in Central Europe is indicated by the homogenous appearance of the Neolithic Package across LBK sites, which are distributed from the Hungarian plain to Alsace in France (Bogucki, 1996). Throughout the LBK's distribution, there is a consistent presence of agriculture associated elements, including ceramics in the linear pottery style, settlements in the form of longhouses and Near Eastern domesticates, including wheat, barley and cattle (Bogucki, 1996, Price et al., 2001). In Central Europe, the complete and quick establishment of agriculture in just over 400 years may have been assisted by the small number of hunter-gatherer populations, whose numbers were decreasing in Central Europe during the Mesolithic (11, 500 – 5, 300 years before present) (Gronenborn, 1999, Soares et al., 2010). However, vestiges of the hunter-gatherer lifestyle remained after the introduction of agriculture, as indicated by the presence of hunter-gatherer ceramic styles, such as Limburg and La Hoguette, among LBK sites in Western-Central Europe (Gronenborn, 1999). Compared to Central Europe, there was a greater prevalence of hunter-gatherers in Northern Europe, which encompasses the modern-day countries of Denmark, Finland, Norway and Sweden (Nygaard, 1989, Price, 1991). The successfully utilisation of marine resources by Northern European hunter-fisher-gatherers (Richards et al., 2003a) made it difficult for the establishment of agriculture, and resulted in the co-occurrence of hunter-gatherer and agriculturist lifestyles during the Neolithic (Nygaard, 1989, Price, 1991). Near Eastern domesticate plants, such as wheat and barley were found in association with the earliest agricultural group in Northern Europe (Price, 1991), the Funnel Beaker culture (3900 – 3300 BC) (Richards et al., 2003a), which is thought to have descended from the LBK. This culture co-existed for approximately 500 years with the Pitted Ware Culture, which were a Northern European, Mesolithic hunter-gatherer population (Nygaard, 1989, Fornander et al., 2008). Pitted Ware culture individuals appear to have obtained much of their food through fishing, as indicated by the presence of seal, dolphin and whale bones at Pitted Ware sites (Nygaard, 1989, Fornander et al., 2008). The hunter-gatherer, Pitted Ware Culture eventually disappeared approximately 4,000 years ago, and the Funnel Beaker culture survived, with agricultural populations then dominating Northern Europe (Malmstrom et al., 2009).

While there is negligible evidence that agriculture developed in Northern or Central Europe prior to the arrival of Near Eastern domesticates during the Neolithic, in contrast, it has been suggested that domesticated local animals formed part of the hunter-gatherer lifestyle in Western Europe, and were later incorporated into a Near-Eastern derived, agricultural lifestyle. The local domestication of animals in Western Europe has been indicated by the recovery of goat skeletal remains from Italy that are smaller than the wild variety, dated to approximately 8, 000 BC (Price, 1987, Zvelebil, 1989), which predates the arrival Near Eastern domesticates in approximately 6, 000 BC (Bogucki, 1996). However, the small goats may have been morphologically misidentified and are potentially juvenile chamois or ibex (Zilhao, 2001). Determining the interaction between hunter-gatherers and agriculturists along the Mediterranean coast is further confounded by the difficulty in dating sites which are often disturbed (located in caves and rock shelters) or submerged (Zeder, 2008). For example, dates which support the hypothesis that animal domestication was occurring in Spain by 8 BC were obtained from disturbed sites and could potentially reflect the overlying Neolithic context (Zilhão, 1993). To determine whether the development of the Neolithic Package along the Westward route was due to the adoption of Near Eastern elements or due to the independent development of agriculture in-situ, this would require the re-analysis of animal morphology and a focus on the dating of undisturbed sites across the Mediterranean coast.

Two opposing hypotheses (demic diffusion and cultural diffusion) have been proposed to explain the patchwork of modes by which the Neolithic Package was adopted across Europe, as inferred from the archaeological record. The demic diffusion hypothesis posits that agricultural practices were introduced to Europe by populations of Near Eastern origin, which replaced the indigenous hunter-gatherers (Childe, 1925, Clark, 1965, Ammerman and Cavalli-Sforza, 1984). According to the demic diffusion hypothesis, the first farmers of Europe would have originated from populations that had been practising agriculture for generations, explaining why the spread of farming in areas such as Central Europe was homogenous and quick (Gronenborn, 1999). In contrast, the gradual transition from a hunter-gatherer to farming lifestyle, as seen in Northern and potentially Western Europe, has been used to support the cultural diffusion hypothesis (Price, 1987, Nygaard, 1989, Zvelebil, 1989, Price, 1991). This model advocates that farming practices were developed by indigenous hunter-gatherers, who either obtained Near-Eastern domesticates by trading with agricultural groups or independently developed plant and animal domestication (Zvelebil, 1986, Zvelebil and Zvelebil, 1988). The degree to which the Neolithic Revolution has contributed to the settlement history of Europe is dependent on the mode by which farming was adopted in Europe. For example, if farming practices spread to Europe through the transmission of ideas

as opposed to people, then the arrival of agriculture would have had little impact on Europe's settlement history. According to this scenario, the modern population structure of Europe would be descended from indigenous European populations and influenced by events which occurred in Europe during the Pleistocene and the later Neolithic (Soares et al., 2010).

Each major event in Europe's history is likely to have left a signature in the genetic diversity of modern and prehistoric European populations. Genetic analysis of human (Haak et al., 2005), plant (Salamini et al., 2002) and animal (Larson et al., 2007) DNA have been applied to investigate different forms of gene flow, such as past dispersals, expansions, colonisations and settlement patterns (Soares et al., 2010). In particular, the impact of past events on the genetic structure of populations has been examined using the non-recombining and near-neutral elements of the human genome, the maternally inherited mitochondrial (mt) and paternally inherited Y chromosome (YDNA) (Jorde et al., 2000). Due to the lack of recombination in these uniparental genomes, sequence differentiation is only caused by the accumulation of mutations, and mutations which share a common ancestor have been termed haplogroups (Wallace, 1995). The time scale for haplogroup divergence is determined by converting the diversity of lineages to an age estimate through the use of a molecular clock (Soares et al., 2010). A molecular clock uses fossil constraints, such as the time of the *Homo-Pan* split and rates of molecular change, also called the mutation rate, to deduce the age when a lineage diverged (Ho and Larson, 2006, Soares et al., 2009). The molecular divergence of mtDNA and YDNA which has occurred among populations has been extensively used to investigate population differentiation (Wallace, 1995, Achilli et al., 2004). The mtDNA and YDNA have a number of features which make them ideal for studying population differentiation, including; 1) uniparental inheritance and 2) reduced effective population size (one-quarter) compared to the autosomal genome, as the mtDNA and YDNA are only inherited from one parent, as opposed to autosomal genes which are maintained on two chromosomes in each parent (Garrigan and Hammer, 2006). The smaller effective population size of these uniparental genomes compared to autosomal genes, results in the mtDNA and YDNA being highly influenced by genetic drift (Garrigan and Hammer, 2006). Of the uniparental genomes, mtDNA in particular has been targeted to investigate ancient DNA (Haak et al., 2005, Larson et al., 2007, Haak et al., 2008, Malmstrom et al., 2009, Haak et al., 2010) because of the high number of copies of the genome within each cell (10^3 – 10^4 copies per cell), compared to single copy nuclear DNA (Shadel and Clayton, 1997).

Genetic analysis of both modern (Torroni et al., 1998, Richards et al., 2000a, Achilli et al., 2004, Rootsi et al., 2004, Achilli et al., 2005, Soares et al., 2010) and ancient human DNA

(Haak et al., 2005, Sampietro et al., 2007, Haak et al., 2008, Bramanti et al., 2009, Malmstrom et al., 2009, Haak et al., 2010) has been used to investigate how agriculture spread to Europe and hence determine the contribution of the Neolithic Revolution to the genetic makeup of Europe. The application of genetic analysis to investigate how agriculture spread to Europe from the Near East is detailed in the following sections.

1.2.2.1 Support for the cultural diffusion hypothesis – Pleistocene settlement events

Proponents of the cultural diffusion hypothesis suggest that the modern European gene pool is essentially autochthonous (i.e. indigenous to the continent), reflecting settlement events which occurred during the Pleistocene. The Pleistocene settlement events which are thought to have contributed to the genetic makeup of Europe include; the original peopling of Europe by anatomically modern humans from the Levant, 41-46 thousand years ago (Mellars, 2006), and the repeopling of Europe following the Last Glacial Maxima (LGM), 19-22 thousand years before present. During the LGM the climate become colder and dryer (Yokoyama et al., 2000, Clark et al., 2004) and resulted in forest areas in Europe retreating to scattered refugia where humans concentrated in the Southern European Peninsula (Franco Cantabrian basin/Iberian Peninsula and the Alps) and the Western Caucasus (Adams and Faure, 1997, Peyron et al., 1998, Forster, 2004). Humans concentrated in these refugia during the LGM because the forests provided a source of plant food, and also attracted animals, such as badger (Sommer and Benecke, 2004), brown bear (Sommer and Benecke, 2005b) and red fox (Sommer and Benecke, 2005a), which could be hunted.

Genetic studies of modern European and Near Eastern populations have indicated that Pleistocene events have greatly contributed to the modern European gene pool, as the majority of common European mtDNA and Y chromosome haplogroups originated during this period (Torroni et al., 1998, Achilli et al., 2004, Rootsi et al., 2004, Achilli et al., 2005, Soares et al., 2010), which are detailed below. However, estimating the age of haplogroup divergence and equating the age when a lineage diverges to the time of population settlement, may overestimate the time of population settlement (Barbujani et al., 1998), such as in the scenario where a population which is carrying a particular haplogroup moves after the time that that the haplogroup originated. Furthermore, the use of single haplogroups to determine population movement is an oversimplification of human demography, and analysing variation in the overall genetic diversity of populations may provide a more realistic picture of Europe's settlement history. Analysis of genetic diversity in modern European populations has also indicated that a substantial proportion of genetic diversity in Europe today is thought to be indigenous to the continent as opposed to from the Near East, and attributed to

Pleistocene events (Richards et al., 2000a, Semino et al., 2000), as will be discussed in the following sections.

The candidate lineages for the initial peopling of Europe include mtDNA haplogroup U (Achilli et al., 2005, Soares et al., 2010) and YDNA haplogroup I (Rootsi et al., 2004) because these lineages have been dated to have originated between the original peopling of Europe and the LGM from molecular dating, and they are thought to have arisen in Europe because they have distributions which are restricted to the continent. Haplogroup U is the oldest mtDNA lineage within Europe, dated by analysis of whole mtDNA genomes to have originated between 56 (Soares et al., 2010) and 58 (Achilli et al., 2005) thousand years before present. Although haplogroup U is the oldest mtDNA lineage within Europe, it is not indigenous to the continent, possessing a broad geographic distribution that ranges from Europe to India and Central Asia (Achilli et al., 2005). However, sub-haplogroups of U, U5 and U8, are thought to reflect the initial peopling of Europe, as both are found almost exclusively in Europe and are estimated to have originated on the continent before the LGM from molecular dating (Thomas et al., 2002, Achilli et al., 2005, Soares et al., 2010). Analysis of whole mtDNA genomes from a wide range of European populations has suggested that U5 originated between 37 (Soares et al., 2010) and 41 (Achilli et al., 2005) thousand years before present, and U8 is estimated to have originated 50 thousand years ago (Soares et al., 2010). Akin to mtDNA lineages U5 and U8, the Y chromosome haplogroup I is also thought to have originated in Europe pre-LGM (Soares et al., 2010). Haplogroup I is widespread in Europe, accounting for approximately 18% of male lineages, and is virtually absent elsewhere in the world (Rootsi et al., 2004). The suggestion that haplogroup I is a marker of the Pleistocene settlement in Europe is further supported by the peak in frequency of haplgroup I amongst populations where hunter-gatherers are known to have survived longest, such as in Northern Europe (Semino et al., 2000). A peak in frequency of a haplogroup refers to the population or area where the haplogroup is most frequent. The Y chromosome haplogroup I is thought to have originated 24 thousand years before present, as estimated from the analysis of short tandem repeats (STR) in the Y chromosome from over 1000 males, sampled across Europe (Rootsi et al., 2004).

In Europe, the next major settlement event after the initial colonisation was the post-LGM repeopling of the continent, primarily from the Southern European refugia (Adams and Faure, 1997, Peyron et al., 1998, Forster, 2004). The post glacial colonisation of Europe is thought to have substantially influenced the modern the structure of the European gene pool, as common European haplogroups, such as haplogroup H, originated during this period, as determined

from molecular dating (Achilli et al., 2004, Soares et al., 2010). Haplogroup H is the most frequent mtDNA lineage in Europe today, and accounts for on average 45% of Europeans (Achilli et al., 2004, Loogvali et al., 2004, Soares et al., 2010). During the LGM (18 thousand years before present), haplogroup H is proposed to have originated in the Near East from the molecular dating of 54 complete mitochondrial genomes which were retrieved from modern European populations (Achilli et al., 2004). Sub-haplogroups, H1 and H3, are potential signals of the post-LGM colonisation of Europe, as their frequency distribution indicates population movement from the West to East of Europe (Achilli et al., 2004, Loogvali et al., 2004, Soares et al., 2010). Both H1 and H3 are estimated to have diverged after the LGM, 13 and 10 thousand years before present from molecular dating of 134 and 50 complete mitochondrial genomes, respectively that were sequenced from modern European populations (Achilli et al., 2004). These sub-haplogroups display a frequency cline, peaking at ~65% in areas around the Southern European refugia, and decreasing to ~15% in the North and East of Europe (Achilli et al., 2004, Roostalu et al., 2007). The declining frequency gradient from the West to East of Europe observed in H1 and H3 is also observed in mtDNA haplogroup V (Torroni et al., 1998, Torroni et al., 2001). Haplogroup V is an autochthonous European haplogroup, which is estimated to have originated post-LGM, 11 thousand years ago, from molecular dating of the mtDNA hypervariable region 1 (HVS1) from 205 modern European individuals (Torroni et al., 1998, Torroni et al., 2001). As with the mtDNA haplogroups, the Y haplogroups implicated in the post-glacial colonisation of Europe form the majority of lineages in the continent today. It has been suggested that sub-clades of haplogroup I (I1 and I2b1) have expanded from the Southern European refugia, and are estimated to have diverged post-LGM (16 thousand years ago) from molecular dating of STR variation in 7, 574 European males, displaying a frequency peak in the Iberian Peninsula (Rootsi et al., 2004). In addition to haplogroup I, the Y chromosome R1 lineage has been implicated in the repeopling of Europe post-LGM (Semino et al., 2000). There are two major clades of R1 in Europe, R1b in the west and R1a in the north-east (Semino et al., 2000). It has been suggested that R1b mirrors mtDNA sub-haplogroups H1 and H3 (Soares et al., 2010), displaying similar frequency clines, with a peak in Western Europe at ~70% (Myres et al., 2011) and a decrease in frequency across Northern and Eastern Europe (Rosser et al., 2000, Semino et al., 2000).

The description of Europe's settlement history reviewed above has focused on estimating the age of haplogroup divergence and equating the age when a lineage diverges to the time of population settlement (Torroni et al., 1998, Achilli et al., 2004, Rootsi et al., 2004, Achilli et al., 2005, Soares et al., 2010). However, as previously mentioned, the dating of haplogroup origins may overestimate the time of population settlement (Barbujani et al., 1998). A

different approach to studying past population events is the application of founder analysis to estimate the contribution of Pleistocene and Neolithic events to the overall genetic diversity of the modern European gene pool (Richards et al., 2000a). Founder analysis compares the genetic diversity in a sink or settled population, in this case Europe, to a likely source population, which is the Near East in this scenario. Genetic diversity is compared between the sink and source populations by searching for founder sequences in the sink population that are identical to, or close derivatives of, sequences in the source population (Richards et al., 2000a). Founder analysis assumes that the past European and Near Eastern population structure is reflected in the modern composition. The application of founder analysis to mtDNA sequences from modern European and Near Eastern populations has repeatedly found that the majority of European maternal genetic diversity is autochthonous (Richards et al., 1996, Torroni et al., 1998, Richards et al., 2000a). Indigenous maternal genetic diversity accounts for ~80% of the European gene pool and has been suggested to reflect Pleistocene events (Richards et al., 1996, Torroni et al., 1998, Richards et al., 2000a). The remaining proportion of maternal genetic diversity in Europe (~20%) is attributed to the settlement of agriculture by individuals from the Near East at the onset of the Neolithic (Richards et al., 1996, Torroni et al., 1998, Richards et al., 2000a). The pattern observed in Europe's maternal gene pool, has also been observed in the paternal gene pool (Semino et al., 2000). By analysing Y chromosome STR markers that are common in Europe and the Near East, the Near Eastern contribution to the paternal gene pool was estimated to be 22%, with the remaining paternal genetic diversity found to be indigenous to Europe and attributed to Pleistocene events (Semino et al., 2000).

Although founder analysis enabled the investigation of settlement history by using genetic diversity and moved away from using only initial haplogroup divergence dates to estimate when past population events occurred, there are problems with founder analysis, in particular relating to the statistical validity of the method and the ability of the analysis to correctly identify population changes. Founder analysis may underrepresent the similarity between sink and source populations. For example, if the source population is not sequenced to an adequate depth to capture the majority of genetic diversity, this may result in failure to identify 'true' founder sequences within the sink population. However, in the studies investigating Europe's settlement history by founder analysis, as the number of individuals in the source population has been increased, from 64 (Richards et al., 1996) to over 1000 (Richards et al., 2000a), the proportion of founder sequences has not increased, which indicates the analysis is not being biased by sample size. A more substantial issue is the statistical or phylogeographic approach taken in founder analysis to investigate settlement history. Due to the statistical limitations of

founder analysis, this method is unable to model certain aspects of past population change and hence requires unrealistic demographic assumptions to be made. For example, in founder analysis each founder sequence can only represent a single migration (Richards et al., 2000a). This assumption is probably unrealistic given the settlement history of Europe, where haplogroups have probably been involved in multiple population movements (Achilli et al., 2004). Furthermore, the analysis assumes a constant population size (Richards et al., 2000a) even though there is strong archaeological evidence to suggest population sizes in Europe exponentially increased following the introduction of agriculture (Gronenborn, 1999, Shennan and Edinborough, 2007). The highlighted issues associated with founder analysis can be addressed by analysing the data within a coalescent population genetic framework. This enables varying population size to be modelled, encompasses the impact of genetic drift and allows sequences to be involved in more than one event (Rosenberg and Nordborg, 2002). A further issue with founder analysis, which is inherent to modern genetic studies, is the assumption that all founder lineages of a settlement event survived until today (Richards et al., 2000a). Sequences involved in past events may be completely erased if a comprehensive population replacement occurs subsequently. To determine the past state of population structure directly, genetic material can be recovered from archaeological human remains.

1.2.2.2 Support for the demic diffusion hypothesis – Neolithic settlement events

The two main hypotheses proposed to explain how agriculture arose in Europe, cultural and demic diffusion, while not mutually exclusive, do have different consequences for Europe's settlement history (Menozzi et al., 1978). Ammerman and Cavalli-Sforza (1973, 1984) were the first to theorise about how the introduction of farming would have influenced the population genetics of Europe under the demic diffusion hypothesis. They described the spread of agriculture into Europe as a 'wave of advance' of Near Eastern origin farmers (Ammerman and Cavalli-Sforza, 1984). Population growth at the periphery of the farmers' range was predicted to have resulted in the steady expansion of agricultural populations in all directions across Europe (Ammerman and Cavalli-Sforza, 1984). The arrival of agriculture as described by the 'wave of advance' model was predicted to have resulted in a gene frequency gradient across Europe (Ammerman and Cavalli-Sforza, 1984). Gene frequencies were expected to peak in the Near East where agriculture developed and decrease across Europe due to the admixture of farmers at the periphery of the wave with indigenous Mesolithic populations (Ammerman and Cavalli-Sforza, 1973, Ammerman and Cavalli-Sforza, 1984).

The forecast presence of a gene gradient in Europe was first confirmed by the principal components analysis (PCA) of 38 independent alleles in a broad range of Near Eastern and

European populations (Menozzi et al., 1978). These alleles consisted of classical genetic markers, such as the histocompatibility locus (HLA-A and HLA-B), blood type locus (alleles A, B_1 and O), Rhesus factor locus and the locus for the MN blood group factors (Menozzi et al., 1978). A multivariate statistic, such as PCA, was used to examine the frequency of these alleles in multiple populations because PCA enables numerous variables to be condensed into a few synthetic components, each of which represents a degree of variation within the data (Jombart et al., 2009). To investigate whether a gene gradient was present, the principal component scores were superimposed on the geographical map of Europe (Menozzi et al., 1978, Cavalli-Sforza et al., 1993, Cavalli-Sforza et al., 1994). This routine was performed for the principal component scores from each of the first three components. The map of the first principal component, which accounted for ~30% of variation in the data, revealed a genetic gradient that peaked in the South East of Europe and declined towards the North West (Menozzi et al., 1978, Cavalli-Sforza et al., 1993, Cavalli-Sforza et al., 1994). The peak of the observed gene cline in the South East of Europe was inferred to reflect the area where agriculture arose on the continent. The direction of the gene cline towards the North West was taken to indicate the expansion route of the Neolithic farmers from the South East of Europe.

However, a gene cline does not in itself indicate when a population expansion occurred. The observed gene cline was linked to the arrival of farming in Europe by demonstrating a strong correlation ($p < 10^{-5}$) between genetic distances, calculated from 59 classical genetic markers which were examined in European and Near Eastern populations, and archaeological time distances, estimated from Mesolithic and early Neolithic sites (Sokal et al., 1991). Analyses of genetic markers other than the classical loci described above have also revealed a South East to North West gene cline in Europe. These include PCAs of 309, 790 genotype single-nucleotide polymorphisms (SNPs) which were examined in 2, 457 Europeans (Lao et al., 2008) and 22 SNPs from the non-recombining portion of the Y-chromosome that were analysed in 1, 007 Europeans (Chikhi et al., 2002).

Support for the demic diffusion hypothesis has been derived from the observation of gene frequency gradients in PCA (Menozzi et al., 1978, Cavalli-Sforza et al., 1993, Cavalli-Sforza et al., 1994, Chikhi et al., 2002, Lao et al., 2008), which have been interpreted as evidence for historical migrations providing support for the demic diffusion hypothesis. However, the presence of gene clines in PCA has been found to be a mathematical artefact of analysing spatially distributed data (Novembre and Stephens, 2008). This was revealed by the production of gene clines, typical of those observed in Europe, from the PCA of genetic data that was simulated under equilibrium population genetic models, without range expansions

and assuming the occurrence of homogenous, short-range migrations across time and space. Novembre and Stephens (2008) found that when PCA was applied to spatially distributed data where the covariance (similarity) between locations tended to decline with geographic distance, it produced a highly structured covariance matrix. In this situation, the eigenvectors, which are derived from the covariance matrix and produce the patterns in the PCA plot, become related to sinusoidal waves of increasing frequency. Hence, the principal components follow an S-wave, with the first component being an orthogonal gradient, and the second and third components being saddle or mound shapes. Therefore, the gene clines observed in Europe are the same as the null pattern observed in the PCA of spatially distributed data from a non-expanding population. In this study, Novembre and Stephens (2008) did not test whether the null pattern could be produced from the application of PCA to genetic data which contained a population expansion signal. PCA of genetic data that was simulated under the demic diffusion hypothesis, with a South East to North West population expansion, surprisingly revealed a gene cline that was perpendicular to the modelled expansion, running South West to North East (Francois et al., 2010). Based on these results, Francois and colleagues (2010) suggested that the gene cline observed in Europe, the South East to North West gradient, may be reflective of either the null pattern or a population expansion which ran perpendicular to the gradient, such as the post-LGM colonisation of Europe. Hence, the results of PCA appear to be open to multiple interpretations when being used to detect past population expansions from geographically distributed data (Novembre and Stephens, 2008), even if a population expansion signal is present in the data (Francois et al., 2010).

A more favourable approach for investigating population history is the use of model based methods, such as Serial SimCoal. Coalescent based analyses simulate a model of population history, the likelihood of which is assessed by comparing the simulated and observed genetic data (Burnham and Anderson, 2002, Anderson et al., 2005, Chan et al., 2006, Ramakrishnan and Hadly, 2009). Genetic data is typically simulated under the coalescent theory (Burnham and Anderson, 2002), which is a retrospective model of population genetics in which offspring randomly choose their parents, and mutational events are non-selective, governed only by genetic drift (Kingman, 1982). The simulated data is compared to the observed genetic information through the use of summary statistics (Anderson et al., 2005, Chan et al., 2006, Ramakrishnan and Hadly, 2009). The type of summary statistic used is dependent on the aspect of population history of interest. For example, to investigate a single populations' history, recommended intrapopulation statistics include average pairwise differences and number of segregating sites (DNA sites that are polymorphic) (Ramakrishnan and Hadly, 2009). Both these statistics reflect the degree to which a population has evolved 'neutrally',

and whether deviations from neutrality have occurred due to events such as population expansion or contraction (Tajima, 1989, Slatkin and Hudson, 1991). To investigate the amount of population divergence between subpopulations, interpopulation statistics are applied, such as the fixation index or $F_{ST.}$ The F_{ST} is a genetic distance measure which is based on the pairwise comparison of sequences from subpopulations of a whole population (Wright, 1951, Slatkin and Hudson, 1991). Estimates of F_{ST} can be expressed in terms of the average divergence between pairs of sequences within subpopulations and the average divergence between pairs of sequences randomly drawn from the whole population (Slatkin and Hudson, 1991). The F_{ST} value increases as the sequence divergence between subpopulations becomes larger than the divergence between randomly drawn sequences from the whole population. The fixation index has been the primary summary statistic used for model comparison analyses of the demic and cultural diffusion hypotheses (Bramanti et al., 2009, Malmstrom et al., 2009, Haak et al., 2010).

The findings from coalescent analysis of ancient mtDNA from early Neolithic farmers and post-LGM hunter-gatherer remains has lent support to the development of farming occurring in accordance with the demic diffusion hypothesis in Central (Bramanti et al., 2009) and Northern Europe (Malmstrom et al., 2009) by showing a lack of genetic continuity between hunter-gatherers and modern European populations. These two studies used very similar approaches to test the likelihood of the demic diffusion hypothesis, and hence will be discussed together. Both studies compared a simulated model to summary statistics generated from mtDNA from the hypervariable region 1 (HVS1) that was retrieved from ancient hunter-gatherer and agriculturist populations, and modern populations from the same regions. The study by Bramanti and colleagues (2009) used ancient mtDNA, extracted from post-LGM (23, 000 – 13, 000 BC), hunter-gatherer remains from Central and Northern Europe (n=22), and compared this to previously published ancient DNA sequences the LBK (5450 – 4775 BC, n=20) (Haak et al., 2005). Malmstrom and co-workers (2009) used mtDNA retrieved from the Mesolithic, hunter-gatherer group, the Pitted Ware Culture (4,800–4,000 years BP, n=19) and compared this to ancient DNA recovered from an agricultural population the Funnel Beaker Culture (5,500–4,500 years BP, n=3). The validity of the ancient human DNA sequences was supported in both studies by the adherence to authentication criteria. These included the independent replication of results per individual by using separate skeletal samples for DNA extraction, confirmation of appropriate molecular behaviour for degraded material (inverse relationship between fragment length and quantity) through the analysis of DNA quantity using real time (q)PCR and monitoring contamination through non-template controls and either cloning of sequences (Bramanti et al., 2009) or in-depth sequencing of

samples by second generation sequencing analysis (Malmstrom et al., 2009). The hunter-gatherer sequences from both Central (Bramanti et al., 2009) and Northern Europe (Malmstrom et al., 2009) were predominantly (on average 81%) classified within haplogroup U. The hapolgroup makeup of the agricultural populations were notably different to the hunter-gatherers, with haplogroups N1a, H, HV, J, K, T, V and U3 found in Central Europe (Haak et al., 2005, Bramanti et al., 2009) and H, J and T identified for the Northern European remains (Malmstrom et al., 2009). Based on pairwise $F_{ST's}$, the genetic makeup of the hunter-gatherers in Central ($p < 10^{-6}$) (Bramanti et al., 2009) and Northern Europe ($p < 0.005$) (Malmstrom et al., 2009) was found to be significantly different to modern populations from these two areas. These results demonstrate a significant level of population differentiation between the hunter-gatherer and modern European populations, and hence did not support the cultural diffusion hypothesis.

To determine if these findings could represent genetic drift over time, the observed F_{ST} values were compared to $F_{ST's}$ calculated from genetic data that was simulated under a model of population continuity (Bramanti et al., 2009, Malmstrom et al., 2009). In this model, which is essentially simulating the cultural diffusion hypothesis, population continuity extended from the arrival of humans in Europe during the Pleistocene (45, 000 years ago), through the development of agriculture in the Neolithic (7, 500 years ago), until today. This model took into account two periods of population growth; the first coincided with the arrival of humans in Europe, and the second occurred after the introduction of agriculture. For each simulation, pairwise F_{ST} values were calculated between each combination of hunter-gatherers, agriculturists and modern populations. The simulations revealed that for the population continuity model, nearly all the pairwise F_{ST} values calculated between hunter-gatherers and modern day populations were smaller than the F_{ST} estimates from the observed data. This indicates that if the model of the cultural diffusion hypothesis was true, we would expect to observe smaller F_{ST} values when comparing hunter-gatherers and current day Europeans, and hence less population differentiation. Based on these results, both studies claimed to reject the null model and hence support the demic diffusion hypothesis.

The manner by which Bramanti et al., (2009) and Malmstrom et al., (2009) used a model based approach to investigate the demic diffusion hypothesis restricted the full potential of the analyses by only testing one model of population history and not assessing the goodness of fit of the model. The power of model-based analyses is derived from the ability to compare alternative population histories and assess the most likely model, given the observed genetic data. In both Bramanti's et al., (2009) and Malmstrom's et al., (2009) studies, only the

cultural diffusion hypothesis was modelled, and hence the likelihood of the demic diffusion hypothesis was not directly determined. Furthermore, these studies essentially reduced model testing to a statistic, such as a t-test, whereby a distribution of possible values is created from the simulations and compared to the observed data. This type of comparison provides no information about the likelihood of the model occurring given the observed data. To determine the maximum likelihood of a model, the simulated summary statistics need to be compared to the observed statistics through the use of rejection algorithms, such as in Approximate Bayesian Computations (ABC). ABC's are used to determine the posterior probability of parameters in the model, such as effective population size, migrations and expansions (Beaumont et al., 2002). The likelihood of different models, given the observed data, can then be compared through the use of tests such as Akaike's Information Criterion (AIC) (Burnham and Anderson, 2002). An AIC value for each model is determined by both the probability of the model occurring given the observed data and the complexity of the model. Complex models (i.e. with more parameters) are penalised in AIC because they tend to fit real life data better than simpler models. Models that are most likely have the lowest AIC scores, and when comparing models a difference of two AIC values has been deemed by the field to reflect a significant differences in the goodness of fit of the model to the observed data (Burnham and Anderson, 2002).

Proper model comparison of the demic and cultural diffusion hypotheses, was performed on the same ancient mtDNA obtained from Central Europe as described above (Haak et al., 2005, Bramanti et al., 2009), with an increased amount of sequences from the LBK (n=42) by Haak and colleagues (2010). Also, to assess the geographic origin of the first farmers in Europe, the ancient sequences were compared to not only modern Central Europeans, but also Near Eastern populations (Haak et al., 2010). The new ancient genetic data was subject to the same authentication criteria as described above, with the addition of the multiplex typing 22 mtDNA haplotype defining SNPs from the coding region. The coding region SNP multiplex amplicon products for the non-template controls were sequenced, which provided an ideal monitoring system for contamination of the ancient samples and non-template controls, as the PCR multiplex directly targets SNPs that define potentially contaminating lineages. This study showed that based on F_{ST} estimates, the significant population differentiation ($p < 0.05$) observed between hunter-gatherers and both Neolithic farmers and modern Central Europeans was best explained by the model of population history which was simulated according to the demic diffusion hypothesis. This hypothesis was modelled by simulating a discontinuous population between hunter-gatherers and modern day Europeans due to the influx of migrants from a Near Eastern deme during the early Neolithic, when farming was introduced to the

continent. The level of migration from the Near East to Europe was modelled between 25-75%. The most likely models, given the observed genetic data, were when migration from the Near East to Europe was high (50-75%).

The model based analyses indicated that the LBK shared genetic similarities with Near Eastern populations and this finding was confirmed by phylogeographic analyses (Haak et al., 2010). The geographic origin of the first farmers was investigated by calculating the genetic distances between the LBK data to modern European and Near Eastern populations. These genetic distances were superimposed on the map of Europe and the Near East. Mapping genetic distances, enabled the identification of populations which had the shortest genetic distance to the ancient agriculturists, and hence were most similar in mtDNA composition. Based on genetic distances, the LBK were more similar to Near Eastern populations, including modern day Turkey, Armenia, Iraq and Iran. While the affinity of the first framers in Europe to modern Near Eastern populations is expected under the demic diffusion hypothesis, it was also expected that the LBK would be a continuous population with modern day Central Europeans. However, based on F_{ST} comparisons there was significant population differentiation ($p < 0.05$) between the early Neolithic farmers and modern day Central Europeans. The lack of mitochondrial genetic continuity between the LBK and modern day Central Europeans may indicate that the first farmers in Central Europe were not fully incorporated into the succeeding populations (Soares et al., 2010). Therefore, demographic processes that occurred after the early Neolithic may have substantially contributed to shaping the genetic makeup of contemporary Central Europeans (Haak et al., 2010).

Population movements that occurred after the early Neolithic and are postulated to have contributed to the genetic makeup of Central Europe include internal migrations within the continent during the Late Neolithic until the early Bronze Age (4000-2000 BC (Anthony, 1990, Soares et al., 2010)). The influence of population events during the Late Neolithic/Early Bronze Age on the genetic makeup of contemporary Europeans has only been investigated by a few studies (Richards et al., 2000a, Haak et al., 2008). Founder analysis of modern mtDNA has revealed that population events during the Bronze Age contributed minimally (2-5%) to the genetic makeup of contemporary Central Europeans (Richards et al., 2000a). However, as discussed previously (Section 1.2.2.1), there are major flaws with the use of founder analysis to interpret population history. The occurrence of population change during the Copper Age has been suggested by the presence of genetic differences in ancient mtDNA retrieved from the early Neolithic culture, the LBK (Haak et al., 2005) and late Neolithic group, the Corded Ware Culture (Haak et al., 2008). In both studies, remains from the LBK and Corded Ware

Culture were recovered from Central Europe. The most dominant haplogroup in the LBK, N1a (25%), was not observed in the Corded Ware Culture. Furthermore, the Corded Ware remains contained haplogroups not found in the LBK, including mtDNA lineages I and X. However, there were mtDNA haplogroups found in both cultures, including H, K and U. Whether these genetic differences represent significant population differentiation between early and late Neolithic cultures in Central Europe is unknown because quantitative analysis to compare the data has not been performed. To determine the degree of population change between the early Neolithic and late Neolithic/early Bronze Age, and hence the influence of this period on the genetic makeup of modern day Central Europeans, it is necessary to analyse ancient DNA from a range of temporally spaced Late Neolithic/Early Bronze Age samples in a model based approach. The analysis of data in this manner would enable the likelihood of population continuity throughout the Neolithic to be examined and to further resolve the influence of the first farmers on current European populations. This approach is applied to the analysis of ancient mtDNA recovered from Central European, late Neolithic/early Bronze Age samples in Chapter 3.

1.2.3 Dietary changes associated with the introduction of agriculture

The introduction of agriculture to Europe, aside from being associated with a population expansion from the Near East (Bramanti et al., 2009, Haak et al., 2010), was also associated with a major change in diet. During the Neolithic, populations in Europe began to obtain their food from domesticated plants and animals that had originated in the Near East (Braidwood et al., 1961), as suggested by the recovery of domesticate animal and plants remains of Near Eastern origin from Neolithic sites across Europe (Zilhão, 1993, Bogucki, 1996, Price et al., 2001). The use of plants among European, Neolithic agriculturists is also indicated by the increasing numbers of plant storage facilities and grinding tools at agricultural sites (Stiner, 2001). While there is debate over whether domesticated animals formed part of the hunter-gatherer assemblage in Western Europe (See Section 1.2.2, (Price, 1987, Zvelebil, 1989)), in general there is an absence of domesticate plant or animal remains recovered from European, hunter-gatherer sites. The faunal assemblages from European Pleistocene sites reveal humans were hunting small to medium sized ungulates, ignoring small mammals, birds, fish, tortoises, and crabs until the terminal Pleistocene/early Mesolithic (Stiner et al., 1999). The skeletal remains in the faunal assemblage are linked to human feeding through the presence of fire damage and/or tool marks, such as percussion fractures from stone hammers (Stiner, 1993). However, the observation of changes in the faunal record between the Mesolithic and

Neolithic in Europe does not directly demonstrate that changes in diet occurred due to the development of agriculture.

Dietary reconstruction of hunter-gatherers and early agriculturists has primarily been achieved through the analysis of stable isotopes and dental morphology. The isotopic composition of an organism can be used to estimate diet because the consumed isotopes become incorporated into the tissues of the consumer (Koch et al., 1994). Dietary reconstruction has commonly been estimated from the analysis of stable isotopes of carbon (the ratio of $^{13}C/^{14}C$ or $\delta^{13}C$ value) and nitrogen (the ratio of $^{14}N/^{15}N$ or $\delta^{15}N$ value) in tissues, such as bone and hair, which reflect the $\delta^{13}C$ and $\delta^{15}N$ values of dietary protein (Tauber, 1981, Macko et al., 1999b, Lillie and Richards, 2000, Richards, 2002). Analysis of carbon and nitrogen isotopes from tissues, such as bone, provide a long-term record of diet, giving the average $\delta^{13}C$ and $\delta^{15}N$ values of all of the protein consumed over the last years of the measured individual's life (~10 years) (Richards, 2002). Isotope values from archaeological remains can be used to estimate diet by comparison with known isotopic values of plants and animals (Tauber, 1981). The value of $\delta^{13}C$ can distinguish the dietary intake of C_3 and C_4 protein sources, which use different biosynthetic pathways to convert $C0_2$ into plant carbon (Macko et al., 1999a). The C_4 plants have $\delta^{13}C$ compositions averaging -12.5%, whereas C_3 plants have average $\delta^{13}C$ values of -26.5% (Koch et al., 1994, Schoeninger, 1995, Richards et al., 2003b). The $\delta^{13}C$ of terrestrial animals tend to reflect the types of plants being consumed (i.e. C_3 or C_4) (Tauber, 1981). The consumption of cereal crops by farmed animals has been shown to produce $\delta^{13}C$ values of approximately -20% (Richards et al., 2003b). In comparison to terrestrial plants and animals, marine plants tend to produce higher $\delta^{13}C$ values (on average -12%) (Richards et al., 2003b). The difference in $\delta^{13}C$ values between terrestrial and marine plants is due to the different sources of carbon used, bicarbonate by marine plants and atmospheric CO_2 by terrestrial plants (Koch et al., 1994). Akin to the $\delta^{13}C$ estimates, the $\delta^{15}N$ value can also be used to determine whether the protein is from a marine or terrestrial source, with an average $\delta^{15}N$ value of 12% for freshwater fish and 4-6% for cattle and red deer (Richards et al., 2000b). Furthermore, the $\delta^{15}N$ value can be used to determine the trophic level of the protein consumed (Richards et al., 2000b), with an approximate enrichment of 3% in $\delta^{15}N$ per trophic level (Macko et al., 1999a). Isotope analysis, while directly revealing the broad source of protein in the diet, is unsuitable for the investigation of more fine scale questions, such as which exact animals or plants formed part of the prehistoric diet.

Stable isotope analysis of European hunter-gatherer and agriculturist skeletal remains has indicated that hunter-gatherers primarily obtained their protein from aquatic based food

sources, whereas farming populations obtained their protein from terrestrial foods (Tauber, 1981, Richards et al., 2001, Richards et al., 2003b). This was indicated by the finding of a marine signature in the $\delta^{13}C$ values (e.g. approximately -12%) of Mesolithic, hunter-gatherer remains which were recovered from coastal regions in both Great Britain (average -16%) (Richards et al., 2003b) and Denmark (average -12%) (Tauber, 1981). Furthermore, the diet of inland hunter-gatherers also appears to have included aquatic foods, which were presumably obtained from freshwater sources (Richards et al., 2001). This was demonstrated by the higher than -20% $\delta^{13}C$ values retrieved from both Mid-Upper Palaeolithic hunter-gatherer remains sampled from Russia, the Czech Republic and Great Britain (Richards et al., 2001), and Mesolithic hunter-gatherer remains recovered from Great Britain (Richards et al., 2003b). A marine signature was also found in the $\delta^{15}N$ values (e.g. approximately 12%) of Mid-Upper Palaeolithic hunter-gatherer remains from inland Europe, which had $\delta^{15}N$ values of between 9-15% (Richards et al., 2001). In contrast, carbon isotope analysis of Neolithic remains has revealed early agriculturists had a terrestrial based diet. This was demonstrated by the finding of a terrestrial signature in the $\delta^{13}C$ values (e.g. -20%) of Neolithic remains from Britain (Richards et al., 2001), in coastal (-20.7%) and inland (-20.8%) settings, and from Denmark (Tauber, 1981) in costal environments (average -25%). However, isotope analysis has revealed that there may have been regional variation across Europe in terms of the impact of agriculture on diet during the Neolithic (Lillie and Richards, 2000). The dietary shift described above, from aquatic to terrestrial, does not appear to have occurred in the Ukraine during the development of agriculture. Carbon and nitrogen isotope analysis of Mesolithic and Neolithic remains from the Dnieper Rapids region of Ukraine revealed a lack of change in the dietary sources of protein across this time period (Lillie and Richards, 2000). In Lillie and Richards study (2000), protein in the diet of Mesolithic and Neolithic individuals was found to be derived from both terrestrial foods, as indicated by $\delta^{13}C$ values ranging from -22% to -24% and freshwater sources, as determined by the finding of $\delta^{15}N$ values between 9-13%.

Dietary changes associated with the introduction of agriculture have also been assessed through tooth wear, also called attrition. The presence and type of macro- and microwear patterns on teeth provide evidence about masticatory behaviour, which is intimately related to diet (Smith, 1984a). Hence, the investigation of attrition patterns in agriculturist and hunter-gatherers has been used to make inferences about prehistoric diet (Larsen, 1995, Schmidt, 2001). Investigations of tooth wear in Pleistocene and Mesolithic hunter-gatherer remains from Europe (Lubell et al., 1994) (Smith, 1984a) and the Near East (Smith, 1984a, Eshed et al., 2006) has revealed heavy amounts of wear, typically in the form of flat occlusal attrition

across the dental arch, where the cusps become obliterated. In comparison, Neolithic agriculturist populations from Europe (Smith, 1984a) and the Near East (Smith, 1984a, Eshed et al., 2006) tend to have less wear, which is concentrated on the molars (Hillson, 1979). Additionally, tooth wear in agriculturist populations is typically angled or cupped, which is seen as scalloping of enamel on the molars, resulting in exposure of the inner tooth tissue, dentine (Smith, 1984a, Eshed et al., 2006). This change in tooth attrition from hunter-gatherers to agriculturists, from heavy to moderate, and flat to angled wear, has been suggested to reflect a reduction in the consistency, abrasiveness or hardness of foods consumed by agricultural populations compared to hunter-gatherers (Smith, 1984a, Larsen, 1995). An agricultural diet has been suggested to be less abrasive than a hunter-gatherer diet due to the cooking of food in water and grinding of grain with sandstone and quartz mortars, and grinding stones (Hillson, 1979, Smith, 1984a). However, the incorporation of grit into the diet due to the grinding of grain has also been attributed to the increased frequency of microwear (pits and scratches) observed in early Neolithic remains recovered from Israel compared to Natufian hunter-gatherer populations from the same area (Mahoney, 2006). In contrast to these findings, the severity of microwear has been found to be greater in Mesolithic hunter-gatherers compared to Neolithic agriculturists in Syria (Molleson and Jones, 1991). The use of tooth wear frequency, severity and pattern to infer diet is confounded by the influence of the age at death of the remains being examined, as tooth attrition tends to increase with an individuals' age (Larsen, 1995). The determination of narrow age estimates from skeletal remains is extremely difficult (Cox and Mays, 2000), except for remains from young individuals (i.e. under 18 years of age) where dental eruption patterns can provide fairly exact age approximations (Ubelaker, 1978). Furthermore, there is a lack of investigations assessing tooth attrition in modern populations (Hillson, 1979) to enable general dietary inferences to be made from tooth wear findings with certainty.

1.2.4 Health changes associated with the introduction of agriculture

Given that humans have spent the majority of their evolution as hunter-gatherers, agriculture represents a very recent ecological change to humans. As hunter-gatherers, the pathogens and parasites harboured by humans are thought to have closely resembled higher apes (Dobson and Carper, 1996). However, since the introduction of agriculture, the pathogen load of humans is thought to have changed and increased (Wolfe et al., 2007), resulting in the declining health of farming populations (Larsen, 1995, Aufderheide et al., 1998). The relationship between an agricultural lifestyle and declining health can be illustrated by the drop in terminal stature (adult height) of European males from on average 179 cms during the

Pleistocene (before 16 000 BC) to 150 cms during the Neolithic (LBK) (Hermanussen, 2003). The average body stature of European males remained between 165 and 170 cm until the end of the 19th century (Hermanussen, 2003). Investigations of stature in modern populations have shown that growth retardation is linked to many environmental factors, including dietary stress (Frisancho et al., 1970). The particular aspects of an agricultural lifestyle which are thought to have contributed to the general decline in health of agriculturist communities compared to hunter-gatherers include a cereal based diet (Tauber, 1981, Richards et al., 2001, Richards et al., 2003b), which was low in nutritional value and diversity (Cordain et al., 2005), increased population density (Shennan and Edinborough, 2007) and close living arrangements with domesticated animals (Taylor et al., 2001, Wolfe et al., 2007). The role that dietary changes and altered living conditions associated with the introduction of agriculture have played in contributing to the worsening health farming populations is discussed in the following sections.

The adoption of a cereal based diet during the Neolithic in Europe (Tauber, 1981, Richards et al., 2001, Stiner, 2001, Richards et al., 2003b), which was nutritionally less diverse (Diamond, 1997, Cordain et al., 2005) than the diet of hunter-gatherers is thought to have resulted in the increased prevalence in the Neolithic skeletal record of oral pathologies (Larsen, 1995, Aufderheide et al., 1998) and signs of malnutrition (Angel, 1966, Goodman, 1989). There is a greater frequency of oral diseases, including caries (decay) and periodontal disease (bone loss caused by inflammation of the gingiva and connective tissues) in European Neolithic remains compared to hunter-gatherers (Aufderheide et al., 1998). When the frequency of caries was compared between five hunter-gatherer groups (Mesolithic and Pleistocene) to five agriculturist groups (Neolithic) that were sampled across Europe and the Near East, there was virtually no caries in the hunter-gatherers and a substantial frequency of the pathology in the farming groups (Smith, 1984b). Caries was present before the introduction of agriculture, however it has only been observed rarely in hunter-gatherers (Smith, 1984b), and early hominid remains, such as *Australopithecus* (Grine et al., 1990). The development of caries and periodontal disease is known to be related to the consumption of carbohydrates, with an increased prevalence of the diseases observed in humans (Sidi and Ashley, 1984) and domesticated animals, such as dogs and cats (Logan, 2006), when they consume an agricultural type diet which is high in carbohydrates.

In addition to contributing to the development of oral diseases, a carbohydrate based diet is thought to have contributed to the increasing prevalence of malnutrition amongst agriculturist compared to hunter-gatherer populations (Larsen, 1995). In particular, the low nutritional

value and diversity of an agricultural type, carbohydrate-based diet compared to the diet of a hunter-gatherer is thought to have contributed to the increased signs of malnutrition amongst agricultural populations (Cordain et al., 2005). Indications that farming communities suffered from nutritional deficiencies more than hunter-gatherers include increased prevalence in agriculturists of cribra orbitalia (Angel, 1966) and enamel hypoplasia (Goodman, 1989). Cribra orbitalia are pitting or lesions in the orbits of the frontal bone that are caused by expanding blood-forming marrow, as the body attempts to increase the amount of blood cells in response to iron-deficiency anaemia (Larsen, 1995). Enamel hypoplasia results in pitting or lines in the enamel, which indicates a disruption to enamel formation occurred in early childhood due to nutritional deficiencies or an infectious disease (Goodman, 1989) . Enamel hypoplasia can also be caused by genetic alterations to the amelogenin gene, however this heritable form of the disease is extremely rare (1 in 10 000) (Barker et al., 2007) and could not account for the prevalence of enamel hypoplasia observed in early farming communities, such the Medieval British population Jewbury, where 25.7% of individuals displayed signs of the pathology (Lilley et al., 1994).

The development of agriculture has also been associated with the development of now widespread infectious diseases such as leprosy, malaria, syphilis and tuberculosis (Taylor et al., 2001, Wolfe et al., 2007). It was only after the adoption of farming that skeletal signs of the above mentioned diseases appeared (Lilley et al., 1994, Meiklejohn and Zvelebil, 1994). Furthermore, Neolithic agricultural populations in Europe during displayed higher frequencies of non-specific skeletal infections, also called periostosis, compared to hunter-gatherers (Lilley et al., 1994, Meiklejohn and Zvelebil, 1994). The increasing prevalence of infectious diseases after the introduction of agriculture has been attributed to an agricultural lifestyle, in particular the close living arrangements with other humans and livestock. The increased population density and close living arrangements with domestic animals is thought to have resulted in larger numbers of potential hosts for infectious disease living in close proximity, enabling pathogens to spread, which were not able to persist in small nomadic groups (Diamond, 1997, Antia et al., 2003, Wolfe et al., 2007). For example, Cholera bacterium does not exist for long periods outside its human host, and spreads via water contaminated with faeces from individuals suffering from the disease (Merrell et al., 2002). The spread of pathogens in farming populations may also have been assisted by the generally unhealthy state of individuals living an agricultural lifestyle, as they tended to suffer from dietary stress, as discussed above. Aside from an agricultural lifestyle resulting in a large number of human hosts for pathogens, the close proximity of humans to large populations of animals is also thought to have contributed to the increased prevalence of infectious diseases.

Modern molecular research has revealed that a number of pathogens in humans, originated in animals (Wolfe et al., 2007), and these infectious agents are termed zoonoses. Examples of zoonoses include measles that originated in cattle (Moss and Griffin, 2006), influenza from ducks and pigs (Pensaert et al., 1981), and smallpox that may have been transferred from cattle or camels (Fenner, 1996). Although the pathogen which causes tuberculosis in humans, *Mycobacterium tuberculosis* (Diamond, 1997, Wolfe et al., 2007), was originally proposed to have originated in cattle, from *Mycobacterium bovis*, comparative genome and phylogenetic analysis has revealed that this was unlikely (Smith et al., 2009). It has even been suggested that human tuberculosis is a 'reverse zoonoses', with *Mycobacterium bovis* representing a reduced version of *Mycobacterium tuberculosis* (Gibbons, 2008).

The increased prevalence of disease amongst agriculturist populations, while well documented from the presence of pathology in skeletal remains, reveals no information about the past diversity of pathogens or how the infectious agents have evolved. As the majority of diseases which are commonplace today, including oral and infectious systemic diseases, are associated with the development of agriculture, and therefore emerged in humans relatively recently, it is crucial we understand how these pathogens originated and evolved. By understanding the past state of pathogens, we can understand how they respond to changing environmental conditions, such as diet and cultural transitions, which may also inform the development of treatments. A potential way to directly study pathogen evolution is the extraction of ancient pathogen DNA from skeletal material or mummified tissue. The application of ancient DNA techniques to recover preserved pathogen genetic material from human remains has made only small advances in the direct investigation of past disease state, as discussed in Section 1.3.3. This PhD proposes to use a new material for the genetic analysis of past diseases, dental calculus, to analyse the changes in human microbiota over time. The properties of dental calculus are described in Section 1.3.3.1.

1.3 Application of ancient DNA to study human evolutionary questions

Ancient DNA has been used to directly study evolutionary processes, such as population replacement, admixture and bottlenecks, which have led to the current state of modern species, such as bison (Shapiro et al., 2004), brown bears (Barnes et al., 2002) and moas (Cooper et al., 1992). The initial ancient DNA studies used bacterial clones to amplify small fragments of DNA extracted from skins of animals, such as the quagga (Higuchi et al., 1984) and Egyptian mummies (Paabo, 1985). The invention of the polymerase chain reaction (PCR)

(Saiki et al., 1985, Mullis and Faloona, 1987) enabled the much more difficult task of amplifying very small quantities of DNA from skeletal remains. The ability to use skeletal remains for genetic analysis was a big step forward, as skeletal remains are much more frequently recovered from the post-mortem environment than mummified tissues (Wood et al., 1992). The first application of PCR to amplifying DNA from skeletal remains was published in 1989, and demonstrated that PCR could be used to retrieve mtDNA from human bones that ranged in age from 300 to 5000 years old (Hagelberg et al., 1989). However, the results from this early study are questionable, as the recovered DNA did not appear to behave as fragmented and degraded DNA would, with extremely long fragments of DNA (600 base pairs) recovered from bone samples that were 750 years old, suggesting the results may have been contaminant modern human DNA. The investigation of ancient human DNA is extremely difficult, due to the high risk of contamination with modern DNA and the inability to then distinguish the ancient from contaminant modern DNA sequences. Hence, the field of ancient DNA has settled upon a set of authentication criteria (Cooper and Poinar, 2000) (discussed below), to ensure (or at least attempt to ensure) that endogenous DNA has been recovered from prehistoric human remains. Many of these authentication criteria have not been applied to the investigation of ancient pathogen DNA from prehistoric human remains (Roberts and Ingham, 2008). The challenges and use of ancient DNA from human remains, for both the analysis of human and pathogen DNA, are discussed in the following sections.

1.3.1 Technical challenges of using ancient DNA

Overall, the amount and integrity of recovered genetic material from preserved remains is a crucial limiting factor in ancient DNA research (Handt et al., 1996, Krings et al., 1997). Due to the lack of DNA repair mechanisms after death, DNA is rapidly degraded into short fragments through hydrolysis, oxidative nucleotide modification, the action of nucleases and other hydrolytic enzymes, in addition to bacterial and fungal attack (Paabo, 1989, Lindahl, 1993, Hoss et al., 1996). These processes can result in modification of the DNA template, such as strandbreaks, baseless sites, miscoding lesions and cross links between strands of DNA, and between DNA and other molecules such as proteins (Lindahl, 1993, Hoss et al., 1996). As DNA becomes fragmented, the number of copies rapidly declines. The low concentration of DNA typically found in archaeological remains, results in ancient DNA being extremely susceptible to contamination with modern DNA, which is comparatively high in concentration (Malmstrom et al., 2005, Malmstrom et al., 2007).

To minimise the impact of modern contamination when dealing with human remains, a number of precautions need to be taken to ensure the authenticity of results (Cooper and Poinar, 2000). These include:

(i) If possible, excavating samples in a manner to prevent contamination with modern human DNA. Such precautions include the wearing of protective clothing and gloves during the excavation, and refraining from washing, treating or examining samples prior to performing DNA analysis.

(ii) A major source of contamination when performing ancient DNA analysis is from the person doing the laboratory work. To minimise the amount of modern human DNA contamination that the sample is exposed to from the researcher, a number of precautions are taken, including wearing protective clothing, such as a body suit, three layers of gloves, face mask, face shield and shoe covers.

(iii) As skeletal samples have often been handled, either at the excavation site or in museums, there is probably exogenous DNA on their surface. Hence, skeletal samples commonly have their outer-surfaces decontaminated prior to DNA extraction. Some common procedures used for decontaminating skeletal samples include mechanically removing the outer surface, exposing the samples to ultraviolet light (UV) (Lindahl, 1993) and immersing or wiping samples in bleach (sodium hypochlorite) (Malmstrom et al., 2007). The latter two methods are known to degrade/destroy DNA (Lindahl, 1993, Ohnishi et al., 2002). However, the efficacy of each of these proposed methods has been questioned. The soaking of skeletal samples in bleach (3% bleach to water) for 5 minutes was shown to remove exogenous DNA from skeletal remains that were purposely contaminated, while the endogenous DNA remained unharmed (Kemp and Smith, 2005). However, other investigations have shown that contaminating bones and teeth with modern DNA is actually quite difficult (Gilbert et al., 2006) and hence the use of purposely contaminated skeletal material to then support the efficiency of bleach soaking as a decontamination procedure is dubious. Neither bleach treatment nor UV irradiation appears to be able to completely remove surface contaminants. A study of human bone samples that had been bleach treated and UV irradiated, was still able to recover multiple, different human sequences from the bone extracts (Gilbert et al., 2005). Hence, no decontamination procedure appears to be 100% effective and therefore using a combination of the decontamination procedures is appropriate to remove as much exogenous contamination as possible.

(iv) A source of modern DNA contamination may come from the laboratory and laboratory equipment. Hence, to reduce contamination on surfaces of the laboratory and the laboratory equipment, both are routinely cleaned with bleach and irradiated with UV.

(v) A major precaution taken to control contamination from modern DNA is the separation of pre- and post-PCR analyses into separate work spaces. In the pre-PCR workspace, all sample decontamination and DNA extraction steps prior to DNA amplification are performed. This workspace is ideally in a clean room area, which is solely dedicated to ancient DNA work, and is located in a physically separated space, where no modern DNA work is conducted. By splitting the pre- and post-PCR analytical steps into separate spaces, this reduces the chance of ancient samples becoming contaminated with high concentration PCR products.

(vi) To monitor the presence of contamination in laboratory reagents, and/or the presence of cross contamination between samples, all extraction and PCR steps are conducted with multiple non-template controls or blanks. These blanks are amplified and run on a gel to check for the presence of contamination and sequenced if necessary (i.e. if an amplicon product is observed on the gel).

(vii) To ensure the authenticity of the sequences recovered from the skeletal material, the findings are replicated. This can involve extracting DNA from a sample twice, either in the same laboratory (on separate days) or in a different ancient DNA laboratory, which enables the impact of laboratory contamination to be assessed as a reason for the result. Furthermore, if independent DNA extraction cannot be performed, potentially due to a limited amount of skeletal tissue being available for DNA analysis, replication of PCR is also useful, as it can be used to monitor for the presence of sporadic contamination and DNA damage.

(viii) The authenticity of sequences obtained from ancient remains can also be assessed by the cloning of specific PCR products (e.g. for HVS1), to detect the presence of multiple individuals in the DNA extract. If only endogenous DNA is recovered, the majority of sequences should be from one individual, with potential sequence heterogeneity derived from DNA damage in the sequences. In contrast to cloning a PCR amplicon from a specific target, the cloning of ‘universal’ DNA sequences, such as those generated from the 16S rRNA gene, is necessary to identify multiple microbial sequences in the ancient DNA extract. Cloning of universal targets for DNA

extracts and non-template controls is useful for monitoring the presence of contaminating sequences in the DNA extracts, which can be identified from the non-template controls.

(ix) The molecular behaviour of DNA obtained from ancient remains is also used to assess whether it is endogenous. DNA from skeletal remains is expected to be fragmented and degraded, with a greater amount of shorter length DNA than longer fragments (Malmstrom et al., 2005, Malmstrom et al., 2007). Quantitative real-time PCR is a common method used to assess if DNA from ancient remains is displaying appropriate molecular behaviour (Malmstrom et al., 2005, Malmstrom et al., 2007). The method quantifies the amount of DNA in a sample for a specific fragment length, prior to PCR amplification.

The recovery of genetic material from preserved hard skeletal remains has become a key procedure for a wide range of research fields including ancient DNA, archaeology, anthropology and forensics. Despite this, very little is known about the amount and physical distribution of DNA in ancient bones and teeth, or the impact of standard analytical approaches on the quantity and quality of the few remaining molecules. This knowledge is especially important when dealing with human remains, which have significant morphological value (Hillson, 1979, Donlon, 1994, Hanihara, 2008), and hence it is necessary to sample both conservatively and effectively. By elucidating which methods are least detrimental and most efficient for the recovery of DNA from skeletal remains, it is possible that DNA may be obtained from skeletal material which is yet to yield positive DNA results. This is especially important for specimens which may contribute significantly to our understanding of in human evolution, such as *Homo floresiensis,* from which DNA has yet to be obtained. An article presented in Nature *News*, entitled ‘Researchers to drill for Hobbit history’ canvassed the potential of re-analysing important hominid remains, when the appropriate methodology was applied, as indicated from the results presented in Chapter 2 (see Appendix A).

1.3.2 Using ancient human DNA to study past demographic events

Ancient DNA analysis of prehistoric human populations has revealed demographic events in human history that were not visible from the genetic analysis of modern populations, such as the replacement of hunter-gatherers by agriculturists in Central and Northern Europe at the onset of the Neolithic period, as discussed in Section 1.2.2 (Haak et al., 2005, Haak et al., 2008, Bramanti et al., 2009, Haak et al., 2010). Initial studies of DNA from human remains

focused on single samples, such as 'The Ice Man' (Handt et al., 1994). However, the genetic analysis of a single prehistoric individual does not provide any information about the wider population, and severely limits the application of ancient DNA to investigation population demography. Some of the recent publications which have investigated ancient human DNA have incorporated multiple individuals, with 42 individuals sampled in Haak and colleagues papers (2010). The publication of studies with large numbers of ancient human remains that have been genetically analysed (Haak et al., 2005, Haak et al., 2008, Bramanti et al., 2009), has primarily occurred due to advances in ancient DNA methodology (Malmstrom et al., 2005, Malmstrom et al., 2007) and the establishment of authentication criteria (Cooper and Poinar, 2000), with strict monitoring of contamination (Gilbert et al., 2005, Gilbert et al., 2006).

The development of large datasets of ancient human DNA has enabled complex demographic models of population history to be tested, by using sophisticated, coalescent based analysis, such as Serial SimCoal (Anderson et al., 2005, Chan et al., 2006) (discussed in Section 1.2.2) and the Bayesian Skyline Plot (Drummond et al., 2005) (discussed below). Both these analyses incorporate ancient genetic data, which provides direct past genetic information. Serial SimCoal has been used to investigate population continuity, such as in Europe between post-LGM hunter-gatherer and Neolithic farming populations (Bramanti et al., 2009, Haak et al., 2010). Serial SimCoal has also been used to elucidate the ancestral population of modern groups, such as the Sardinians (Ghirotto et al., 2010). It was suspected that the modern Sardinian populations, Ogliastra and Gallura, had distinct ancestors (Ghirotto et al., 2010). Ogliastra was thought to be an isolated population descended from the Bronze Age Sardinians, whereas Gallura was thought to have had significant population input from the Italian mainland. The authors used mtDNA from Bronze Age Sardinians (n=23) and contemporary individuals from Ogliastra and Gallura (n=254) to assess the likelihood of population continuity in Sardinia between the Bronze Age and today. The ancient samples were subject to stringent contamination controls (e.g. dedicated ancient DNA work space, independent replication of results, wearing of protective clothing and use of non-template controls). With the use of Serial SimCoal, the authors were able to extensively assess alternative population histories by modelling seven different demographic scenarios. The ancient and modern genetic data was compared using both intrapopulation (e.g. haplotype diversity, number of segregating sites, Tajima's D) and interpopulation (F_{STS}) summary statistics, providing information about both changes in effective population size and population continuity. The models were compared to the data using ABC and determination of the most likely model was determined by estimation of a Bayesian p-value, as opposed to

using maximum likelihood methods and AIC scores, as discussed in Section 1.2.3. The Bayesian p-value corresponds to the probability that simulated data using the estimated posterior distributions of the parameters are more extreme that the observed data, hence a non-significant value indicates the observed summary statistics are very similar to the simulated summary statistics. However, the Bayesian p-value has been found to have limited power to reject more simplistic models, particularly when there is limited data (Bertorelle et al., 2010). The study showed that the most likely model given the ancient genetic data was that the Ogliastra population were descendants of the Bronze Age Sardinians. The Gallura population was shown to have most likely experienced immigration (presumably from the mainland), and this resulted in it being most likely to have a larger ancestral population compared to Ogliastra, which is in agreement with the archaeological record. This study highlighted the use of serial simulated coalescent analyses to show strong genealogical differences between prehistoric and modern inhabitants of nearby territories.

Another analytical approach used to investigate past demographic events is the Skyline Plot, implemented in the program Beast (Drummond et al., 2005). The Skyline Plot estimates effective population size directly from genetic diversity through time. The Skyline plot has been successfully used to reconstruct population history in bison, using ancient mtDNA sampled over the last 150, 000 years (Drummond et al., 2005). The original Skyline Plot analysis of this same bison ancient DNA dataset had revealed that a decrease in the genetic diversity of bison occurred approximately 37, 000 years before present (Shapiro et al., 2004). The revised Bayesian Skyline Plot analysis, which instead of inferring demographic history from an estimated genealogy, sampled the gene sequences directly, revealed a reduction in genetic diversity also occurred approximately 25, 000 years ago and then another at 15, 000 years ago, which coincides with the expansion of humans into North America (Drummond et al., 2005). To date, Bayesian Skyline Plot analysis has yet to be applied (or at least published) for ancient human genetic data. This is presumably because for Skyline Plot analysis to produce meaningful results, large datasets are required, preferentially with temporal spaced samples. However, suitable ancient human DNA datasets for Skyline Plot analyses are increasing (e.g (Bramanti et al., 2009, Haak et al., 2010)).

1.3.3 Using ancient bacterial DNA to study the evolution of pathogens

Ancient DNA has been touted as the only way to directly study the evolution of human-associated pathogens (Salo et al., 1994), as analysis of skeletal pathologies rarely enables the exact identification of a specific pathogen (Wood et al., 1992) and provides no information about the potential changes in genetic structure of the pathogen over time. To date ancient

DNA research has focused on attempting to extract pathogens from skeletal remains that morphologically present with a specific disease, such as tuberculosis (Salo et al., 1994, Rothschild et al., 2001, Zink et al., 2001, Brosch et al., 2002, Taylor et al., 2005), leprosy (Haas et al., 2000, Montiel et al., 2003), malaria (Sallares, 2001) and syphilis (Kolman et al., 1999). Among ancient pathogen DNA studies, based on the frequency of reporting, the greatest success appears to be the recovery of the bacterium which causes human tuberculosis (*Mycobacterium tuberculosis*) that has reportedly been amplified from multiple different skeletal remains and mummified tissue (Salo et al., 1994, Rothschild et al., 2001, Zink et al., 2001, Brosch et al., 2002, Taylor et al., 2005), and even from samples that were 2, 200 years old (Taylor et al., 2005). However, it is difficult to reliably identify *Mycobacterium tuberculosis* from other species of the *Mycobacterium* genus, which includes soil bacteria, due to the close sequence similarity of all species in the genus, (94.3% for the 16S gene (Smith et al., 2009)). Furthermore, the extremely close sequence similarity (99.95%) of *Mycobacterium tuberculosis* to the *Mycobacterium* pathogens present in goats, cattle and rodents (Smith et al., 2009), makes it nearly impossible to correctly identify the human associated mycobacterium from short and fragmented pieces of ancient DNA. Aside from the issue of being able to correctly identify the pathogen from ancient DNA, skeletal remains may also not be a good reserve of ancient pathogen DNA. A study by Bouwman and Brown (2005), in which they attempted to amplify the spirochete bacteria which causes syphilis, *Treponema pallidum,* from skeletal remains that morphologically presented with the disease, which ranged in age from the 9th to 19th century, reported that they could not reliably amplify treponema sequences. This study may not have been able to recover the pathogen because by the time a skeletal pathology of syphilis is observed, which occurs during the tertiary stage of the disease, there is basically no pathogen left in the bones (von Hunnius, 2007). The lack of *Treponema pallidum* in skeletal elements presenting with pathological lesions was revealed by the investigation of rabbits purposely infected with syphilis (von Hunnius, 2007).

The authenticity of findings from ancient DNA studies of pathogens has often been questioned due to the lack of commonly used ancient DNA authentication criteria (Gilbert et al., 2004). A review of 65 ancient pathogen DNA studies, published between 1995 and 2006 (Roberts and Ingham, 2008), revealed that the majority of investigations were found to have not conducted research in accordance with ancient DNA authentication criteria. For example, 80% of studies failed to discuss or consider the excavation conditions, only 18% used protective clothing during ancient DNA analysis, and a mere 25% used a dedicated ancient DNA facility or workroom in which to perform pre-PCR analytical steps. An interesting statistic from this review paper was the lack of consideration by the majority of investigators

(92%) of the influence of contamination from the environment, such as by soil bacteria, on the results. Determination of the presence of environmental contamination in the samples would be extremely important for studies on the mycobacterium pathogens, including tuberculosis and leprosy, which have a high sequence similarity to soil microbes.

An ancient pathogen DNA study which has reported to have adhered to strict authentication and contamination control procedures is Haensch and colleagues (2010) investigation, in which they successfully recovered the Gram-negative bacterium, *Yersinia pestis* from plaque victims (14th-17th century, from northern, southern and central Europe). The contamination control procedures implemented in this study included using a dedicated ancient DNA facility, separate pre- and post-PCR work spaces, and performing strict sample decontamination procedures (surface removed and UV irradiation). To authenticate the results, the majority of samples were extracted twice on different days, non-template controls were run in parallel with extractions and PCRs, and soil samples that were taken from near the excavated skeletal remains were assessed for the presence of *Y. pestis*. Furthermore, a non-PCR based method was used to independently validate the ancient DNA findings. The non-PCR test involved an immunochromatographic-based assay to detect the presence of the F1 envelope glycoprotein, which is specific to Y. pestis. The authors were able to successfully amplify *Y. pestis* from 10 of the 76 samples assessed, and from all these samples they were also able to show the presence of *Y. pestis* protein from the F1 analysis. The *Y. pestis* sequences do not appear to be contaminants, as *Y. pestis* was not found in the blanks, soil samples or in any of the 28 skeletal remains analysed that were known not to have plague (buried before the plaque occurred) and were recovered from the same graveyards as the plague victims. This study demonstrates that ancient pathogen DNA can be recovered from skeletal remains and the results can be authenticated, however the low success rate (13%) indicates that skeletal remains are not ideal reserves of preserved ancient bacterial DNA.

The studies discussed so far have focused on identifying a specific human-associated pathogen. Hence, this approach yields no knowledge about the wider diversity of commensal microbes within the human host. The human microbiota consists of all the microbial lineages within our body (Ley et al., 2008), which make up 90% of the 10^{14} cells in humans (Marsh, 2010). The majority of these microbes are bacteria, with archaea making up a small portion (Ley et al., 2008). These commensal microbial communities directly influence our health, having the ability to cause oral diseases (Aas et al., 2005), and have the ability to modify our genomes, with elements of commensal microbial genomes found within the human genome (Lander et al., 2001). However, we remain largely ignorant of the response of human

associated microbiota to major past bio-cultural transitions in human history. Our lack of knowledge about the past state of human microbiota results from the lack of a preserved source of commensal bacteria which is suitable for genetic analysis. In this PhD (Chapter 4), I have identified dental calculus as a preserved source of ancient oral microbial DNA, which is an important component of the human microbiota.

1.3.3.1 Dental calculus

Dental calculus is a highly mineralised form of plaque, which is extremely dense in bacteria ($2x10^8$/mg) (Scheie, 1994) and occurs both above and below the gum line, which is referred to as supra- and subgingival calculus (Jin and Yip, 2002), respectively. The properties of supra- and subgingival calculus are described in Table 1. The bacteria in plaque become mineralised as calculus forms, with calcium phosphate found both within and between the cells which locks the bacteria in calculus in a crystalline structure that is similar to bone (Jin and Yip, 2002). Scanning electron microscopy analysis of archaeological dental calculus deposits has shown that the morphology of bacteria within dental calculus is well preserved (Dobney and Brothwell, 1986), as can be seen in Figure 1. Furthermore, transmission electron microscopy analysis has revealed the presence of ancient bacterial DNA within archaeological dental calculus from Neolithic human remains (Preus et al., 2011). Preus and colleagues (2011) study showed that dental calculus contained ancient bacterial DNA by exposing the samples to UV irradiation to produce thymine dimers, which were then identified using H3 antibodies which are specific to fragments of DNA containing thymine dimers that are at least 20-30 bases long. Although microscopic analysis of archaeological dental calculus has indicated that ancient oral microbial DNA is preserved in calculus, the types of bacterial species in the calculus deposit cannot be determined from microscopic analysis. Hence, ancient DNA analysis of archaeological dental calculus samples offers a novel means can to investigate the past composition and diversity of oral microbiota, as described in Chapter 4.

	Supragingival	Subgingival
Location	Coronal to the gingival margin. Most abundant on the lingual surfaces of mandibular incisors, buccal surfaces of the maxillary first molars and adjacent to salivary glands.	Below the gingival margin on root surface. No difference between teeth.
Colour	Yellow/white	Brown/black
Distribution	Adjacent to salivary ducts	Random
Composition	Concentration of Ca, Mg, Fl, Sr and Zn lower than subgingival. Carbonate and Mn higher than subgingival.	Opposite to supraginigival
Mineral content and source	Average mineral content is 37%. This is obtained from the saliva.	Average mineral content is 58%. This is obtained from the gingival cervical fluid.
Crystal type and size	Primarily octacalcium phosphate (OCP) and carbonate hydroxyapatite, which are small and needle shaped. There are small amounts of dicalcium phosphate dihydrate (DCPD), which are long and ribbon like.	Mainly magnesium substituted tricalcium phosphate, which is also called whitlock, there is no dicalcium phosphate dihydrate and in general the crystals are small.
Formation	Heterogeneous nucleation and crystal growth, with variable calcification.	A combination of heterogeneous nucleation and crystal growth, with homogenous crystal growth.
Microorganisms	Contains both calcified and non-calcified microorganisms. The microbes are arranged perpendicular to the tooth.	Mainly calcified microorganisms. No pattern in arrangement.
Pathogenic Potential	Can be associated with caries and periodontal disease	Can be associated with periodontal disease

Table 1.1. Characteristics of supragingival and subgingival calculus.

NOTE:
This figure is included on page 41 of the print copy of the thesis held in the University of Adelaide Library.

Figure 1.1. Calculus from the macroscopic to microscopic scale. a) Mandibular molar from a medieval UK population with supra- and subgingival calculus on the buccal surface; b) another specimen from the same population, showing remarkable microscopic incremental growth lines in calculus, magnification x50 (scanning electron micrograph); c) further specimen showing a large unidentified 'golf ball shaped' colony of rod-shaped bacteria, magnification x3000 (scanning electron micrograph). 1b) and c) are courtesy of K. Dobney (Lilley et al., 1994) .

To understand why dental calculus is an ideal reserve for preserved ancient oral microbial DNA, the features of dental calculus will be described in the following sections, including how the calculus deposit forms and what it is made up of, both the cellular and non-cellular contents.

Formation of Plaque

The development of plaque, also called an oral biofilm, is the precursor to calculus formation. As soon as a tooth erupts, or soon after it has been cleaned, a conditioning biofilm covers the tooth surface (also called the acquired pellicle) (Marsh, 2010). The biofilm is comprised of proteins and glycoproteins, which for supragingival plaque are obtained from the saliva, and for subgingival plaque are obtained from the gingival cervical fluid (GCF) (Bowden and Hamilton, 1998). While all oral bacterial species have the ability to adhere to surfaces of the mouth (otherwise they are transported by saliva to the digestive tract) (Kolenbrander et al., 2010), only a few bacterial species are able to adhere to the biofilm on the tooth surface initially through the action of surface molecules and appendages of the bacteria (Bowden and Hamilton, 1998). Common primary colonisers included streptococcus and actinomyces species (Diaz et al., 2006). The adherence of microbes differs between the various tooth surfaces. The surface of the crown is colonised by bacteria that can resist the abrasive forces of saliva flow and mastication, such as streptococci (Marsh, 2003). In the crevices of the gingiva, protection and bathing with the nutritionally rich GCF increases the range of colonising bacterial species, and there is an increased frequency of periodontal-disease

associated pathogens, such as *Porphyromonas gingivalis* (Hillson, 1996, Kolenbrander et al., 2010). Following the binding of initial colonisers to the tooth surface, secondary colonisers bind to receptors on the primary colonising bacteria (Kolenbrander et al., 2006). The attachment of the of the secondary to primary colonisers produces short-range interactions (co-adhesion) between bacteria (Hillson, 1996). There are many specific interactions between initial, secondary and late bacterial colonisers in oral biofilms (thousands have been described) (Kolenbrander et al., 2006), such as streptococci binding with *Veillonella* species (Palmer et al., 2006). In contrast, there are certain bacteria in the oral biofilm, such as species in the *Fusobacterium* genus, which are able to bind to initial, early and late colonisers, and hence form a 'bridge' within the plaque (Kolenbrander and London, 1993). The growth of adherent cells on the biofilm changes the oral environment to promote further growth and succession of bacteria.

Formation of Calculus - Extracellular Matrix

The calcification of microorganisms in plaque is reliant on the development of an extracellular matrix and mineral deposition (Yin and Jip 2000). The extracellular matrix makes up the bulk of calculus and is primarily (80%) composed of inorganic salts, two-thirds of which are in crystal form (Yin and Jip 2000). The major elements of the extracellular matrix are calcium (Ca) and phosphate (P) (Lieverse, 1999). The minor elements of calculus include carbonate, sodium (Na), magnesium (Mg) and fluoride (Fl) (Lieverse, 1999). The crystal form of inorganic salts in calculus closely resembles bone, enamel, dentine and cementum, as all contain carbonate hydroxyapatite (CHA) (Hillson, 1979). Dental calculus contains three additional types of crystals to bone and teeth, these include dicalcium phosphate dihydrate (DCPD), octacalcium phosphate (OCP) and magnesium substituted tricalcium phosphate, which is also called whitlock (Gonzales and Sognnaes, 1960, Zander et al., 1960, Hillson, 1996).

The mineralisation of microbes in calculus occurs in stages. Firstly small crystals (typically DCPD and OCP) are deposited in the intermicrobial spaces (Roberts-Harry and Clerehugh, 2000, Roberts-Harry et al., 2000), when the plaque is super-saturated with Ca and P (Hillson, 1979). Gradually the bacteria become calcified, beginning with the outer surface and followed by mineralisation of the whole cell (Zander et al., 1960). The complete calcification of microbes in calculus can occur as quickly as 38 hours after plaque deposition (Schroeder, 1969). In mature calculus deposits, such as those which have been present for over six months, larger crystals are deposited including hydroxyapatite and whitlock (Roberts-Harry and Clerehugh, 2000, Roberts-Harry et al., 2000). The type of crystals in supra- and

subgingival calculus differs due to their varying sources of Ca and P (Table 1). Supragingival calculus obtains ions from the saliva and subgingival from the GCF (Roberts-Harry et al., 2000). The higher mineral content in GCF may be responsible for the greater levels of calcification in sub- compared to supragingival calculus (Friskopp, 1983).

Cellular Component

The microbial makeup of the oral environment generally reflects the microbial composition of dental calculus (Marsh, 2005). The determination of the microbial composition of the oral environment was originally achieved through the use of cell-culture methods (Marsh, 2010), which have largely been superseded by genetic analysis because of the cheaper costs and the ability to identify bacterial species that were unculturable (Marsh, 2003). The taxonomic genetic analysis of the oral environment has primarily been achieved by sequencing the small subunit (16S) of the ribosomal (r)RNA gene. This gene was targeted because it is present in all cellular organisms and it is highly phylogenetically informative (Woese and Fox, 1977). The genetic analysis of subgingival plaque, using the 16S gene, has revealed that plaque contains over 700 bacterial and archaeal species, and 50% of these had not previously been found because the microbes were unculturable (Paster et al., 2001). The genetic analysis of oral samples has revealed that the human mouth contains a suite of phyla (Figure 1.2), which is dominated by Firmicutes (36.7%). Although at the phylum level, the microbial composition of the mouth varies minimally between the different surfaces, such the tongue, teeth and mucosa, there is variation between the different oral surfaces at the species or operational taxonomic unit (OTU, 97% similarity) (Paster et al., 2001, Aas et al., 2005, Aas et al., 2008, Faveri et al., 2008, Lazarevic et al., 2010).

The determination of the bacterial composition of dental calculus has not received great attention from molecular studies, which have instead focused on analysing dental plaque (Choi et al., 2000, Paster et al., 2001, Aas et al., 2005, Hintao et al., 2007). This may be due to the greater ease of sampling, since the removal of calculus from a tooth would generally need to be performed by a dental professional. The microbial composition of dental plaque and calculus appears to be similar from both cell-culture studies (Sidaway, 1978) and genetic analysis (see Chapter 4). The microbial composition of dental plaque in modern populations will be described as an indication of the general microbial composition of dental calculus in modern populations.

NOTE:
This figure is included on page 44 of the print copy of the thesis held in the University of Adelaide Library.

Figure 1.2. The phylum level, microbial composition of modern oral samples from the Human Oral Microbiome Database (HOMD) (Dewhirst et al., 2010).

Dental plaque is diverse, with 12-27 species typically found in a deposit (Aas et al., 2005). A study of 185 plaque samples revealed the presence of five bacterial complexes (Socransky et al., 1998); 1) *Bacteroides forsythus*, *Porphyromonas gingivalis* and *Treponema denticola*, 2) *Fusobacterium nucleatum*, *Periodonticum* subspecies, *Prevotella intermedia, Prevotella nigrescens* and *Peptostreptococcus micros*, 3) *Streptococcus sanguis, S. oralis, S. mitis, S. gordonii* and *S. intermedius*, 4) *Capnocytophaga* species, in addition to *Campylobacter concisus, Eikenella corrodens* and *Aggregatibacter actinomycetemcomitans*, 5) *Veillonella parvula* and *Actinomyces odontolyticus*. Aside from these complexes, genetic analysis has revealed the presence of many other bacterial species with 700 microbial species identified in subgingival plaque (Paster et al., 2001). Of particular interest to this PhD are those microbes which are associated with oral diseases and influenced by diet, and hence were probably affected by the introduction of agriculture during the Neolithic.

The most common oral diseases are dental caries and periodontal disease, and the development of both has been found to be related to a carbohydrate based diet (see Section 1.2.4). Caries is an extremely common pathology, affecting 60-90% of school-aged children in industrialised countries, whilst periodontal disease occurs in 5-20% of the adult population worldwide (Petersen et al., 2005). Although both oral diseases can occur at the same time and even on the same tooth (De Soete et al., 2005), the aetiology of caries differs substantially from periodontal disease. Caries is a response to a prolonged lowering of salivary pH due to the consumption of sucrose (Marsh, 2010). A cariogenic plaque contains bacteria which are able to use fermentable carbohydrates and convert them to lactic acid (De Soete et al., 2005).

The production of lactic acid creates a locally acidic environment, which attracts acid tolerant bacteria (Aas et al., 2008). The bacteria which have been found to be associated with decay include *S. mutans*, *S. sobrinus*, low pH tolerant streptococci, such as *S. oralis, S. mitis, S. anginuosus,* in addition to *Lactobacilli*, *Actinomyces*, *Bifidobacterium*, *Propionibacterium* and *Veillonella* (Becker et al., 2002, Munson et al., 2004, Beighton, 2005, Aas et al., 2008). Carious plaques in general have been found to have lower microbial diversity compared to the plaque from healthy individuals (Gross et al., 2010).

In comparison to dental caries, periodontal disease is associated with a locally alkaline environment, which facilitates the deposition of bacteria at the gingival margin and these bacteria produce an inflammatory state (Loesche, 2007). The bacteria associated with periodontal disease include Gram-negative (e.g. *Porphyromonas gingivalis, Treponema denticola, Prevotella intermedia, Aggregatibacter actinomycetemcomitans,* and *Fusobacterium nucleatum*) and Gram-positive taxa (e.g. Clostridia species, such as *Megasphaera, Dialister, Peptococcaceae, Filifactor, Peptostreptococcus,* and Archaea genera *Methanobrevibacter*) (Lepp et al., 2004, Kumar et al., 2005, Filoche et al., 2010). The presence of periodontal disease also has far broader medical implications than oral health. Periodontal disease is a risk factor in the development of heart disease and stroke (Dave and Van Dyke, 2008), early-onset diabetes (Grossi and Genco, 1998), osteoporosis (Gomes-Filho et al., 2007) and premature births (Jeffcoat et al., 2003). Periodontal disease-associated pathogens are thought to contribute to the pathogenesis of systemic diseases (such as diabetes, heart disease and arthritis) by producing a prolonged inflammatory state (Dave and Van Dyke, 2008). Furthermore, periodontal disease-associated pathogens have been found to disseminate from the mouth to other areas of the body and where these pathogens can then produce an inflammatory state locally, as suggested by the finding of *Aggregatibacter actinomycetemcomitans* and *Porphyromonas gingivalis* in athersclerotic plaques (Li et al., 2000). Hence, the identification of key periodontal disease-causing pathogens in ancient dental calculus samples, which are also implicated in systemic diseases, can be used to infer the health of an individual. Together these facts strongly suggest that dental calculus will record multiple facets of the diet and disease-state of individual skeletons. The additional benefit of using dental calculus is that it is considered a ‘non-valuable’ scientific resource compared to human bone or teeth. Indeed, dental calculus has been known to be physically removed by museum curators to improve the visual appearance of specimens.

1.4 Thesis structure

This PhD is written as a mixture of publications and chapters. The contents of the chapters and their relevant Appendices are as follows:

Chapter 2 – Survival and recovery of DNA from ancient bones and teeth

This published manuscript assesses the impact of common sampling techniques on the retrieval of DNA from human skeletal remains. Additionally, the amount of DNA in different bone and tooth tissues, across a broad range of environments and time frames is used to reveal DNA degradation patterns. This study showed that common sampling techniques were detrimental to the recovery of DNA from skeletal remains and that DNA degradation rates were influenced by the post-mortem environment and tissue type. The reference is:

Adler, C. J., Haak, W., Donlon, D., Cooper, A. Survival and recovery of DNA from ancient teeth and bones. *Journal of Archaeological Science*. (2010) doi 10.1016/j.jas.2010.11.010.

The implications of this article to the field of ancient DNA research have been covered in Nature *News* as can be seen in Appendix A.

Chapter 3 - Ancient DNA reveals discontinuity between successive Central European, Neolithic cultures and their varied impact on modern mtDNA lineages

This chapter investigates the population structure of Central European Neolithic cultures in order to test population continuity since the introduction of farming in the early Neolithic until today. To assess this, ancient human mtDNA was recovered from five Neolithic and Bronze Age cultures and compared to modern European data from the same regions through the use of population genetic analysis. This study showed that population continuity was unlikely since the introduction of agriculture during the early Neolithic until today. It was found that population change between the Early Neolithic and Late Neolithic/Early Bronze Age has had a greater impact on the population structure of contemporary Central Europeans than the initial introduction of farming.

I also contributed to the paper by Haak and colleagues (2010) presented in Appendix B, which is related to the study presented in Chapter 3. This study focuses on the genetic make-up of the first farmers in Central Europe and their geographical relationship to

modern day populations. I contributed to this paper by performing experimental work (qPCR), analyses (Serial SimCoal), writing and editing parts of the paper.

Chapter 4 - Ancient DNA from oral microbiota records the impact of dietary change on the evolution of human oral diseases

This paper has been submitted to *Nature*, and is currently under review. It shows that dental calculus (calcified plaque), recovered from archaeological human remains (dated between 5500BC – 850AD) contains a remarkable preserved genetic record of ancient oral microbiota. The preliminary results show that, prior to the Industrial Revolution (150-200 years) the diversity of oral bacteria present in the mouths of early farming groups (Neolithic, Bronze Age and Medieval Europeans) was significantly higher than that found in the present day. In contrast, modern Europeans show a significantly restricted suite of bacteria, dominated by taxa involved in decay. This proof-of-concept work provides a new method for investigating the evolutionary history of human microbiota.

Chapter 5 – Conclusion

The closing chapter summarises the main results and discusses how the findings address the overall aims of this thesis, which are to use ancient genetic data to investigate the impact of farming on the population structure and health of Europeans, and also to assess the methodologies used to recover DNA from skeletal remains. The limitations of the research and future directions are also discussed.

REFERENCES

Aas, J A, Griffen, A L, Dardis, S R, Lee, A M, Olsen, I, Dewhirst, F E, Leys, E J & Paster, B J 2008. Bacteria of dental caries in primary and permanent teeth in children and young adults. *J Clin Microbiol,* 46**,** 1407-17.

Aas, J A, Paster, B J, Stokes, L N, Olsen, I & Dewhirst, F E 2005. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol,* 43**,** 5721-32.

Achilli, A, Rengo, C, Battaglia, V, Pala, M, Olivieri, A, Fornarino, S, Magri, C, Scozzari, R, Babudri, N, Santachiara-Benerecetti, A S, Bandelt, H J, Semino, O & Torroni, A 2005. Saami and Berbers--an unexpected mitochondrial DNA link. *Am J Hum Genet,* 76**,** 883-6.

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, Loogvali, 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**,** 910-8.

Adams, J M & Faure, H 1997. Preliminary vegetation maps of the world since the last glacial maximum: an aid to archaeological understanding. *Journal of Arcaeological Sciences,* 24**,** 623-647.

Akkermans, P M M G, Cappers, R, Cavallo, C, Nieuwenhuyse, O, Nilhamn, B & Otte, I N 2006. Investigating the Early Pottery Neolithic of Northern Syria: New evidence from Tell Sabi Abyad. *American Journal of Archaeology,* 110**,** 123-156.

Ammerman, A J & Cavalli-Sforza, L L 1971. Measuring the Rate of Spread of Early Farming in Europe. *Man New Series* 6**,** 674–688.

Ammerman, A J & Cavalli-Sforza, L L 1973. A population model for the diffusion of early farming in Europe. *In:* RENFREW, C. (ed.) *The explanation of culture change.* London: Duckworth.

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

Anderson, C N, Ramakrishnan, U, Chan, Y L & Hadly, E A 2005. Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics,* 21**,** 1733-4.

Angel, J L 1966. Porotic hyperostosis, anemias, malarias, and marshes in the prehistoric Eastern Mediterranean. *Science,* 153**,** 760-3.

Anthony, D W 1990. Migration in archeology: The baby and the bathwater. *American Anthropologist* 92**,** 895-914.

Antia, R, Regoes, R R, Koella, J C & Bergstrom, C T 2003. The role of evolution in the emergence of infectious diseases. *Nature,* 426**,** 658-61.

Asouti, E 2006. Beyond the Pre-Pottery Neolithic B interaction sphere. *Journal of World Prehistory,* 20**,** 87–126.

Aufderheide, A C, Rodriguez-Martin, C & Langsjoen, O 1998. *The Cambridge encyclopedia of human paleopathology* Cambridge Cambridge University Press.

Bar-Yosef, O 1998. The Natufian culture in the Levant, threshold to the origins of agriculture. *Evolutionary Anthropology: Issues, News, and Reviews,* 6**,** 159-177.

Bar-Yosef, O & Belfer-Cohen, A 1989. The origins of sedentism and farming communities in the Levant. *Journal of World Prehistory,* 3**,** 447-498.

Bar-Yosef, O & Belfer-Cohen, A 1992. From foraging to farming in the Mediterranean Levant. *In:* GEBAUER, A. B. & PRICE, T. D. (eds.) *Transitions to Agriculture in Prehistory.* Madison: Prehistory Press.

Bar-Yosef, O & Goren, N 1973. Natufian remains in Hayonim Cave. *Paléorient,* 1**,** 49-68

Barbujani, G, Bertorelle, G & Chikhi, L 1998. Evidence for Paleolithic and Neolithic gene flow in Europe. *Am J Hum Genet,* 62**,** 488-92.

Barker, G, Barton, H, Bird, M, Daly, P, Datan, I, Dykes, A, Farr, L, Gilbertson, D, Harrisson, B, Hunt, C, Higham, T, Kealhofer, L, Krigbaum, J, Lewis, H, Mclaren, S, Paz, V, Pike, A, Piper, P, Pyatt, B, Rabett, R, Reynolds, T, Rose, J, Rushworth, G, Stephens, M, Stringer, C, Thompson, J & Turney, C 2007. The 'human revolution' in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). *J Hum Evol,* 52**,** 243-61.

Barnes, I, Matheus, P, Shapiro, B, Jensen, D & Cooper, A 2002. Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science,* 295**,** 2267-70.

Beaumont, M A, Zhang, W & Balding, D J 2002. Approximate Bayesian computation in population genetics. *Genetics,* 162**,** 2025-35.

Becker, M R, Paster, B J, Leys, E J, Moeschberger, M L, Kenyon, S G, Galvin, J L, Boches, S K, Dewhirst, F E & Griffen, A L 2002. Molecular analysis of bacterial species associated with childhood caries. *J Clin Microbiol,* 40**,** 1001-9.

Beighton, D 2005. The complex oral microflora of high-risk individuals and groups and its role in the caries process. *Community Dent Oral Epidemiol,* 33**,** 248-55.

Bentley, R A 2006. Strontium Isotopes from the Earth to the Archaeological Skeleton: A Review. *J Archaeol Method Theory,* 13.

Bertorelle, G, Benazzo, A & Mona, S 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol Ecol,* 19**,** 2609-25.

Bogucki, P 1996. The spread of early farming in Europe. *American Scientist,* 84**,** 242-253

Bouwman, A S & Brown, T A 2005. The limits of biomolecular palaeopathology: ancient DNA cannot be used to study venereal syphilis *Journal of Archaeological Science,* 32**,** 703-713.

Bowden, G H & Hamilton, I R 1998. Survival of oral bacteria. *Crit Rev Oral Biol Med,* 9**,** 54-85.

Braidwood, R J, Howe, B & Reed, C A 1961. The Iranian Prehistoric Project: New problems arise as more is learned of the first attempts at food production and settled village life. *Science,* 133**,** 2008-10.

Bramanti, B, Thomas, M G, Haak, W, Unterlaender, M, Jores, P, Tambets, K, Antanaitis-Jacobs, I, Haidle, M N, Jankauskas, R, Kind, C J, Lueth, F, Terberger, T, Hiller, J, Matsumura, S, Forster, P & Burger, J 2009. Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. *Science*.

Breasted, J H 1906-7. *Ancient Records of Egypt: Historical Documents from the Earliest Times to the Persian Conquest* Chicago, University of Chicago Press.

Brosch, R, Gordon, S V, Marmiesse, M, Brodin, P, Buchrieser, C, Eiglmeier, K, Garnier, T, Gutierrez, C, Hewinson, G, Kremer, K, Parsons, L M, Pym, A S, Samper, S, Van Soolingen, D & Cole, S T 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proc Natl Acad Sci U S A,* 99**,** 3684-9.

Burnham, K P & Anderson, D R 2002. *Model Selection and Inference: A Practical Information - Theoretic Approach,* New York, Springer-Verlag.

Cavalli-Sforza, L L, Menozzi, P & Piazza, A 1993. Demic expansions and human evolution. *Science,* 259**,** 639-46.

Cavalli-Sforza, L L, Menozzi, P & Piazza, A 1994. *The history and geography of human genes,* Princeton, Princeton University Press.

Chan, Y L, Anderson, C N & Hadly, E A 2006. Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. *PLoS Genet,* 2**,** e59.

Chikhi, L, Nichols, R A, Barbujani, G & Beaumont, M A 2002. Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci U S A,* 99**,** 11008-13.

Childe, V G 1925. *The dawn of European civilisation.,* London, Kegan Paul.

Childe, V G 1935. *New Light on the Most Ancient East.,* London, Routledge and Kegan Paul.

Choi, B K, Park, S H, Yoo, Y J, Choi, S H, Chai, J K, Cho, K S & Kim, C K 2000. Detection of major putative periodontopathogens in Korean advanced adult periodontitis patients using a nucleic acid-based approach. *J Periodontol,* 71**,** 1387-94.

Clark, G 1965. Radiocarbon dating and the spread of farming economy. *Antiquity,* 39**,** 45-48.

Clark, P U, Mccabe, A M, Mix, A C & Weaver, A J 2004. Rapid rise of sea level 19,000 years ago and its global implications. *Science,* 304**,** 1141-4.

Cohen, M N 1977. *The Food Crises in Prehistory: Overpopulation and the Origins of Agriculture,* New Haven, Yale University Press.

Cohen, M N & Armelagos, G J 1984. *Paleopathology at the Origins of Agriculture,* Orlando, Academic.

Cooper, A, Mourer-Chauvire, C, Chambers, G K, Von Haeseler, A, Wilson, A C & Paabo, S 1992. Independent origins of New Zealand moas and kiwis. *Proc Natl Acad Sci U S A,* 89**,** 8741-4.

Cooper, A & Poinar, H N 2000. Ancient DNA: do it right or not at all. *Science,* 289**,** 1139.

Cordain, L, Eaton, S B, Sebastian, A, Mann, N, Lindeberg, S, Watkins, B A, O'keefe, J H & Brand-Miller, J 2005. Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr,* 81**,** 341-54.

Cox, M & Mays, S 2000. *Human osteology in archaeology and forensic science* London, Greenwich Medical Media.

Darwin, C 1868. *The Variation of Animals and Plants Under Domestication,* New York, Appelton.

Dave, S & Van Dyke, T 2008. The link between periodontal disease and cardiovascular disease is probably inflammation. *Oral Dis,* 14**,** 95-101.

Davis, B a S, Brewer, S, Stevenson, A C & Guiot, J 2003. The temperature of Europe during the Holocene reconstructed from pollen data. *Quaternary Science Reviews,* 22**,** 1701-1716.

Davis, S J M 1981. The effects of temperature change and domestication on the body size of late Pleistocene to Holocene mammals of Israel. *Paleobiology,* 7**,** 101-114.

De Soete, M, Dekeyser, C, Pauwels, M, Teughels, W, Van Steenberghe, D & Quirynen, M 2005. Increase in cariogenic bacteria after initial periodontal therapy. *J Dent Res,* 84**,** 48-53.

Denham, T P, Haberle, S G, Lentfer, C, Fullagar, R, Field, J, Therin, M, Porch, N & Winsborough, B 2003. Origins of agriculture at Kuk Swamp in the highlands of New Guinea. *Science,* 301**,** 189-93.

Dewhirst, F E, Chen, T, Izard, J, Paster, B J, Tanner, A C, Yu, W H, Lakshmanan, A & Wade, W G 2010. The human oral microbiome. *J Bacteriol,* 192**,** 5002-17.

Diamond, J 1997. *Guns, germs, and steel : the fates of human societies,* New York, W.W. Norton & Co.

Diamond, J 2002. Evolution, consequences and future of plant and animal domestication. *Nature,* 418**,** 700-7.

Diaz, P I, Chalmers, N I, Rickard, A H, Kong, C, Milburn, C L, Palmer, R J, Jr. & Kolenbrander, P E 2006. Molecular characterization of subject-specific oral microflora during initial colonization of enamel. *Appl Environ Microbiol,* 72**,** 2837-48.

Dobney, K M & Brothwell, D R 1986. Dental calculus: its relevance to ancient diet and oral ecology. *In:* CRUWYS, E. & FOLEY, R. A. (eds.) *Teeth and Anthropology.* B.A.R. International Series.

Dobson, A P & Carper, E R 1996. Infectious Diseases and Human Population History. *BioScience* 46**,** 115-126

Donlon, D 1994. Aboriginal skeletal collections and research in physical anthropology: An historical perspective. *Australian Archaeology,* 39**,** 73-82

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**,** 1185-92.

Dupanloup, I, Bertorelle, G, Chikhi, L & Barbujani, G 2004. Estimating the impact of prehistoric admixture on the genome of Europeans. *Mol Biol Evol,* 21**,** 1361-72.

Eshed, V, Gopher, A & Hershkovitz, I 2006. Tooth wear and dental pathology at the advent of agriculture: new evidence from the Levant. *Am J Phys Anthropol,* 130**,** 145-59.

Faveri, M, Mayer, M P, Feres, M, De Figueiredo, L C, Dewhirst, F E & Paster, B J 2008. Microbiological diversity of generalized aggressive periodontitis by 16S rRNA clonal analysis. *Oral Microbiol Immunol,* 23**,** 112-8.

Fenner, F 1996. Poxvirus. *In:* FIELDS, B. N. (ed.) *Fields Virology.* Philadelphia: Lippincott-Raven Publishers.

Filoche, S, Wong, L & Sissons, C H 2010. Oral biofilms: emerging concepts in microbial ecology. *J Dent Res,* 89**,** 8-18.

Flannery, K V 1973. The origins of agriculture. *An. Rev. Anthropol.,* 2**,** 271 – 310.

Formicola, V, Milanesi, Q & Scarsini, C 1987. Evidence of spinal tuberculosis at the beginning of the fourth millennium BC from Arene Candide cave (Liguria, Italy). *Am J Phys Anthropol,* 72**,** 1-6.

Fornander, E, Eriksson, G & Lidén, K 2008. Wild at heart: Approaching Pitted Ware identity, economy and cosmology through stable isotopes in skeletal material from the Neolithic site Korsnäs in Eastern Central Sweden. *Journal of Anthropological Archaeology,* 27**,** 281–297.

Forster, P 2004. Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Philos Trans R Soc Lond B Biol Sci,* 359**,** 255-64; discussion 264.

Francois, O, Currat, M, Ray, N, Han, E, Excoffier, L & Novembre, J 2010. Principal component analysis under population genetic models of range expansion and admixture. *Mol Biol Evol,* 27**,** 1257-68.

Frisancho, A R, Garn, S M & Ascoli, W 1970. Childhood retardation resulting in reduction of adult body size due to lesser adolescent skeletal delay. *Am J Phys Anthropol,* 33**,** 325-36.

Friskopp, J 1983. Ultrastructure of nondecalcified supragingival and subgingival calculus. *J Periodontol,* 54**,** 542-50.

Galor, O & Moav, O 2002. Natural selection and the origin of economic growth. *The Quarterly Journal of Economics* 117**,** 1133–1191.

Gamble, C 1999. *The Palaeolithic societies of Europe,* Cambridge Cambridge University Press.

Garrigan, D & Hammer, M F 2006. Reconstructing human origins in the genomic era. *Nat Rev Genet,* 7**,** 669-80.

Gepts, P & Papa, R 2001. *Evolution during Domestication*, John Wiley & Sons, Ltd.

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**,** 875-86.

Gibbons, A 2008. American Association of Physical Anthropologists meeting. Tuberculosis jumped from humans to cows, not vice versa. *Science,* 320**,** 608.

Gilbert, M T, Cuccui, J, White, W, Lynnerup, N, Titball, R W, Cooper, A & Prentice, M B 2004. Absence of Yersinia pestis-specific DNA in human teeth from five European excavations of putative plague victims. *Microbiology,* 150**,** 341-54.

Gilbert, M T P, Hansen, A J, Willerslev, E, Turner-Walker, G & Collins, M 2006. Insights into the processes behind the contamination of degraded human teeth and bone samples with exogenous sources of DNA. *International Journal of Osteoarchaeology,* 16**,** 156-164.

Gilbert, M T P, Rudbeck, L, Willersleva, E, Hansen, A J, Smith, C, Penkman, K E H, Prangenberg, K, Nielsen-Marsh, C M, Jans, M E, Arthur, P, Nlynnerup, N, Turner-Walker, G, Biddle, M, Kjølbye-Biddle, B & Collins, M J 2005. Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. *J Archaeol Sci,* 32**,** 785-793.

Gimbutas, M 1963. The Indo-Europeans: Archeological Problems *American Anthropologist,* 65**,** 815-836.

Gomes-Filho, I S, Passos Jde, S, Cruz, S S, Vianna, M I, Cerqueira Ede, M, Oliveira, D C, Dos Santos, C A, Coelho, J M, Sampaio, F P, Freitas, C O & De Oliveira, N F 2007. The association between postmenopausal osteoporosis and periodontal disease. *J Periodontol,* 78**,** 1731-40.

Gonzales, F & Sognnaes, R F 1960. Electron microscopy of dental calculus. *Science,* 131**,** 156-8.

Goodman, A H 1989. Dental enamel hypoplasias in prehistoric populations. *Adv Dent Res,* 3**,** 265-71.

Gopher, A, Abbo, S & Lev-Yadun, S 2002. The 'when', the 'where' and the 'why' of the Neolithic revolution in the Levant. *Documenta Praehistorica,* 28**,** 49–62.

Grine, F E, Gwinnett, A J & Oaks, J H 1990. Early hominid dental pathology: interproximal caries in 1.5 million-year-old Paranthropus robustus from Swartkrans. *Arch Oral Biol,* 35**,** 381-6.

Gronenborn, D 1999. A variation on a basic theme: The transition to farming in southern Central Europe *Journal of World Prehistory,* 13**,** 123-210.

Gross, E L, Leys, E J, Gasparovich, S R, Firestone, N D, Schwartzbaum, J A, Janies, D A, Asnani, K & Griffen, A L 2010. Bacterial 16S sequence analysis of severe caries in young permanent teeth. *J Clin Microbiol,* 48**,** 4121-8.

Grossi, S G & Genco, R J 1998. Periodontal disease and diabetes mellitus: a two-way relationship. *Ann Periodontol,* 3**,** 51-61.

Guthrie, R D 2003. Rapid body size decline in Alaskan Pleistocene horses before extinction. *Nature,* 426**,** 169-71.

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 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biol,* 8**,** e1000536.

Haak, W, Brandt, G, De Jong, H N, Meyer, C, Ganslmeier, R, Heyd, V, Hawkesworth, C, Pike, A W, Meller, H & Alt, K W 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**,** 18226-31.

Haak, W, Forster, P, Bramanti, B, Matsumura, S, Brandt, G, Tanzer, 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**,** 1016-8.

Haas, C J, Zink, A, Palfi, G, Szeimies, U & Nerlich, A G 2000. Detection of leprosy in ancient human skeletal remains by molecular identification of Mycobacterium leprae. *Am J Clin Pathol,* 114**,** 428-36.

Haensch, S, Bianucci, R, Signoli, M, Rajerison, M, Schultz, M, Kacki, S, Vermunt, M, Weston, D A, Hurst, D, Achtman, M, Carniel, E & Bramanti, B 2010. Distinct clones of Yersinia pestis caused the black death. *PLoS Pathog,* 6.

Hagelberg, E, Sykes, B & Hedges, R 1989. Ancient bone DNA amplified. *Nature,* 342**,** 485.

Handt, O, Krings, M, Ward, R H & Paabo, S 1996. The retrieval of ancient human DNA sequences. *Am J Hum Genet,* 59**,** 368-76.

Handt, O, Richards, M, Trommsdorff, M, Kilger, C, Simanainen, J, Georgiev, O, Bauer, K, Stone, A, Hedges, R, Schaffner, W & Et Al. 1994. Molecular genetic analyses of the Tyrolean Ice Man. *Science,* 264**,** 1775-8.

Hanihara, T 2008. Morphological variation of major human populations based on nonmetric dental traits. *Am J Phys Anthropol,* 136**,** 169-82.

Harlan, J R & Zohary, D 1966. Distribution of wild wheats and barley. *Science,* 153**,** 1074-80.

Hayden, B 1990. Nimrods, piscators, pluckers and planters: the emergence of food production. *Journal of Anthropological Research,* 9**,** 31–69.

Hedges, R E M, Housley, R A, Bronk, C R & Van Klinken, G J 1992. Radiocarbon dates from the Oxford AMS system: Archaeometry Datelist 15. *Archaeometry,* 34**,** 337–357.

Hermanussen, M 2003. Stature of early Europeans. *Hormones (Athens),* 2**,** 175-8.

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**,** 282-4.

Hillman, G C & Davies, M S 1990. Measured domestication rates in wild wheat and barley under primitive cultivation, and their archaeological implications. *Journal of World Prehistory,* 4**,** 157 - 222

Hillson, S 1996. *Dental anthropology,* Cambridge, Cambridge University Press.

Hillson, S W 1979. Diet and dental disease. *World Archaeol,* 11**,** 147-62.

Hintao, J, Teanpaisan, R, Chongsuvivatwong, V, Ratarasan, C & Dahlen, G 2007. The microbiological profiles of saliva, supragingival and subgingival plaque and dental caries in adults with and without type 2 diabetes mellitus. *Oral Microbiol Immunol,* 22**,** 175-81.

Ho, S Y & Larson, G 2006. Molecular clocks: when times are a-changin'. *Trends Genet,* 22**,** 79-83.

Hoss, M, Jaruga, P, Zastawny, T H, Dizdaroglu, M & Paabo, S 1996. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res,* 24**,** 1304-7.

Jeffcoat, M K, Hauth, J C, Geurs, N C, Reddy, M S, Cliver, S P, Hodgkins, P M & Goldenberg, R L 2003. Periodontal disease and preterm birth: results of a pilot intervention study. *J Periodontol,* 74**,** 1214-8.

Jin, Y & Yip, H K 2002. Supragingival calculus: formation and control. *Crit Rev Oral Biol Med,* 13**,** 426-41.

Jombart, T, Pontier, D & Dufour, A B 2009. Genetic markers in the playground of multivariate analysis. *Heredity,* 102**,** 330-41.

Jorde, L B, Watkins, W S, Bamshad, M J, Dixon, M E, Ricker, C E, Seielstad, M T & Batzer, M A 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet,* 66**,** 979-88.

Kemp, B M & Smith, D G 2005. Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Sci Int,* 154**,** 53-61.

Kenyon, K M 1956. Excavations at Jericho 1956. *Palestine Exploration Quarterly,* 88**,** 67-82.

Kingman, J F C 1982. On the genealogy of large populations *Journal of Applied Probability,* 19**,** 27-43.

Koch, P L, Fogel, M L & Tuross, N 1994. Tracing the diets of fossil animals using stable isotope. *In:* LAJTHA, K. & MICHENER, R. H. (eds.) *Stable isotopes in ecology and environmental science.* Oxford Blackwell Scientic Publications.

Kohler-Rollefson, I, Gillespie, W & Metzger, M 1988. The fauna from Neolithic ‘Ain Ghazal. *In:* N, G. A. & G, G. H. (eds.) *The Prehistory of Jordan, British Archaeology Reports.* Oxford.

Kolenbrander, P E & London, J 1993. Adhere today, here tomorrow: oral bacterial adherence. *J Bacteriol,* 175**,** 3247-52.

Kolenbrander, P E, Palmer, R J, Jr., Periasamy, S & Jakubovics, N S 2010. Oral multispecies biofilm development and the key role of cell-cell distance. *Nat Rev Microbiol,* 8**,** 471-80.

Kolenbrander, P E, Palmer, R J, Jr., Rickard, A H, Jakubovics, N S, Chalmers, N I & Diaz, P I 2006. Bacterial interactions and successions during plaque development. *Periodontol 2000,* 42**,** 47-79.

Kolman, C J, Centurion-Lara, A, Lukehart, S A, Owsley, D W & Tuross, N 1999. Identification of Treponema pallidum subspecies pallidum in a 200-year-old skeletal specimen. *J Infect Dis,* 180**,** 2060-3.

Krings, M, Stone, A, Schmitz, R W, Krainitzki, H, Stoneking, M & Paabo, S 1997. Neandertal DNA sequences and the origin of modern humans. *Cell,* 90**,** 19-30.

Kuijt, I & Goring-Morris, N 2002. Foraging, farming, and social complexity in the pre-pottery Neolithic of the Southern Levant: A review and synthesis *Journal of World Prehistory,* 16**,** 361 - 440

Kumar, P S, Griffen, A L, Moeschberger, M L & Leys, E J 2005. Identification of candidate periodontal pathogens and beneficial species by quantitative 16S clonal analysis. *J Clin Microbiol,* 43**,** 3944-55.

Kutzbach, J E 1981. Monsoon Climate of the Early Holocene: Climate Experiment with the Earth's Orbital Parameters for 9000 Years Ago. *Science,* 214**,** 59-61.

Lander, E S, Linton, L M, Birren, B, Nusbaum, C, Zody, M C, Baldwin, J, Devon, K, Dewar, K, Doyle, M, Fitzhugh, W, Funke, R, Gage, D, Harris, K, Heaford, A, Howland, J, Kann, L, Lehoczky, J, Levine, R, Mcewan, P, Mckernan, K, Meldrim, J, Mesirov, J P, Miranda, C, Morris, W, Naylor, J, Raymond, C, Rosetti, M, Santos, R, Sheridan, A, Sougnez, C, Stange-Thomann, N, Stojanovic, N, Subramanian, A, Wyman, D, Rogers, J, Sulston, J, Ainscough, R, Beck, S, Bentley, D, Burton, J, Clee, C, Carter, N, Coulson, A, Deadman, R, Deloukas, P, Dunham, A, Dunham, I, Durbin, R, French, L, Grafham, D, Gregory, S, Hubbard, T, Humphray, S, Hunt, A, Jones, M, Lloyd, C, Mcmurray, A, Matthews, L, Mercer, S, Milne, S, Mullikin, J C, Mungall, A, Plumb, R, Ross, M, Shownkeen, R, Sims, S, Waterston, R H, Wilson, R K, Hillier, L W, Mcpherson, J D, Marra, M A, Mardis, E R, Fulton, L A, Chinwalla, A T, Pepin, K H, Gish, W R, Chissoe, S L, Wendl, M C, Delehaunty, K D, Miner, T L, Delehaunty, A, Kramer, J B, Cook, L L, Fulton, R S, Johnson, D L, Minx, P J, Clifton, S W, Hawkins, T, Branscomb, E, Predki, P, Richardson, P, Wenning, S, Slezak, T, Doggett, N, Cheng, J F, Olsen, A, Lucas, S, Elkin, C, Uberbacher, E, Frazier, M, et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

Lao, O, Lu, T T, Nothnagel, M, Junge, O, Freitag-Wolf, S, Caliebe, A, Balascakova, M, Bertranpetit, J, Bindoff, L A, Comas, D, Holmlund, G, Kouvatsi, A, Macek, M, Mollet, I, Parson, W, Palo, J, Ploski, R, Sajantila, A, Tagliabracci, A, Gether, U, Werge, T, Rivadeneira, F, Hofman, A, Uitterlinden, A G, Gieger, C, Wichmann, H E, Ruther, A, Schreiber, S, Becker, C, Nurnberg, P, Nelson, M R, Krawczak, M & Kayser, M 2008. Correlation between genetic and geographic structure in Europe. *Curr Biol,* 18**,** 1241-8.

Larsen, C S 1995. Biological changes in human populations with agriculture. *Ann Rev of Anthr,* 24**,** 185-213.

Larson, G, Albarella, U, Dobney, K, Rowley-Conwy, P, Schibler, J, Tresset, A, Vigne, J D, Edwards, C J, Schlumbaum, A, Dinu, A, Balacsescu, A, Dolman, G, Tagliacozzo, A, Manaseryan, N, Miracle, P, Van Wijngaarden-Bakker, L, Masseti, M, Bradley, D G & Cooper, A 2007. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc Natl Acad Sci U S A,* 104**,** 15276-81.

Lazarevic, V, Whiteson, K, Hernandez, D, Francois, P & Schrenzel, J 2010. Study of inter- and intra-individual variations in the salivary microbiota. *BMC Genomics,* 11**,** 523.

Lepp, P W, Brinig, M M, Ouverney, C C, Palm, K, Armitage, G C & Relman, D A 2004. Methanogenic Archaea and human periodontal disease. *Proc Natl Acad Sci U S A,* 101**,** 6176-81.

Ley, R E, Lozupone, C A, Hamady, M, Knight, R & Gordon, J I 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol,* 6**,** 776-88.

Li, X, Kolltveit, K M, Tronstad, L & Olsen, I 2000. Systemic diseases caused by oral infection. *Clin Microbiol Rev,* 13**,** 547-58.

Lieverse, A R 1999. Diet and the aetiology of dental calculus *International Journal of Osteoarchaeology,* 9**,** 219 - 232.

Lilley, J, Stroud, G & Brothwell, D 1994. The Jewish burial ground at Jewbury *In:* ADDYMAN, P. V. & KINSLER, V. A. (eds.) *The Arcaheology of York.* York: Council for British Archaeology.

Lillie, M C & Richards, M 2000. Stable isotope analysis and dental evidence of diet at the Mesolithic–Neolithic transition in Ukraine. *Journal of Archaeological Science,* 27**,** 965-972

Lindahl, T 1993. Instability and decay of the primary structure of DNA. *Nature,* 362**,** 709-15.

Loesche, W 2007. Dental caries and periodontitis: contrasting two infections that have medical implications. *Infect Dis Clin North Am,* 21**,** 471-502, vii.

Logan, E I 2006. Dietary influences on periodontal health in dogs and cats. *Vet Clin North Am Small Anim Pract,* 36**,** 1385-401, ix.

Londo, J P, Chiang, Y C, Hung, K H, Chiang, T Y & Schaal, B A 2006. Phylogeography of Asian wild rice, Oryza rufipogon, reveals multiple independent domestications of cultivated rice, Oryza sativa. *Proc Natl Acad Sci U S A,* 103**,** 9578-83.

Loogvali, E L, Roostalu, U, Malyarchuk, B A, Derenko, M V, Kivisild, T, Metspalu, E, Tambets, K, Reidla, M, Tolk, H V, Parik, J, Pennarun, E, Laos, S, Lunkina, A, Golubenko, M, Barac, L, Pericic, M, Balanovsky, O P, Gusar, V, Khusnutdinova, E K, Stepanov, V, Puzyrev, V, Rudan, P, Balanovska, E V, Grechanina, E, Richard, C, Moisan, J P, Chaventre, A, Anagnou, N P, Pappa, K I, Michalodimitrakis, E N, Claustres, M, Golge, M, Mikerezi, I, Usanga, E & Villems, R 2004. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol,* 21**,** 2012-21.

Macko, S A, Engel, M H, Andrusevich, V, Lubec, G, O'connell, T C & Hedges, R E 1999a. Documenting the diet in ancient human populations through stable isotope analysis of hair. *Philos Trans R Soc Lond B Biol Sci,* 354**,** 65-75; discussion 75-6.

Macko, S A, Lubec, G, Teschler-Nicola, M, Andrusevich, V & Engel, M H 1999b. The Ice Man's diet as reflected by the stable nitrogen and carbon isotopic composition of his hair. *FASEB J,* 13**,** 559-62.

Mahoney, P 2006. Dental microwear from Natufian hunter-gatherers and early Neolithic farmers: comparisons within and between samples. *Am J Phys Anthropol,* 130**,** 308-19.

Malmstrom, H, Gilbert, M T, Thomas, M G, Brandstrom, M, Stora, J, Molnar, P, Andersen, P K, Bendixen, C, Holmlund, G, Gotherstrom, A & Willerslev, E 2009. Ancient DNA Reveals Lack of Continuity between Neolithic Hunter-Gatherers and Contemporary Scandinavians. *Curr Biol.*

Malmstrom, H, Stora, J, Dalen, L, Holmlund, G & Gotherstrom, A 2005. Extensive human DNA contamination in extracts from ancient dog bones and teeth. *Mol Biol Evol,* 22**,** 2040-7.

Malmstrom, H, Svensson, E M, Gilbert, M T, Willerslev, E, Gotherstrom, 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.

Marsh, P D 2003. Plaque as a biofilm: pharmacological principles of drug delivery and action in the sub- and supragingival environment. *Oral Dis,* 9 Suppl 1**,** 16-22.

Marsh, P D 2005. Dental plaque: biological significance of a biofilm and community life-style. *J Clin Periodontol,* 32 Suppl 6**,** 7-15.

Marsh, P D 2010. Microbiology of dental plaque biofilms and their role in oral health and caries. *Dent Clin North Am,* 54**,** 441-54.

Mcdougall, I, Brown, F H & Fleagle, J G 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature,* 433**,** 733-6.

Meiklejohn, C & Zvelebil, M 1994. Health status of European populations at the agricultural transition and the implications for the adoption of farming. *In:* BUSH, H. & ZVELEBIL, M. (eds.) *Health in Past Societies.* Biocultural Interpretations of Human Skeletal Remains in Archaeological Contexts.

Mellars, P 2006. A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature,* 439**,** 931-5.

Menozzi, P, Piazza, A & Cavalli-Sforza, L 1978. Synthetic maps of human gene frequencies in Europeans. *Science,* 201**,** 786-92.

Merrell, D S, Butler, S M, Qadri, F, Dolganov, N A, Alam, A, Cohen, M B, Calderwood, S B, Schoolnik, G K & Camilli, A 2002. Host-induced epidemic spread of the cholera bacterium. *Nature,* 417**,** 642-5.

Miller, E, Ragsdale, B D & Ortner, D J 1996. Accuracy in dry bone diagnosis: A comment on palaeopathological methods *International Journal of Osteoarchaeology* 6**,** 221 - 229

Milton, K 1999. Nutritional characteristics of wild primate foods: do the diets of our closest living relatives have lessons for us? *Nutrition,* 15**,** 488-98.

Molleson, T & Jones, K 1991. Dental evidence for dietary changes at Abu Hureyra. *Journal of Archaeological Science,* 24**,** 455–468.

Montiel, R, Garcia, C, Canadas, M P, Isidro, A, Guijo, J M & Malgosa, A 2003. DNA sequences of Mycobacterium leprae recovered from ancient bones. *FEMS Microbiol Lett,* 226**,** 413-4.

Moss, W J & Griffin, D E 2006. Global measles elimination. *Nat Rev Microbiol,* 4**,** 900-8.

Mullis, K B & Faloona, F A 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol,* 155**,** 335-50.

Munson, M A, Banerjee, A, Watson, T F & Wade, W G 2004. Molecular analysis of the microflora associated with dental caries. *J Clin Microbiol,* 42**,** 3023-9.

Myres, N M, Rootsi, S, Lin, A A, Jarve, M, King, R J, Kutuev, I, Cabrera, V M, Khusnutdinova, E K, Pshenichnov, A, Yunusbayev, B, Balanovsky, O, Balanovska, E, Rudan, P, Baldovic, M, Herrera, R J, Chiaroni, J, Di Cristofaro, J, Villems, R, Kivisild, T & Underhill, P A 2011. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur J Hum Genet,* 19**,** 95-101.

Neumann, K 2003. Anthropology. New Guinea: a cradle of agriculture. *Science,* 301**,** 180-1.

North, D C & Thomas, R P 1977. The first economic revolution. *Economic History Review,* 30**,** 229–241.

Novembre, J & Stephens, M 2008. Interpreting principal component analyses of spatial population genetic variation. *Nat Genet,* 40**,** 646-9.

Nygaard, S E 1989. The Stone Age of Northern Scandinavia: a review. *Journal of World Prehistory,* 3**,** 71-116.

Ohnishi, S, Murata, M & Kawanishi, S 2002. DNA damage induced by hypochlorite and hypobromite with reference to inflammation-associated carcinogenesis. *Cancer Lett,* 178**,** 37-42.

Paabo, S 1985. Molecular cloning of Ancient Egyptian mummy DNA. *Nature,* 314**,** 644-5.

Paabo, S 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci U S A,* 86**,** 1939-43.

Palmer, R J, Jr., Diaz, P I & Kolenbrander, P E 2006. Rapid succession within the Veillonella population of a developing human oral biofilm in situ. *J Bacteriol,* 188**,** 4117-24.

Paster, B J, Boches, S K, Galvin, J L, Ericson, R E, Lau, C N, Levanos, V A, Sahasrabudhe, A & Dewhirst, F E 2001. Bacterial diversity in human subgingival plaque. *J Bacteriol,* 183**,** 3770-83.

Pensaert, M, Ottis, K, Vandeputte, J, Kaplan, M M & Bachmann, P A 1981. Evidence for the natural transmission of influenza A virus from wild ducts to swine and its potential importance for man. *Bull World Health Organ,* 59**,** 75-8.

Perrot, J 1962. Palestine-Syria-Cilicia. *In:* BRAIDWOOD, R. J. & WILLEY, G. R. (eds.) *Courses Toward Urban Life.* Chicago: Aldine Publishing Co.

Petersen, P E, Bourgeois, D, Ogawa, H, Estupinan-Day, S & Ndiaye, C 2005. The global burden of oral diseases and risks to oral health. *Bull World Health Organ,* 83**,** 661-9.

Peyron, O, Guiot, J, Cheddadi, R, Tarasov, P, Reille, M, De Beaulieu, J-L, Bottema, S & Andrieu, V 1998. Climatic reconstruction in Europe for 18,000 Y.B.P. from pollen data. *Quatenary Research,* 49**,** 183-196.

Pinhasi, R, Fort, J & Ammerman, A J 2005. Tracing the origin and spread of agriculture in Europe. *PLoS Biol,* 3**,** e410.

Pope, K O, Pohl, M E, Jones, J G, Lentz, D L, Von Nagy, C, Vega, F J & Quitmyer, I R 2001. Origin and environmental setting of ancient agriculture in the lowlands of Mesoamerica. *Science,* 292**,** 1370-3.

Preus, H R, Marvik, O J, Selvig, K A & Bennike, P 2011. Ancient bacterial DNA (aDNA) in dental calculus from archaeological human remains *Journal of Archaeological Science*.

Price, D T 1991. The Mesolithic of northern Europe. *Annual Review of Anthropology,* 20**,** 211-233.

Price, T D 1987. The Mesolithic of Western Europe. *Journal of World Prehistory,* 1**,** 225-305.

Price, T D, Bentley, R A, Lüning, J, Gronenborn, D & Wahl, J 2001. Prehistoric human migration in the Linearbandkeramik of Central Europe. *Antiquity,* 75**,** 593–603

Price, T D, Knipper, C, Grupe, G & Smrcka, V 2004. Strontium isotopes and prehistoric human migration: The Bell Beaker period in central Europe. *European Journal of Archaeology,* 7**,** 9-40.

Ramakrishnan, U & Hadly, E A 2009. Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Mol Ecol,* 18**,** 1310-30.

Reed, C A 1959. Animal Domestication in the Prehistoric Near East: The origins and history of domestication are beginning to emerge from archeological excavations. *Science,* 130**,** 1629-39.

Renfrew, C 1987. *Archaeology and Language: the Puzzle of Indo-European Origins,* London, Jonathan Cape.

Richards, M, Corte-Real, H, Forster, P, Macaulay, V, Wilkinson-Herbots, H, Demaine, A, Papiha, S, Hedges, R, Bandelt, H J & Sykes, B 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet,* 59**,** 185-203.

Richards, M, Macaulay, V, Hickey, E, Vega, E, Sykes, B, Guida, V, Rengo, C, Sellitto, D, Cruciani, F, Kivisild, T, Villems, R, Thomas, M, Rychkov, S, Rychkov, O, Rychkov, Y, 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 Rienzo, A, Novelletto, A, Oppenheim, A, Norby, S, Al-Zaheri, N, Santachiara-Benerecetti, S, Scozari, R, Torroni, A & Bandelt, H J 2000a. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet,* 67**,** 1251-76.

Richards, M P 2002. A brief review of the archaeological evidence for Palaeolithic and Neolithic subsistence. *Eur J Clin Nutr,* 56**,** 16 p following 1262.

Richards, M P, Pettitt, P B, Stiner, M C & Trinkaus, E 2001. Stable isotope evidence for increasing dietary breadth in the European mid-Upper Paleolithic. *Proc Natl Acad Sci U S A,* 98**,** 6528-32.

Richards, M P, Pettitt, P B, Trinkaus, E, Smith, F H, Paunovic, M & Karavanic, I 2000b. Neanderthal diet at Vindija and Neanderthal predation: the evidence from stable isotopes. *Proc Natl Acad Sci U S A,* 97**,** 7663-6.

Richards, M P, Price, D T & Koch, E 2003a. Mesolithic and Neolithic subsistence in Denmark: new stable isotope data. *Current Anthropology,* 44**,** 288-294.

Richards, M P, Schulting, R J & Hedges, R E 2003b. Archaeology: sharp shift in diet at onset of Neolithic. *Nature,* 425**,** 366.

Richerson, P J, Boyd, R & Bettinger, R L 2001. Was agriculture impossible during the Pleistocene but mandatory during the Holocene? A climate change hypothesis. *American Antiquity,* 66**,** 387-411.

Roberts-Harry, E A & Clerehugh, V 2000. Subgingival calculus: where are we now? A comparative review. *J Dent,* 28**,** 93-102.

Roberts-Harry, E A, Clerehugh, V, Shore, R C, Kirkham, J & Robinson, C 2000. Morphology and elemental composition of subgingival calculus in two ethnic groups. *J Periodontol,* 71**,** 1401-11.

Roberts, C & Ingham, S 2008. Using ancient DNA analysis in palaeopathology: A critical analysis of published papers, with recommendations for future work. *International Journal of Osteoarchaeology,* 18**,** 600–613.

Roberts, C A, Lewis, M E & Manchester, K 2002. The past and present of leprosy. *BAR International Series.* Oxford.

Roostalu, U, Kutuev, I, Loogvali, E L, Metspalu, E, Tambets, K, Reidla, M, Khusnutdinova, E K, Usanga, E, Kivisild, T & Villems, R 2007. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Mol Biol Evol,* 24**,** 436-48.

Rootsi, S, Magri, C, Kivisild, T, Benuzzi, G, Help, H, Bermisheva, M, Kutuev, I, Barac, L, Pericic, M, Balanovsky, O, Pshenichnov, A, Dion, D, Grobei, M, Zhivotovsky, L A, Battaglia, V, Achilli, A, Al-Zahery, N, Parik, J, King, R, Cinnioglu, C, Khusnutdinova, E, Rudan, P, Balanovska, E, Scheffrahn, W, Simonescu, M, Brehm, A, Goncalves, R, Rosa, A, Moisan, J P, Chaventre, A, Ferak, V, Furedi, S, Oefner, P J, Shen, P, Beckman, L, Mikerezi, I, Terzic, 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**,** 128-37.

Rosenberg, N A & Nordborg, M 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat Rev Genet,* 3**,** 380-90.

Rosser, Z H, 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, Corte-Real, H B, De Knijff, P, Decorte, R, Dubrova, Y E, Evgrafov, O, Gilissen, A, Glisic, S, Golge, 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, Norby, S, Pandya, A, Parik, J, Patsalis, P C, Pereira, L, Peterlin, B, Pielberg, G, Prata, M J, Previdere, 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**,** 1526-43.

Rothschild, B M, Martin, L D, Lev, G, Bercovier, H, Bar-Gal, G K, Greenblatt, C, Donoghue, H, Spigelman, M & Brittain, D 2001. Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. *Clin Infect Dis,* 33**,** 305-11.

Saiki, R K, Scharf, S, Faloona, F, Mullis, K B, Horn, G T, Erlich, H A & Arnheim, N 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science,* 230**,** 1350-4.

Salamini, F, Ozkan, H, Brandolini, A, Schafer-Pregl, R & Martin, W 2002. Genetics and geography of wild cereal domestication in the near east. *Nat Rev Genet,* 3**,** 429-41.

Sallares, R a G, S. 2001. Biomolecular archaeology of Malaria. *Ancient Biomolecules,* 3**,** 195-213.

Salo, W L, Aufderheide, A C, Buikstra, J & Holcomb, T A 1994. Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. *Proc Natl Acad Sci U S A,* 91**,** 2091-4.

Sampietro, M L, Lao, O, Caramelli, D, Lari, M, Pou, R, Marti, M, Bertranpetit, J & Lalueza-Fox, C 2007. Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc Biol Sci,* 274**,** 2161-7.

Sauer, C O 1952. *Agricultural origins and their dispersals,* New York, American Geographical Society.

Savard, M, Nesbitt, M & Jones, M K 2006. The role of wild grasses in subsistence and sedentism: new evidence from the northern Fertile Crescent. *World Archaeology,* 38**,** 179 - 196.

Scheie, A A 1994. Mechanisms of dental plaque formation. *Adv Dent Res,* 8**,** 246-53.

Schmidt, C 2001. Dental microwear evidence for a dietary shift between two nonmaize-reliant prehistoric human populations from Indiana. *Am J Phys Anthropol,* 114**,** 139-45.

Schoeninger, M J 1995. Stable isotope studies in human evolution. *Evolutionary Anthropology,* 4**,** 83–98.

Schroeder, H E 1969. Formation and inhibition of dental calculus. *J Periodontol,* 40**,** 643-6.

Semino, O, Passarino, G, Brega, A, Fellous, M & Santachiara-Benerecetti, A S 1996. A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am J Hum Genet,* 59**,** 964-8.

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**,** 1155-9.

Shadel, G S & Clayton, D A 1997. Mitochondrial DNA maintenance in vertebrates. *Annu Rev Biochem,* 66**,** 409-35.

Shapiro, B, Drummond, A J, Rambaut, A, Wilson, M C, Matheus, P E, Sher, A V, Pybus, O G, Gilbert, M T, Barnes, I, Binladen, J, Willerslev, E, Hansen, A J, Baryshnikov, G F, Burns, J A, Davydov, S, Driver, J C, Froese, D G, Harington, C R, Keddie, G, Kosintsev, P, Kunz, M L, Martin, L D, Stephenson, R O, Storer, J, Tedford, R, Zimov, S & Cooper, A 2004. Rise and fall of the Beringian steppe bison. *Science,* 306**,** 1561-5.

Shennan, S & Edinborough, K 2007. Prehistoric population history: from the Late Glacial to the Late Neolithic in Central and Northern Europe. *Journal of Archaeological Science,* 34**,** 1339-1345.

Sidaway, D A 1978. A microbiological study of dental calculus. I. The microbial flora of mature calculus. *J Periodontal Res,* 13**,** 349-59.

Sidi, A D & Ashley, F P 1984. Influence of frequent sugar intakes on experimental gingivitis. *J Periodontol,* 55**,** 419-23.

Slatkin, M & Hudson, R R 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics,* 129**,** 555-62.

Smith, B 2001. The transition to food production. *In:* IN FEINMAN, G. M., AND PRICE, T. D (ed.) *Archaeology at the Millennium: A Sourcebook.* New York: Kluwer Academic/Plenum.

Smith, B D 1998. *The Emergence of Agriculture,* New York, Scientific American Library.

Smith, B H 1984a. Patterns of molar wear in hunger-gatherers and agriculturalists. *Am J Phys Anthropol,* 63**,** 39-56.

Smith, H B 1984b. Patterns of molar wear in hunter–gatherers and agriculturalists. *American Journal of Physical Anthroplogists,* 63**,** 39-56.

Smith, N H, Hewinson, R G, Kremer, K, Brosch, R & Gordon, S V 2009. Myths and misconceptions: the origin and evolution of Mycobacterium tuberculosis. *Nat Rev Microbiol,* 7**,** 537-44.

Smith, V L 1975. The primitive hunter culture, pleistocene extinction, and the rise of agriculture. *The Journal of Political Economy,* 83**,** 727-756

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**,** R174-83.

Soares, P, Ermini, L, Thomson, N, Mormina, M, Rito, T, Rohl, 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**,** 740-59.

Socransky, S S, Haffajee, A D, Cugini, M A, Smith, C & Kent, R L, Jr. 1998. Microbial complexes in subgingival plaque. *J Clin Periodontol,* 25**,** 134-44.

Sokal, R R, Oden, N L & Wilson, C 1991. Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature,* 351**,** 143-5.

Sommer, R & Benecke, N 2004. Late- and Post-Glacial history of the Mustelidae in Europe. *Mammal Review,* 34**,** 249–284.

Sommer, R S & Benecke, N 2005a. Late-Pleistocene and early Holocene history of the canid fauna of Europe (Canidae). *Mammalian Biology,* 70**,** 227–241.

Sommer, R S & Benecke, N 2005b. The recolonization of Europe by brown bears Ursus arctos Linnaeus, 1758 after the Last Glacial Maximum. *Mammal Review,* 35**,** 156–164.

Stekelis, M & Yizraeli, T 1963. Excavations at Nahal Oren (preliminary report). *Israel Exploration Journal* 13**,** 1-12.

Stiner, M C 1993. Small Animal Exploitation and its Relation to Hunting, Scavenging, and Gathering in the Italian Mousterian. *In:* PETERKIN, G. L., BRICKER, H. & MELLARS, P. (eds.) *Hunting and Animal Exploitation in the Later Palaeolithic and Mesolithic of Eurasia.* Washington, DC: Archaeological Papers of the American Anthropological Association.

Stiner, M C 2001. Thirty years on the "broad spectrum revolution" and paleolithic demography. *Proc Natl Acad Sci U S A,* 98**,** 6993-6.

Stiner, M C, Munro, N D, Surovell, T A, Tchernov, E & Bar-Yosef, O 1999. Paleolithic population growth pulses evidenced by small animal exploitation. *Science,* 283**,** 190-4.

Tajima, F 1989. The effect of change in population size on DNA polymorphism. *Genetics,* 123**,** 597-601.

Tauber, H 1981. 13C evidence for dietary habits of prehistoric man in Denmark. *Nature,* 292**,** 332-3.

Taylor, G M, Young, D B & Mays, S A 2005. Genotypic analysis of the earliest known prehistoric case of tuberculosis in Britain. *J Clin Microbiol,* 43**,** 2236-40.

Taylor, L H, Latham, S M & Woolhouse, M E 2001. Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci,* 356**,** 983-9.

Testart, A, Forbis, R G, Hayden, B, Ingold, T, Perlman, S M, Pokotylo, D L, Rowley-Conwy, P & Stuart, D E 1982. The Significance of Food Storage Among Hunter-Gatherers: Residence Patterns, Population Densities, and Social Inequalities. *Current Anthropology,* 23**,** 523-537

Thomas, M G, Weale, M E, Jones, A L, Richards, M, Smith, A, Redhead, N, Torroni, A, Scozzari, R, Gratrix, F, Tarekegn, A, Wilson, J F, Capelli, C, Bradman, N & Goldstein, D B 2002. Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am J Hum Genet,* 70**,** 1411-20.

Torroni, A, Bandelt, H J, D'urbano, L, Lahermo, P, Moral, P, Sellitto, D, Rengo, C, Forster, P, Savontaus, M L, Bonne-Tamir, B & Scozzari, R 1998. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet,* 62**,** 1137-52.

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, Koivumaki, 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**,** 844-52.

Ubelaker, D 1978. *Human Skeletal Remains,* Chicago, Aldine Publishing Co.

Von Hunnius, T E, Yang, D.,, Eng, B., Waye, J. S., and Saunders, S. 2007. Digging deeper into the limits of ancient DNA research on syphilis *Journal of Archaeological Science,* 34**,** 2091-2100.

Wallace, D C 1995. 1994 William Allan Award Address. Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *Am J Hum Genet,* 57**,** 201-23.

Weisdorf, J L 2004. *Stone Age economics: the origins of agriculture and the emergence of non-food specialists,* Mimeo, University of Copenhagen.

Weisdorf, J L 2005. From foraging to farming: Explaining the neolithic revolution. *Journal of Economic Surveys,* 19**,** 561-586.

Willcox, G 1996. Evidence for plant exploitation and vegetation history from three Early Neolithic pre-pottery sites on the Euphrates (Syria) *Vegetation History and Archaeobotany,* 5**,** 143-152.

Woese, C R & Fox, G E 1977. The concept of cellular evolution. *J Mol Evol,* 10**,** 1-6.

Wolfe, N D, Dunavan, C P & Diamond, J 2007. Origins of major human infectious diseases. *Nature,* 447**,** 279-83.

Wood, J W, Milner, G R, Harpending, H C & Weiss, K M 1992. The osteological paradox: Problems of inferring prehistoric health from skeletal samples. *Current Anthropology,* 33**,** 343–370.

Wright, S 1951. The genetical structure of populations. *Annuals of Eugenics,* 15**,** 323-354.

Yokoyama, Y, Lambeck, K, De Deckker, P, Johnston, P & Fifield, L K 2000. Timing of the Last Glacial Maximum from observed sea-level minima. *Nature,* 406**,** 713-6.

Zander, H A, Hazen, S P & Scott, D B 1960. Mineralization of dental calculus. *Proc Soc Exp Biol Med,* 103**,** 257-60.

Zeder, M A 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc Natl Acad Sci U S A,* 105**,** 11597-604.

Zhang, D D, Brecke, P, Lee, H F, He, Y Q & Zhang, J 2007. Global climate change, war, and population decline in recent human history. *Proc Natl Acad Sci U S A,* 104**,** 19214-9.

Zilhao, J 2001. Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proc Natl Acad Sci U S A,* 98**,** 14180-5.

Zilhão, J 1993. The spread of agro-pastoral economies across Mediterranean Europe. *Journal of Medical Archaeology,* 6**,** 5-63.

Zink, A, Haas, C J, Reischl, U, Szeimies, U & Nerlich, A G 2001. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *J Med Microbiol,* 50**,** 355-66.

Zong, Y, Chen, Z, Innes, J B, Chen, C, Wang, Z & Wang, H 2007. Fire and flood management of coastal swamp enabled first rice paddy cultivation in east China. *Nature,* 449**,** 459-62.

Zvelebil, M 1986. Review of Ammerman & Cavalli-Sforza (1984). *Journal of Archaeologucal Sciences,* 13**,** 93-95.

Zvelebil, M 1989. On the transition to farming in Europe, or what was spreading with the Neolithic. *Antiquity,* 63**,** 379–383.

Zvelebil, M & Zvelebil, K V 1988. Agricultural transition and Indo-European dispersals. *Antiquity,* 62**,** 574-583.

CHAPTER TWO

Survival and recovery of DNA from ancient teeth and bones

Adler, C.J., Haak, W., D. Donlon, D., Cooper, A. and The Genographic Consortium (2011) Survival and recovery of DNA from ancient teeth and bones.
Journal of Archaeological Science, v. 38 (5), pp. 956-964, May 2011

NOTE:
This publication is included on pages 63-90 in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.1016/j.jas.2010.11.010

CHAPTER THREE

Ancient DNA reveals that population dynamics during the Late Neolithic and Early Bronze Age substantially influenced the population structure of modern Central Europeans

CHAPTER THREE

Ancient DNA reveals that population dynamics during the Late Neolithic and Early Bronze Age substantially influenced the population structure of modern Central Europeans

3.1 Introduction

The Neolithic period (5500 BC – 2200 BC) in Central Europe (Austria, Germany, Hungary, Poland, Slovakia and Switzerland) was a highly dynamic period of human history (Gronenborn 1999) and population events which occurred during this time are thought to have shaped the current genetic structure of Central Europe (Soares et al. 2010). During the Early Neolithic period (5500 BC – 4000 BC), the hunter-gatherer lifestyle that had been practised by indigenous Europeans since the original settlement of the continent 41-46,000 years ago (Mellars 2006), was largely replaced by an agriculturalist lifestyle (Braidwood et al. 1961). The origins of agriculture in Europe have been the subject of a long-standing running debate particularly whether the development of farming was the result of mass migration (Childe 1925; Clark 1965; Ammerman, Cavalli-Sforza 1984; Zeder 2008) or a gradual cultural exchange (Price 1987; Nygaard 1989; Zvelebil 1989; Price 1991). The development of agriculture in Central Europe has been proposed to have been partly caused by the movement of farming populations of Near Eastern descent from the Carpathian Basin along the Danubian rivers into Central Europe (Childe 1925; Clark 1965; Ammerman, Cavalli-Sforza 1984; Gronenborn 1999; Price et al. 2001). However, the influx of Near Eastern migrants to Europe during the Early Neolithic period has been suggested to have contributed minimally to the population structure of modern-day Europeans (Richards et al. 1996; Torroni et al. 1998; Richards et al. 2000). Founder analysis of modern mitochondrial DNA (mtDNA), which compared the genetic diversity between a source (Near East) and a sink (Europe) population, has indicated that the majority of the European maternal genetic diversity (~80%) is indigenous to the continent, presumably reflecting population events that occurred during the Palaeolithic (Richards et al. 1996; Torroni et al. 1998; Richards et al. 2000), before the advent of agriculture. The settlement events that occurred during the Palaeolithic include the original

peopling of Europe by anatomically modern humans (Mellars 2006), and the repeopling of Europe following the Last Glacial Maximum (LGM) (Adams, Faure 1997; Peyron et al. 1998; Forster 2004). The LGM (~19-22,000 years before present), caused the European climate to become colder and dryer (Yokoyama et al. 2000; Clark et al. 2004), and is thought to have resulted in the retreat of forests to scattered refugia in the Southern European Peninsula (Franco-Cantabrian basin/Iberian Peninsula and the Alps) and the Western Caucasus, where human populations became concentrated (Adams, Faure 1997; Peyron et al. 1998; Forster 2004).

Population changes after the Early Neolithic period have been considered too insubstantial to have influenced the structure of modern day Central Europeans (Richards et al. 2000). For example, founder analysis of modern mtDNA, comparing the genetic diversity of European and Near Eastern populations, has estimated that population events occurring between the Bronze Age and today only contributed between 2 and 5% to the genetic makeup of contemporary Central Europeans (Richards et al. 2000). However, founder analysis may under-represent the role of the most recent migrations (e.g. Palaeolithic versus Bronze Age), as each founder sequence can only represent a single migration in this type of analysis (Richards et al. 2000). Also, the study by Richards and colleagues (2010), modelled the recent (post-Neolithic) migrations in Europe as being from the Near East to Europe, and did not examine the influence of recent population dynamics within Europe on the current genetic makeup of modern Europeans. Hence, currently there is little information about the influence of post Early Neolithic migrations in Europe on the modern population structure.

Archaeological evidence suggests that there were internal multiple migrations in Europe following the Early Neolithic period (Gimbutas, 1963, Price et al., 1994, Price et al., 1998, Price et al., 2004) especially during the Late Neolithic/Early Bronze Age period (3800 - 2000 BC (Anthony, 1990, Soares et al., 2010)). Strontium isotope analysis of skeletal remains from Late Neolithic cultures, such as the Bell Beakers (Price et al., 1994, Grupe et al., 1997) have revealed that individuals were highly mobile. Stable strontium (Sr) isotope ratios of $^{87}Sr/^{86}Sr$ in mineralised tissue have been found to reflect the geology of the area where an individual lived (Price et al., 1994). The strontium isotope ratios in tooth enamel and cortical bone, which form at different ages, reflect the geology where the individual was living as an infant and adult respectively (Price et al., 1994), and differences between these tissues can therefore be used to trace movement of an individual during their lifetime. Strontium isotope analysis of enamel and cortical bone from 69 individuals that were buried in Bell Beaker style graves

from Bavaria revealed that 17-25% of individuals had spent their childhood in a different place to the last 7-10 years of their life (Grupe et al., 1997). The movement of individuals may not reflect the migration of a group, because an archaeological culture, such as the Bell Beakers, may be a 'lifestyle' as opposed to a population. However, there is evidence for population change during the Middle to Late Neolithic from changes in human anatomy and population density. For example population change between the Early Neolithic and Late Neolithic/Early Bronze Age, been indicated by the modification in limb morphology between these periods in Central Europe (Gallagher et al., 2009). An analysis of upper limb bones revealed that the clavicles of Late Neolithic and Early Bronze Age individuals were a similar length to each other, and to Mesolithic hunter-gatherers, however were significantly longer than that observed amongst Early Neolithic, LBK individuals (Gallagher et al., 2009). Archaeological evidence also indirectly suggests that the Late Neolithic/Early Bronze Age was a dynamic period. For example, population density increased during this period (Shennan and Edinborough, 2007) and numerous cultures emerge in the archaeological record (Champion et al., 1984, Whittle, 1996). Suites of characteristics are used to describe European Neolithic cultures, including pottery style, settlement type, subsistence systems, burial practises and grave goods (Gimbutas, 1963, Renfrew, 1987, Bogucki and Grygiel, 1993, Gronenborn, 1999). It is unclear from the archaeological record whether the development of distinct European cultures reflects the *in situ* emergence of regionalised ceramic traditions (Bogucki and Grygiel, 1993) or the expansion and migration of people from surrounding areas. It is important to note that the presence of numerous cultures in Central Europe does not in itself demonstrate population movement. Furthermore, it must be recognised that although multiple cultures may be evident at certain archaeological sites, some or all of these groups could be living there episodically and not contributing to the future generations at that site.

Research has shown that there is a lack of mitochondrial genetic continuity between modern day Central Europeans, and both Palaeolithic hunter-gatherers from the same region (Bramanti et al. 2009) and the first farmers in Central Europe, the Linear Pottery Culture (*Linearbandkeramik,* LBK) (Haak et al. 2010). This indicates that internal migrations within Europe during the Late Neolithic/Early Bronze Age may have substantially influenced the population structure of modern Europeans. Ancient mtDNA retrieved from skeletal remains of the early Neolithic, LBK (Haak et al. 2010) and post-LGM hunter-gatherers (Bramanti et al. 2009), was found to be significantly different ($p < 0.05$) from each other, and also from

modern day Central European populations, based on genetic distance comparisons through analysis of molecular variance (AMOVA). The genetic discontinuity observed between modern Central Europeans and both Early Neolithic and Palaeolithic populations, based on genetic distances, is also observed when haplogroup distributions are examined. The major differences between modern Central European populations, and the Neolithic and Palaeolithic populations examined so far from this area (Haak et al. 2005; Bramanti et al. 2009; Haak et al. 2010), is the low frequency of haplogroup H in the past compared to today. Haplogroup H is now the most frequent mtDNA lineage in Europe, accounting for 45% of Europeans, on average (Achilli et al. 2004; Loogvali et al. 2004; Soares et al. 2010), with the highest frequency observed in Western European populations, and a declining frequency cline observed from the west to east of the continent (Achilli et al. 2004; Loogvali et al. 2004). Haplogroup H was completely absent from the post-LGM, Central (Bramanti et al. 2009) and Northern (Malmstrom et al. 2009) European hunter-gatherers sampled, which instead contained a high frequency of haplogroup U (on average 82% for both studies), primarily sub-haplogroups U4 and U5. Haplogroup H occurred at a frequency of 19% in the LBK remains (Haak et al. 2005; Haak et al. 2010), with the most dominant mtDNA lineage being haplogroup N1a (25%). The genetic discontinuity between prehistoric hunter-gatherers and agriculturists in Central Europe (Bramanti et al. 2009; Haak et al. 2010) suggests that the genetic differences observed between modern day Central Europeans and the Early Neolithic populations may indicate that Early Neolithic farming communities were also not fully incorporated into succeeding populations (Soares et al. 2010).

To assess the likelihood of population continuity from the Early Neolithic until today in Central Europe, ancient mtDNA was recovered from skeletal remains representing five Central European cultures that span the Early Neolithic to the Late Neolithic/Early Bronze Age. These included the LBK (5450–4775 cal BC, n=8), Rössen (4475-4250 cal BC, n=10), Corded Ware (2700-2000 cal BC, n=13), Bell Beaker (2500-2050 cal BC, n=8) and Unetice (2050-1800 cal BC, n=17) cultures (descriptions provided in Table 1). The skeletal samples used in this study had exceptional macroscopic preservation and were collected over the last ten years under ‘DNA-free’ conditions (discussed in Section 3.2 Methods). Analysis was performed on the samples extracted in this study (n=56) and published ancient DNA sequences (n=52) from LBK (Haak et al. 2005; Haak et al. 2010) and Corded Ware culture (Haak et al. 2008) remains, which were also recovered from Central Europe. Bayesian Serial SimCoal analysis were used to examine the relationship among the Early Neolithic, Late

Neolithic/Early Bronze Age cultures from Central Europe and modern Central European populations (Anderson et al. 2005a; Chan et al. 2006).

3.2 Methods

All ancient DNA work was performed at the Australian Centre for Ancient DNA (ACAD) at The University of Adelaide. ACAD is a dedicated, physically isolated laboratory that is approximately a 15 minute walk from the laboratory used for post-PCR analyses. Strict protocols were followed to minimise the amount of modern human DNA in the ancient DNA laboratory. These included the wearing of freshly laundered clothes, a full body suit, shoe covers, boots, facemask, face shield and triple gloving. Furthermore, there is no personnel movement from the post-PCR environment to the ancient DNA laboratory within a single day. All surfaces in the ancient DNA laboratory are routinely wiped with bleach, decon and isopropanol. In addition, the ancient DNA laboratory is irradiated with ultraviolet (UV) light (30–50 nm) overnight for approximately 4 hours, in an attempt to degrade DNA on laboratory surfaces and equipment. All consumables, disposables, tools and instruments are externally bleached and UV irradiated before entering the ancient DNA laboratory and then subjected to routine cleaning before, during and after use.

3.2.1 Sample description and preparation

A description of all samples used in this study, including their geographic location, cultural affinity and anatomical identity, is provided in Table 2. For 56 individuals, the hypervariable region 1 (HVS1) was amplified and independently replicated, using two, anatomically isolated skeletal samples, which were commonly two teeth. Of the 56 individuals for which ancient DNA was extracted and the results independently replicated, 29 of the individuals were extracted by Christina Adler and the remaining 27 were extracted by Dr Wolfgang Haak. We extracted DNA from all skeletal samples in the Australian Centre for Ancient DNA laboratory.

The ancient skeletal samples were decontaminated by exposing the samples to high-intensity UV irradiation for 20 minutes per side (i.e. buccal and lingual for teeth, and anterior and posterior for bones), followed by wiping with bleach (3% (w/v) sodium hypochlorite) and then mechanically removing the outer surface. For teeth, the outer surface was mechanically removed using a dental burr (1000 RPM). The surface layer of bone samples was removed using a disposable fine Dremel carborundum cutting disc (413/36). The decontaminated

skeletal samples were then powdered to a fine grain prior DNA extraction, by either drilling with a dental burr at 1000 RPM, or pulverizing in a bone mill (Sartorius Microdismembrator) at 3300 RPM for 30 seconds. A detailed description of the powdering techniques used can be found in Adler and colleagues (2011) study (Chapter 2).

3.2.2 DNA extraction, PCR, real time PCR and sequencing

We used between 0.05–0.2 g of powdered skeletal material for DNA extraction. A non-template control was included every two to three extractions. The skeletal samples and non-template controls were extracted using a phenol/chloroform/isoamyl alcohol, pH 8.0 (25:24:1) method (Haak et al. 2005; Haak et al. 2008). The DNA was desalted and concentrated using Amicon Ultra-4 filter units (50 kDA, Millipore) with 15 ml of sterile, filtered water. The final extraction volume was 100 µl, and extracts were stored at -20°C.

The mtDNA haplogroup for all skeletal samples analysed in this study was determined by amplification of part of the mtDNA control region (HVS1, nucleotide position 15997-16409), using a suite (between 6 and 8) of overlapping primers (Haak et al. 2005; Haak et al. 2008). The PCR conditions used for all primer sets are as follows; Amplitaq Gold (Applied Biosystems) at 2U in 25µl volumes using 1x Buffer Gold, 2.5mM $MgCl_2$, 0.25mM of each dNTP (Fermentas), 400µM of each primer, 1mg/ml RSA (Sigma-Aldrich) and 2µl of DNA extract. The same thermocycling conditions were used for all the different primer sets, and consisted of an initial enzyme activation at 95°C for 6 minutes, followed by 45 cycles of denaturation at 95°C for 30s, annealing at 58°C for 30s and elongation at 72°C for 30s and a single final extension time at 60°C for 10 minutes. To check the HVS1 determined haplogroup, the mtDNA coding region was also partially amplified. The coding region mtDNA haplogroup was determined through the use of a highly-sensitive, single-base extension (SBE) multiplex PCR, which amplifies 22 mtDNA coding region single nucleotide polymorphisms (SNPs) (Haak et al. 2010). The control and coding region mtDNA haplogroups for all individuals analysed in this study are detailed in Table 2.

Real time PCR (qPCR) was used to assess the molecular behaviour (i.e. quantity and quality) of mtDNA in the ancient DNA extracts. To assess the molecular behaviour of the extracted ancient DNA, the amount of mtDNA at two different fragment lengths (141 and 179 bp, HVS1) was quantified using qPCR. The specificity and efficiency of the primers used for qPCR are presented in Table 3. The specificity of the primers to a single binding site was assessed using a post qPCR melt curve, which enables the dissociation kinetics of the primers

to be visualised. Standard curves for each primer pair (141 and 179 bp) were created from amplicons of the intended target, which were generated from modern human DNA (Adler et al. 2011) (see Chapter 2). All qPCRs were carried out in a 10 µl volume, which contained Express SYBR® Green ER Supermix Universal (Invitrogen), rabbit serum albumin (0.4 mg/ml), forward and reverse primers (1 µM), distilled water and 1 µl of DNA extract, non-template control or positive control. The qPCR thermocycling conditions were as follows; 95 ^{o}C/5 min and 50 cycles of 95 ^{o}C/10 s, 58 ^{o}C/20 s and 72 ^{o}C/15 s. All qPCR's of the ancient DNA extracts were run in triplicate. The qPCR's for the extraction and PCR blanks, and PCR standards (positive control) were run in duplicate. The quantification results for samples extracted by Christina Adler can be found in Table 4. qPCR's were performed on the Rotor-Gene 6000 and analysed with the Rotor-Gene 6000 Series Software 1.7.

All PCRs and qPCR's were visually examined by staining 3.5% agarose TBE gels, which had undergone electrophoresis, with ethidium bromide. All successful amplicon products (i.e. those which produced a visible band on an ethidium bromide stained 3.5% agarose gel) were purified using 5 µl of amplified PCR product, exonuclease I (0.8 U/µl) and shrimp alkaline phosphatase (1 U/µl). The mixture was heated to 37 °C for 40 minutes and then heat inactivated at 80 °C for 10 minutes. The purified amplicons were directly sequenced using a BigDye Terminator 3.1 Kit (Applied Biosystems), according to the manufacturer's instructions. Sequencing products were purified using a Multiscreen$_{HTS}$ Vacuum Manifold (Millipore), according to the manufacturer's protocol. Sequencing products were separated on the 3130xl Genetic Analyzer (Applied Biosystems) and the resulting sequences were edited using Sequencher (version 4.7).

3.2.3 Authentication criteria

Due to the high risk of contamination when dealing with human remains for ancient genetic analysis, precautions were taken to prevent contamination of the ancient samples with modern human DNA. (*i*) The samples were excavated over the last ten years under 'DNA-free' conditions. These conditions included no washing, treating or examining of the samples before genetic analysis. (*ii*) All preparation and analytical steps prior to DNA amplification were conducted in a clean-room area, solely dedicated to ancient human DNA work, located in a physically separated building without any modern DNA work (pre-PCR area). Amplification and sequencing were done in the post-PCR lab. (*iii*) A small number of samples were extracted on a single day (e.g. between five and eight). (*iv*) Contamination was

monitored by running non-template controls in parallel, including extraction and PCR blanks. (*v*) All individuals were sampled twice from anatomically independent regions and treated independently (i.e. extractions performed on separate days). At least eight independent PCR reactions (i.e. a minimum of four overlapping fragments for each of the two extractions) were carried out per individual. (*vi*) All replicable polymorphic sites in combination were consistent with known mtDNA haplogroups, which assisted in ruling out post-mortem damage as a potential source for erroneous sequences. (*vii*) The mtDNA SNP multiplex (Haak et al. 2010) for the coding region was used to check the control region haplogroup assignment. The non-template controls were amplified using the coding region SNP multiplex and the PCR products were sequenced, which provided an ideal monitoring system for contamination of the ancient samples and non-template controls, as the PCR multiplex directly targets SNPs that define potentially contaminating lineages. (*vii*) Degraded DNA is expected to contain a greater amount of shorter fragments of DNA than longer fragments (Malmstrom et al. 2005; Malmstrom et al. 2007). QPCR was used to assess if the ancient DNA exhibited appropriate molecular behaviour for a degraded DNA sample.

3.2.4 Statistical analysis

To determine if the ancient DNA extracted in this study was behaving appropriately for a degraded genetic sample, the relationship between DNA quantity and fragment length (141 and 179 bp) was assessed from the qPCR data by using a Shapiro-Wilk W test. For both fragment lengths, the quantity of DNA in the samples was found to be non-normally distributed ($p < 0.05$) so a non-parametric statistic, the Wilcoxon Signed-Ranks test, was used to compare the amount of DNA recovered at 141 and 179 bp. The Wilcoxon-Signed Ranks test revealed that there was a significantly ($p<0.000$) greater amount of shorter (141 bp) compared to longer (170 bp) mtDNA fragments, in the ancient DNA extracts (Table 5), which is expected of a degraded genetic sample.

Principal Components Analysis (PCA)

PCA is an exploratory form of data reduction, suitable for extracting information from genetic markers, as the analysis attempts to display all variation within the data on a small number of axes, and makes very few assumptions about the data being analysed (Jombart et al. 2009). PCA was used to describe and visualize the maternal genetic relationship amongst the five Neolithic and Bronze Age cultures investigated (LBK, Rössen, Corded Ware, Bell Beaker and Unetice), and compare the prehistoric populations to extant populations, including 41 modern

European and 14 contemporary Near Eastern populations. PCA was performed on the frequency of mtDNA haplogroups in the above described prehistoric and modern populations. The haplogroup frequencies for the extant populations was taken from “MURKA: Mitochondrial DNA database and integrated software”, which currently contains 97,523 HVS-I records from published sources, and is maintained by Oleg Balanovsky and Elena Balanovska at the Russian Academy of Medical Sciences, Moscow. To minimize statistical noise caused by rare haplogroups, PCA was restricted to 19 haplogroups, which had a frequency above 1% in extant European and Near Eastern populations. The haplogroups above 1% included H, HV, J, T, I, N1a, K, V, W, X, U2, U3, U4, U5a, U5b, the sum of African haplogroups (L, M1), the sum of East Eurasian haplogroups (A, B, C, D, F, G, Z), and the sum of all other (rare, under 1%) haplogroups.

In PCA, the frequencies of the above detailed haplogroups for each population were reduced to synthetic principal components, which each represent a degree of variation within the data. The number of principal components extracted was determined by a scree plot, which is a plot of the eigenvalue for each principal component (Table 6). The eigenvalue is a measure of the amount of variation accounted for amongst objects, which is amongst populations in this study, for each component. The point at which the variation accounted for by a component represents noise, as opposed to true structure, is indicated by a sharp decline in the eigenvalue score between two successive components, or drop in the eigenvalue score for a component below one (Jombart, Pontier, Dufour 2009). The first three principal components had eigenvalue scores above one (Table 6), and combined, these principal components accounted for approximately 49% of variation among the ancient and modern populations based on mtDNA haplogroup frequencies. The principal component scores for the first three components for the Neolithic and Bronze Age cultures, and for the modern populations investigated, are presented in Table 7. How the frequency of mtDNA haplogroups affects each component in PCA is determined by the vector or component loadings on the descriptors, which in this study is the haplogroups. The vector loadings for each descriptor (i.e. haplogroup) represent the correlation between the descriptor and each principal component (Jombart et al. 2010). The vector loadings for the 19 haplogroup descriptors used in PCA are presented in Table 8. PCA was performed in SPSS version 17.0.

Population Genetic Analysis

Intra and inter population genetic statistics were calculated for the five prehistoric cultures, including LBK (*n*=49), Rössen (*n*=10), Corded Ware (*n*=24), Bell Beaker (*n*=8) and Unetice (*n*=17), and for extant Central European populations (n=1757). The modern data was obtained from published sources and was restricted to Central European populations, including Austria, Germany, Poland and Switzerland (see Table 9 for details). For statistical analysis, due to the length of the modern sequences included, we trimmed all the ancient sequences to span nucleotide position 16069 to 16365. The intrapopulation statistics (Table 10) included nucleotide diversity (Tajima 1983), number of segregating sites, mean number of pairwise differences (Tajima 1983) and the transition/transversion ratio. The interpopulation statistics (Table 10) calculated included population specific pairwise genetic distances (F_{ST}s) on which the analysis of molecular variance (AMOVA) (Excoffier et al. 1992) is calculated. The genetic distances were estimated using the Kimura-2-parameter model (Kimura 1980), which accounts for differences in the rate of transitions and transversions in the sequence data. AMOVA was used as a hypotheses testing tool, to estimate population differentiation among the Neolithic and Bronze Age cultures, and to compare the past cultures to the amalgamated modern Central European population (Table 11). AMOVA is a hierarchal form of analysis, testing three tiers of genetic difference; between groups, among populations within groups and within all populations (Excoffier, Smouse, Quattro 1992). To compare pre-defined groups by AMOVA, a null distribution is calculated by re-sampling the data, which for this study was performed 1023 times. For the AMOVA, a gamma distribution was assumed, with a shape parameter of 0.22 applied, as estimated from calculations using human mtDNA by Wilder et al. (2004). All population genetic statistics were calculated in Arlequin version 3.1 (Excoffier et al. 2005)

Bayesian Serial SimCoal (BayeSSC) Analysis

The influence of Late Neolithic/Early Bronze Age population dynamics on the genetic structure of modern European populations was assessed by using simulated coalescent analysis to model 14 demographic histories, with the program BayeSSC (Anderson et al. 2005b; Chan, Anderson, Hadly 2006). BayeSSC is appropriate for modelling a demographic history which includes multiple populations that are temporally distributed (Anderson et al. 2005b; Chan, Anderson, Hadly 2006). BayeSSC was used to simulate genetic data according to the coalescent theory, in which offspring randomly choose their parents and mutational

events are non-selective, being governed only by genetic drift (Kingman 1982). Sequence data was generated under a number of scenarios in BayeSSC (described below). The simulated sequence data was used to calculate summary statistics. For this study, the summary statistics calculated were pairwise F_{ST}s between the prehistoric and modern populations, because F_{ST}s are thought to be the most informative genetic statistic to investigate population differentiation (Ramakrishnan, Hadly 2009). To determine the most likely population model, the pairwise F_{ST}s generated from the simulated data were compared to the observed data through the use of Approximate Bayesian Computations (ABC).

A total of five hypotheses, which encompassed 14 separate models, were simulated in BayeSSC, to determine the most likely population structure of the Early Neolithic and Bronze Age cultures. A key aim was to assess the degree of population continuity in Central Europe from the Early Neolithic until today (Table 12, Figure 2). The five hypotheses tested are outlined below.

<u>One-population model (H_o):</u> A single population was present in Central Europe, from the early Neolithic, throughout the Late Neolithic/Early Bronze Age, from which modern Central Europeans have descended. The single population model was used as the null hypothesis.

<u>Five-population model (H_1):</u> Each Neolithic and Bronze Age culture was a separate population and each could be the most likely ancestral group to contemporary Central Europeans.

<u>Three-population model (H_2):</u> The five Neolithic and Bronze Age cultures investigated represent three populations, which are grouped by most likely origin from archaeological information. The three populations include the Early Neolithic cultures (LBK and Rössen), the cultures which are potentially of Kurgan origin (Corded Ware and Unetice), and the single culture that is thought to have a Western European origin, the Bell Beakers. The likelihood that the Early Neolithic, Kurgan or Western cultures were the ancestors of modern Central European was assessed.

<u>Three-population model (H_3):</u> The five Neolithic and Bronze Age cultures were assigned to one of three groups based on temporal intervals. The Early Neolithic cultures (LBK and Rössen) were grouped together, the two contemporaneous Late Neolithic cultures (Corded Ware and Bell Beaker) were banded, and the Bronze Age population (Unetice) was kept

separate. The likelihood that each of the three populations was the ancestor to modern Central Europeans was evaluated.

Two-population model (H_4): The five Neolithic populations were divided into two populations, Early Neolithic (LBK and Rössen) and Late Neolithic/Early Bronze Age cultures (Corded Ware, Bell Beaker and Unetice). The likelihood that each population was the most likely ancestral population to modern Central Europeans was evaluated.

For the 14 models assessed, all populations were restricted to coalescing between 45, 000 and 8, 000 years ago, which spans from when human first colonised Europe and the onset of the Neolithic (8 - 45 KY ago). All populations were modelled as having undergone exponential growth from a small, Palaeolithic population, with an effective population size between 50 and 5000. The rate of exponential growth for each population was determined algebraically, as expanding from a Palaeolithic population to either a Neolithic culture/s or the modern Central European population (depending on the model, see Figure 2). The onset of exponential growth was modelled to begin at some point between the initial colonisation of Europe by humans during the Palaeolithic and the onset of the Neolithic. A uniform distribution was applied to estimate the effective population size for the Palaeolithic population (50 – 5, 000), the Neolithic and Bronze Age cultures (1, 000 – 100, 000), and the modern day Central European population (10, 000 – 12, 000, 000). A mutation rate of 0.0000075 mutations per site per generation was applied (Ho, Endicott 2008). The following steps detail how BayeSSC was performed and the likelihood of each model assessed.

1. Each model was simulated for 50, 000 generations using BayesSSC. To compare the simulated and observed data, five pairwise F_{ST}s were chosen which reflect population differentiation between each of the Neolithic and Bronze Age cultures and the modern Central European population (see Table 10). The simulated and observed F_{ST}s were compared using Approximate Bayesian Computation (ABC) (Beaumont et al. 2002), in which the top 1% of simulations were retained. Posterior distributions for each of the parameters with a prior were assessed. ABC was performed in R version 2.10.1.

2. To optimise the models, maximum credibility procedures were performed. This involved re-running BayesSSC for 50, 000 generations with restricted priors. These restrictions were based on the posterior distributions of each parameter, and centred on parameter values that had the highest likelihood. Again the five chosen F_{ST}s were compared between

the simulated and observed data using ABC, and we retained the top 1% of simulations most similar to the observed data.

3. The likelihood of each model was determined using Akaike's Information Criterion (AIC). As AIC is a maximum likelihood method (Burnham, Anderson 2002a), and the coalescent modelling performed so far was Bayesian, this required all models to be re-run without priors. The models were re-run in BayeSSC for 1, 000 generations without any priors on the parameters, which were instead replaced with the parameter value that gave the highest posterior probability. Again the five F_{ST}s were compared between the simulated and observed data using ABC, and the top 1% of simulations most similar to the observed data accepted.

4. The maximum likelihood of each model was used to assess which model was most probable, penalizing the number of parameters through the use of AIC:

   $$AIC = 2k - 2\text{In}\ (L)$$

   where k is the number of parameters in the model and L is the likelihood. A difference in AIC values of two was taken as a significant difference between models, in accordance with previously outlined criteria (Burnham, Anderson 2002b; Ramakrishnan, Hadly 2009).

3.3 Results

All statistical and population genetic analyses were performed on 276 bp of HVS1. These analyses revealed that among the Early Neolithic and Late Neolithic/Early Bronze Age cultures examined, there was a total of 42 haplogroups (Table 10). The level of nucleotide diversity found in the Early Neolithic and Late Neolithic/Early Bronze Age cultures was comparable to the level of diversity found in the combined modern Central European population (0.0186, Table 10). However, it must be noted that the similarity in nucleotide diversity between the modern and ancient samples may be biased by the small sample sizes examined in some of the ancient cultures, for example with only eight individuals assessed from the Bell Beaker culture. The use of coalescent-based analyses to model the likely diversity of these populations was designed to minimise the effect of any biases due to small sample sizes.

Analysis of ancient mtDNA from Central European Early Neolithic (LBK and Rössen, n=59) and Late Neolithic/Early Bronze Age (Bell Beaker, Corded Ware and Unetice n=49) skeletal remains has shown a degree of genetic discontinuity between these periods, based on their haplogroup distributions (PCA, Figure 1) and genetic distance comparisons (F_{ST}s, Table 10 and AMOVAs, Table 11). Coalescent modelling (Table 12) indicates that Late Neolithic/Early Bronze Age cultures are more likely to be the ancestors of modern Central Europeans than the Early Neolithic cultures suggesting that population changes during the Late Neolithic/Early Bronze Age in Central Europe influenced the genetic makeup of contemporary Central European populations. These results indicate that for some reason the Early Neolithic cultures in Central Europe were not completely incorporated into succeeding generations in the Late Neolithic/Early Bronze Age.

PCA was used to describe the haplogroup distribution of the prehistoric cultures, and explore the similarities or differences among the ancient cultures, and modern European and Near Eastern populations in terms of haplogroup composition (Figure 1). For each ancient and modern population, the frequency of 19 common haplogroups was used for PCA. The first two components of the PCA accounted for 36.6% of variation in the haplogroup data. In the PCA plot (Figure 1a), the extant Near Eastern populations (e.g. Jordan, Syria and Iran) clustered in the lower left quadrant of the plot (low principal component 1 and 2 scores), North Western European populations (e.g. Scotland, Ireland and Finland) clustered in the upper right quadrant (high principal component 1 and 2 scores) and Eastern European populations (e.g. Bosnians, Albanians and Latvians) clustered in the lower right quadrant (high principal component 1 and low principal component 2 scores). The clustering pattern of extant populations in the PCA plot (Figure 1a) has been previously observed in PCA plots of modern European and Near Eastern populations that have been analysed using classical genetic markers (Menozzi et al. 1978; Cavalli-Sforza et al. 1993; Cavalli-Sforza et al. 1994), autosomal genotype SNPs (Lao et al. 2008) and Y-chromosome SNPs (Chikhi et al. 2002). This clustering pattern of the modern European and Near Eastern populations has been described as a horseshoe-type shape and is typical of the ‘null distribution’ observed when PCA is performed on spatially distributed, homogenous populations that have experienced short-range migrations (Novembre, Stephens 2008).

The PCA plot revealed that the Early Neolithic cultures could be differentiated from the Late Neolithic/Early Bronze Age cultures through haplogroup frequencies. The mtDNA haplogroup distribution of the Early Neolithic cultures, LBK and Rössen, included haplogroups N1a, W, K, T and J, as indicated by the component loadings in Figure 1b. The mtDNA haplogroup distribution of the Early Neolithic cultures was most similar to extant Near Eastern populations, such as Turkey and Armenia (Figure 1a), as previously observed in Haak and colleagues (2010) study. However, the haplogroup distribution of both Early Neolithic cultures was outside the range observed for modern European and Near Eastern populations (Figure 1a). In contrast to the Early Neolithic cultures, which all clustered on the left-side of the PCA plot, the Late Neolithic/Early Bronze Age cultures (Corded Ware, Bell Beaker and Unetice) all clustered on the right-side of the PCA plot (Figure 1a). The haplogroup distribution of the Corded Ware and Unetice cultures included mtDNA lineages U, I and T, as indicated by component loadings in the PCA plot (Figure 1b) and clustered most closely to modern Eastern European populations, such as Czechoslovakia and Latvia. However, while the Corded Ware culture clustered within extant populations, the Unetice culture clustered outside the modern population haplogroup distribution (Figure 1a). This resulted from the having a higher frequency of the most common European mtDNA lineage, haplogroup H (15%), compared to the Unetice culture (6%). The Bell Beaker culture also fell within the haplogroup distribution of extant populations (Figure 1a), with a very high frequency of haplogroup H (88%), leading to clustering of this culture with modern Western European populations, such as Sardinia and Spain, which also have a high frequency of haplogroup H (Torroni et al. 1998; Achilli et al. 2004).

Genetic differentiation between the Early Neolithic and Late Neolithic/Early Bronze Age cultures observed in the PCA plot was significantly supported by genetic distance comparisons of the cultures in AMOVA (Table 11). Using the ancient (n=108) and modern Central European mtDNA data (n=1757), a total of 13 population structures were tested using AMOVA to address hypotheses H_1-H_4 (Table 11, Figure 2). AMOVA was used to assess the likelihood that any of the Early Neolithic or Late Neolithic/Early Bronze Age cultures was a continuous population with modern Central Europeans (Figure 2). Also, this analysis was used to test whether the five prehistoric cultures represented; five separate populations (H_1), three populations grouped according to cultural affinity (H_2), three populations banded by temporal spacing (H_3), or two populations split into the Early Neolithic and Late Neolithic/Early Bronze Age periods (H_2) (Figure 2). There was significant ($p<0.01$) support

for all population structures tested where modern Central Europeans were considered to be one population with any of the Late Neolithic/Early Bronze Age cultures (Bell Beakers, Corded Ware or Unetice, H_1-H_4, Table 11). Of these significantly supported models, the greatest amount of variation between groups (3.93%) was observed when the prehistoric cultures were banded into three populations (H_3 - Early Neolithic, Late Neolithic and Early Bronze Age) and modern Central Europeans were considered to be one population with the Late Neolithic cultures, the Bell Beakers and Corded Ware (Table 11). The only prehistoric cultures that did not significantly group with extant Central Europeans were the Early Neolithic cultures (LBK with/without Rössen, Table 11). The genetic difference between the Early Neolithic cultures and modern Central Europeans was observed in AMOVA for all the hypotheses tested; H_1 (p=0.095), H_2 (p=0.123), H_3 (p=0.071) and H_4 (p=0.103) (Table 11). Hypothesis testing using AMOVA was not able to distinguish the number of populations within the five ancient cultures investigated, as there was significant support for the presence of between five and two populations in the dataset.

Simulated coalescent analyses were carried out to determine the most likely demographic history of the Neolithic and modern Central European populations and these analyses demonstrated that the Late Neolithic/Early Bronze Age cultures are more likely to be the ancestors of extant Central Europeans than the Early Neolithic cultures. To test hypotheses H_o - H_4 (Figure 2), 14 demographic models were simulated in BayeSSC. To determine the most likely model, five pairwise F_{ST}s (calculated between the combined modern Central European population and each ancient culture) were generated from the simulated and observed genetic data and compared to each other by ABC. The maximum likelihood of each model was assessed by determining the model's AIC score, with a higher score indicating a less probable model (Burnham, Anderson 2002a). The null hypothesis (H_o), of a single continuous population being present from the Early Neolithic throughout the Late Neolithic/Early Bronze Age until today in Central Europe, was found to be unlikely as it had the second highest AIC value of the 14 models simulated (AIC=61.9, Table 12, Figure 2). The most likely models to explain the observed genetic data were when the extant Central European populations were descendants of the Late Neolithic/Early Bronze Age cultures. The AIC scores (Table 12) for models with different Late Neolithic/Early Bronze Age cultures as ancestors to modern Central Europeans were similar; Bell Beaker alone (H_2, AIC=50.4), Bell Beakers and Corded Ware (H_3, AIC=53.1) or Bell Beaker, Corded Ware and Unetice (H_4, AIC=54.8). As indicated by the AMOVA, the most unlikely models were those where the Early Neolithic

cultures (LBK and Rössen) were a continuous population with contemporary Central Europeans (AIC=58-62, Table 12). Also, as previously found from the AMOVA results, the number of populations present among the five Early Neolithic and Late Neolithic/Early Bronze Age cultures, was unclear. However, there was greater support for between three and two populations being present as opposed to five, with these models having lower AIC scores than models where each prehistoric culture represented a separate population, as under H_1 (Table 12). The coalescent modelling results indicate that the observed maternal genetic composition of the Early Neolithic and Late Neolithic/Early Bronze Age cultures cannot be accounted for by genetic drift over time.

3.4 Discussion

Internal population changes in Europe between the Early Neolithic and Late Neolithic/Early Bronze Age appear to have contributed substantially to the population structure of extant Central Europeans, as all the Late Neolithic/Early Bronze Age cultures examined (Bell Beakers, Corded Ware or Unetice) were more likely to be the ancestors of modern Central Europeans (AIC scores ranging from 50.4 – 54.8, Table 12), than either of the Early Neolithic cultures (LBK and Rössen, AIC scores ranging from 60.1 – 62.6, Table 12). This finding is in contrast to modern population genetic studies which have previously estimated that population dynamics after the Early Neolithic contributed minimally (2-5%) to the population structure of modern Europeans (Richards et al. 2000). However, as previously noted, Richards and colleagues study (2000) assessed the post-Neolithic contribution of migrants from the Near East into Europe, as opposed to investigating the influence of internal migrations on the continent.

Previous genetic studies (Bramanti et al., 2009, Haak et al., 2010) have shown a significant separation between the Early Neolithic and today ($p<0.05$), however these studies were unable to directly identify if the Early Neolithic populations were partially replaced or when this potentially occurred. This study has not only confirmed these previous findings using both coalescent analysis (H_0, Table 12) and AMOVA, but also found that the Early Neolithic populations were not fully incorporated into the populations existing in the Late Neolithic/Early Bronze Age (Table 12). This population change is supported by the observed change in haplogroup distribution (Figure 1) between these periods. For example, the frequency of haplogroup N1a decreases between the Early Neolithic and Late Neolithic/Early

Bronze Age, from 25% among the LBK and 10% among the Rössen culture to 0% in the Late Neolithic/Early Bronze Age cultures (Figure 1, Table 2). The frequency of N1a haplogroup remains extremely rare (0.2%) in modern Central European populations (Haak et al., 2005). The lack of continuity between the Early and Late Neolithic may be because individuals from the LBK were only living in these sites episodically. However, the archaeological evidence for cultural continuity between the LBK and the subsequent Rössen culture (Bogucki and Grygiel, 1993) argues against this.

The distinct haplogroup makeup of the Late Neolithic/Early Bronze Age cultures compared to the Early Neolithic cultures (Figure 1b), may also resolve when Palaeolithic mtDNA lineages, such as haplogroup H and U, expanded or re-expanded into Central Europe. The high frequency of haplogroup H among Late Neolithic/Early Bronze Age cultures (up to 88% among the Bell Beakers compared to an average 20% for LBK and Rössen), may indicate that this mtDNA lineage moved into Central Europe during this period. Molecular clock studies suggest that haplogroup H originated during the LGM (18 thousand years before present) in the Near East (Achilli et al. 2004). The frequency clines of sub-haplogroups H1 and H3, which are highest in frequency in Western Europe and decrease in frequency Eastward across Europe, has been suggested to reflect a population expansion from the Iberian Peninsula during the post-LGM colonisation of Europe (Achilli et al. 2004; Loogvali et al. 2004; Soares et al. 2010). The expansion of haplogroup H in Central Europe may have occurred much later than previous post-LGM estimates, as there was a low frequency of haplogroup H in Central Europe during the post-LGM period and the early Neolithic, as indicated by ancient DNA analysis of post-LGM hunter-gatherer (Bramanti et al. 2009) and LBK remains (Haak et al. 2005; Haak et al. 2010), respectively. The high frequency of haplogroup H among the Late Neolithic culture, the Bell Beakers, suggests that the expansion of this lineage to Central Europe may have occurred after the Early Neolithic.

The high frequency of haplogroup H among individuals from the Bell Beaker culture also lends support to the hypothesis that this lineage at some point in history expanded from the west of Europe across the continent. The Bell Beaker culture was the only Late Neolithic/Early Bronze Age culture investigated which has been proposed to have originated in the west of Europe from archaeological data (Gimbutas 1963), and this culture had by far the highest frequency of haplogroup H and was most similar in haplogroup distribution to modern Western European populations (Figure 1a). The Bell Beakers may have been the vehicle for the expansion of haplogroup H into Central Europe, as they are thought to be a

highly mobile group, from both the analysis of strontium isotope ratios (Price, Grupe, Schroter 1994; Price, Grupe, Schroter 1998; Price et al. 2001; Price et al. 2004) and the widespread distribution of Bell Beaker artefacts from the Iberian Peninsula, to Scotland and across Central Europe (Vander Linden 2007). However, the Bell Beakers were probably not the only culture to have brought haplogroup H from the West into Central Europe, as there may have been multiple expansions of the H lineage from the Iberian Peninsula. To determine why haplogroup H is dominant in Central Europe today, it would be necessary to analyse ancient DNA from Neolithic human remains distributed from the Iberian Peninsula to Central Europe. If haplogroup H did expand from Western Europe during the Neolithic, higher frequencies of the H lineage would be expected earlier in the Neolithic in the West compared to in Central Europe. This proposition has been supported by ancient DNA analysis of middle Neolithic (3500 – 3000 BC) remains from Spain (n=11) that were found to have a high frequency of haplogroup H at 36% (Sampietro et al. 2007).

The presence of haplogroup U amongst the Late Neolithic/Early Bronze Age cultures may also represent the re-expansion of a Palaeolithic lineage into Central Europe. As previously discussed, haplogroup U, in particular subhaplogroup U5, has been suggested to be a marker of Mesolithic hunter-gatherer populations, being found frequently in both Central (Bramanti et al. 2009) and Northern European (Malmstrom et al. 2009), post-LGM hunter-gatherer remains, at 63% and 32%, respectively. Subhaplogroup U5 is thought to have originated in Europe, being found almost exclusively on the continent (Thomas et al. 2002; Achilli et al. 2005; Soares et al. 2010), and is estimated to have diverged sometime between 37 (Soares et al. 2010) and 41 (Achilli et al. 2005) thousand years before present from the analysis of whole mtDNA genomes in a wide range of European populations. While the U5 mtDNA lineage was virtually absent amongst the Early Neolithic, LBK and Rössen cultures (recorded at a frequency of only 1.7%), it was observed in the Late Neolithic/Early Bronze Age cultures, Corded Ware and Unetice, at a frequency of 18% and 29%, respectively (Table 2). The re-emergence of subhaplogroup U5 during the Late Neolithic/Early Bronze Age in Central Europe, which is particularly associated with the Corded Ware and Unetice cultures, may reflect the Mesolithic origins of these cultures. However, determining where the Mesolithic ancestors of these cultures arose is difficult. The Corded Ware and Unetice cultures may be the descendants of Mesolithic hunter-gatherers from Eastern Europe and are hypothesised to be descendants of the Kurgan culture which was present in the Eastern Steppe. The Corded Ware and Unetice cultures display similarities to the Kurgan culture in pottery styles (e.g.

impressed ceramics) (Gimbutas 1963) and burial types (i.e. pit graves) (Bogucki, Grygiel 1993; Gallagher, Gunther, Bruchhaus 2009). It is difficult to directly associate the proposed archaeological origins of the prehistoric cultures with the genetic findings in the absence of accompanying ancient genetic data from the proposed regions of origin of the cultures.

3.5 Conclusion

Population dynamics during the Late Neolithic/Early Bronze Age appear to have substantially influenced the maternal population genetic structure of modern Central Europeans. This has been demonstrated by analysing ancient mtDNA from Late Neolithic/Early Bronze Age cultures. Coalescent analyses of this data has revealed that population continuity from the Early Neolithic until today in Central Europe is unlikely. Furthermore, during the Late Neolithic and Early Bronze Age period, Palaeolithic mtDNA haplogroups, such as haplogroups H and U, which were infrequent in the early Neolithic, again became frequent. The finding that population changes during the Late Neolithic/Early Bronze Age in Central Europe were substantial enough to leave a mark in the modern mitochondrial genetic landscape highlights the complexity of past human movements. It also highlights the inability to use only major past events, such as the introduction of agriculture, the initial peopling of Europe or the re-colonisation of Europe following the LGM, to explain the current population structure of modern Central Europeans.

REFERENCES

Achilli A, Rengo C, Battaglia V, et al. 2005. Saami and Berbers--an unexpected mitochondrial DNA link. Am J Hum Genet 76:883-886.

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

Adams JM, Faure H. 1997. Preliminary vegetation maps of the world since the last glacial maximum: an aid to archaeological understanding. Journal of Arcaeological Sciences 24:623-647.

Adler CJ, Haak W, Donlon D, Cooper A. 2011. Survival and recovery of DNA from ancient teeth and bones. Journal of Archaeological Science 38:956-964

Ammerman AJ, Cavalli-Sforza LL. 1984. The Neolithic Transition and the Genetics of Populations in Europe. Princeton: Princeton University Press.

Anderson CN, Ramakrishnan U, Chan YL, Hadly EA. 2005a. Serial SimCoal: a population genetics model for data from multiple populations and points in time. Bioinformatics 21:1733-1734.

Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA. 2005b. Serial SimCoal: A population genetic model for data from multiple populations and points in time. Bioinformatics 21:1733-1734.

Anthony DW. 1990. Migration in archeology: The baby and the bathwater. American Anthropologist 92:895-914.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025-2035.

Bogucki P, Grygiel R. 1993. The first farmers of Central Europe: A survey article Journal of Field Archaeology 20:399-426.

Braidwood RJ, Howe B, Reed CA. 1961. The Iranian Prehistoric Project: New problems arise as more is learned of the first attempts at food production and settled village life. Science 133:2008-2010.

Bramanti B, Thomas MG, Haak W, et al. 2009. Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. Science.

Burnham KP, Anderson DR. 2002a. Model Selection and Inference: A Practical Information - Theoretic Approach. New York: Springer-Verlag.

Burnham KP, Anderson DR. 2002b. Model Selection and Inference: A Practical Information — Theoretic Approach. New York.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1993. Demic expansions and human evolution. Science 259:639-646.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton: Princeton University Press.

Champion T, Gamble C, Shennan S, Whittle A. 1984. Prehistoric Europe. London: Academic Press.

Chan YL, Anderson CN, Hadly EA. 2006. Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. PLoS Genet 2:e59.

Chikhi L, Nichols RA, Barbujani G, Beaumont MA. 2002. Y genetic data support the Neolithic demic diffusion model. Proc Natl Acad Sci U S A 99:11008-11013.

Childe VG. 1925. The dawn of European civilisation. London: Kegan Paul.

Clark G. 1965. Radiocarbon dating and the spread of farming economy. Antiquity 39:45-48.

Clark PU, McCabe AM, Mix AC, Weaver AJ. 2004. Rapid rise of sea level 19,000 years ago and its global implications. Science 304:1141-1144.

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.
Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.
Fokkens H, Achterkamp Y, Kuijpers M. 2008. Bracers or bracelets? About the functionality and meaning of Bell Beaker wrist-guards. Proceedings of the Prehistoric Society 74:109-140.
Forster P. 2004. Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. Philos Trans R Soc Lond B Biol Sci 359:255-264; discussion 264.
Gallagher A, Gunther MM, Bruchhaus H. 2009. Population continuity, demic diffusion and Neolithic origins in central-southern Germany: the evidence from body proportions. Homo 60:95-126.
Gimbutas M. 1963. The Indo-Europeans: Archeological Problems American Anthropologist 65:815-836.
Gronenborn D. 1999. A variation on a basic theme: The transition to farming in southern Central Europe Journal of World Prehistory 13:123-210.
Grupe G, Price DT, Schröter P, Söllner F, Johnson CM, Beard BL. 1997. Mobility of Bell Beaker people revealed by strontium isotope ratios of tooth and bone: a study of southern Bavarian skeletal remains. Applied Geochemistry 12:517-525.
Haak W, Balanovsky O, Sanchez JJ, et al. 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. PLoS Biol 8:e1000536.
Haak W, Brandt G, de Jong HN, 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 105:18226-18231.
Haak W, Forster P, Bramanti B, et al. 2005. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. Science 310:1016-1018.
Ho SY, 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.
Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94.
Jombart T, Pontier D, Dufour AB. 2009. Genetic markers in the playground of multivariate analysis. Heredity 102:330-341.
Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111-120.
Kingman JFC. 1982. On the genealogy of large populations Journal of Applied Probability 19:27-43.
Lao O, Lu TT, Nothnagel M, et al. 2008. Correlation between genetic and geographic structure in Europe. Curr Biol 18:1241-1248.
Loogvali EL, Roostalu U, Malyarchuk BA, et al. 2004. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. Mol Biol Evol 21:2012-2021.
Malmstrom H, Gilbert MT, Thomas MG, et al. 2009. Ancient DNA Reveals Lack of Continuity between Neolithic Hunter-Gatherers and Contemporary Scandinavians. Curr Biol.
Malmstrom H, Stora J, Dalen L, Holmlund G, Gotherstrom A. 2005. Extensive human DNA contamination in extracts from ancient dog bones and teeth. Mol Biol Evol 22:2040-2047.
Malmstrom H, Svensson EM, Gilbert MT, Willerslev E, Gotherstrom 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.

Mellars P. 2006. A new radiocarbon revolution and the dispersal of modern humans in Eurasia. Nature 439:931-935.
Menozzi P, Piazza A, Cavalli-Sforza L. 1978. Synthetic maps of human gene frequencies in Europeans. Science 201:786-792.
Novembre J, Stephens M. 2008. Interpreting principal component analyses of spatial population genetic variation. Nat Genet 40:646-649.
Nygaard SE. 1989. The Stone Age of Northern Scandinavia: a review. Journal of World Prehistory 3:71-116.
Peyron O, Guiot J, Cheddadi R, Tarasov P, Reille M, de Beaulieu J-L, Bottema S, Andrieu V. 1998. Climatic reconstruction in Europe for 18,000 Y.B.P. from pollen data. Quatenary Research 49:183-196.
Piontek J, Marciniak A. 1992. Anthropological structure and culture adaptive strategies of the Neolithic Populations in Central Europe. International Journal of Anthropology 7:71-86.
Price DT. 1991. The Mesolithic of northern Europe. Annual Review of Anthropology 20:211-233.
Price TD. 1987. The Mesolithic of Western Europe. Journal of World Prehistory 1:225-305.
Price TD, Bentley RA, Lüning J, Gronenborn D, Wahl J. 2001. Prehistoric human migration in the Linearbandkeramik of Central Europe. Antiquity 75:593–603
Price TD, Grupe G, Schroter P. 1994. Reconstruction of migration patterns in the Bell Beaker period by stable strontium isotope analysis. Applied Geochemistry 9:413–417.
Price TD, Grupe G, Schroter P. 1998. Migration and mobility in the Bell Beaker period in central Europe. Antiquity 72:405–411.
Price TD, Knipper C, Grupe G, Smrcka V. 2004. Strontium isotopes and prehistoric human migration: The Bell Beaker period in central Europe. European Journal of Archaeology 7:9-40.
Ramakrishnan U, Hadly EA. 2009. Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. Mol Ecol 18:1310-1330.
Renfrew C. 1987. Archaeology and Language: the Puzzle of Indo-European Origins. London: Jonathan Cape.
Richards M, Corte-Real H, Forster P, et al. 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59:185-203.
Richards M, Macaulay V, Hickey E, et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251-1276.
Sampietro ML, Lao O, Caramelli D, Lari M, Pou R, Marti M, Bertranpetit J, Lalueza-Fox C. 2007. Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. Proc Biol Sci 274:2161-2167.
Shennan S, Edinborough K. 2007. Prehistoric population history: from the Late Glacial to the Late Neolithic in Central and Northern Europe Journal of Archaeological Science 34:1339-1345
Sherratt A. 1982. The development of neolithic and copper age settlement in the great hungarian plain part 1: The regional setting. Oxford Journal of Archaeology 3:287-316.
Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt HJ, Torroni A, Richards MB. 2010. The archaeogenetics of Europe. Curr Biol 20:R174-183.
Starling NJ. 1983. Neolithic settlement patterns in Central Germany. Oxford Journal of Archaeology 2:1-11.
Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437-460.

Thomas MG, Weale ME, Jones AL, et al. 2002. Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. Am J Hum Genet 70:1411-1420.

Torroni A, Bandelt HJ, D'Urbano L, et al. 1998. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62:1137-1152.

Vander Linden M. 2007. What linked the Bell Beakers in third millennium BC Europe? . Antiquity 81:343–352.

Whittle A. 1996. Europe in the Neolithic. Cambridge: Cambridge University Press.

Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF. 2004. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. Nat Genet 36:1122-1125.

Yokoyama Y, Lambeck K, De Deckker P, Johnston P, Fifield LK. 2000. Timing of the Last Glacial Maximum from observed sea-level minima. Nature 406:713-716.

Zeder MA. 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. Proc Natl Acad Sci U S A 105:11597-11604.

Zvelebil M. 1989. On the transition to farming in Europe, or what was spreading with the Neolithic. Antiquity 63:379–383.

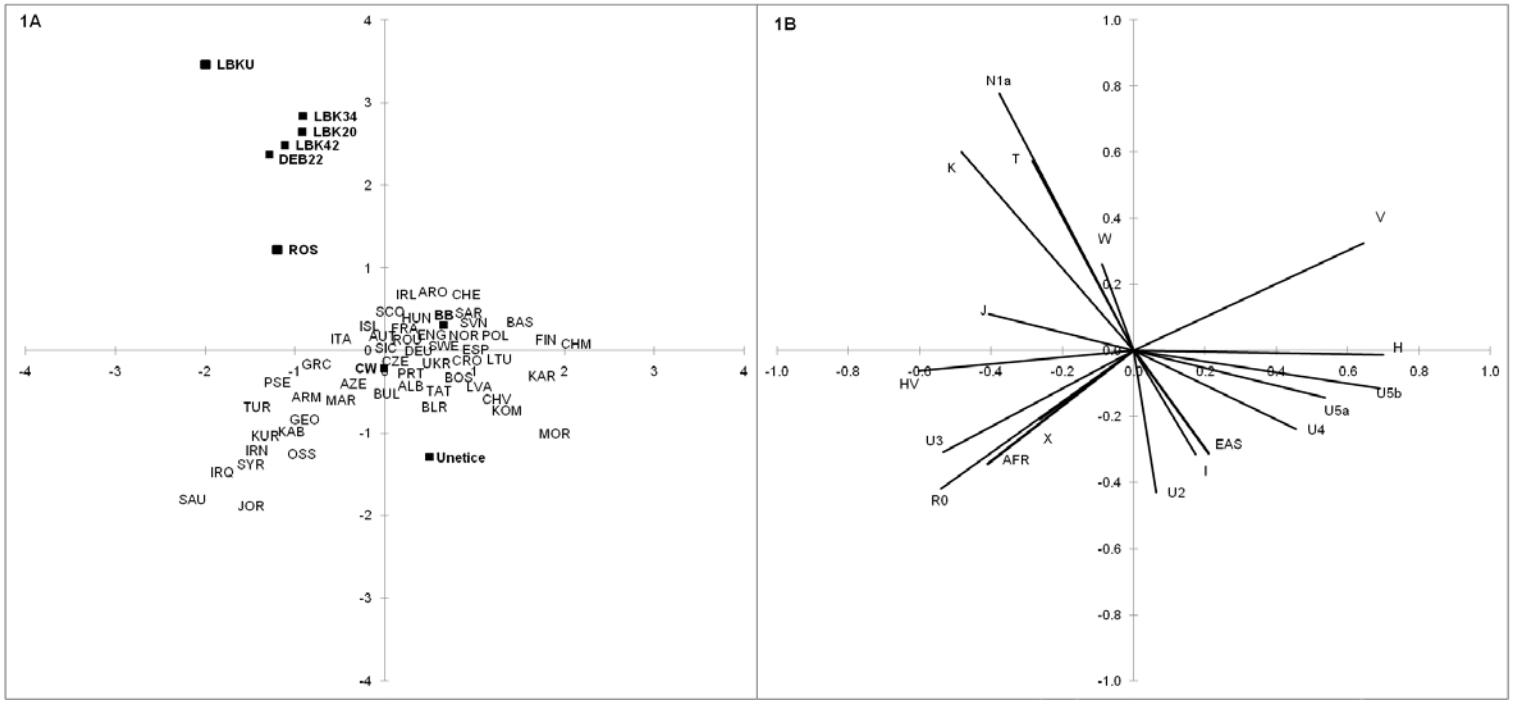

Figure 1. A) Principal components analysis (PCA) for the five and Bronze Age cultures investigated (highlighted by square makers), 41 modern European populations and 14 Near Eastern populations. PCA was performed on the populations, using haplogroup frequencies as descriptors. The haplogroups included 19 common lineages found in Europe and the Near East, in addition to the sum of African haplogroups and East Eurasian haplogroups, and the sum of all other rare lineages. B) The component loadings (vector loadings) indicate which haplogroups are causing populations to cluster in the adjacent Figure 1A. The PCA accounted for 36.6% of variation in the data, 21.7% on Principal Component 1 (x-axis) and 14.9% on Principal Component 2 (y-axis). Country codes can be found in Table 9.

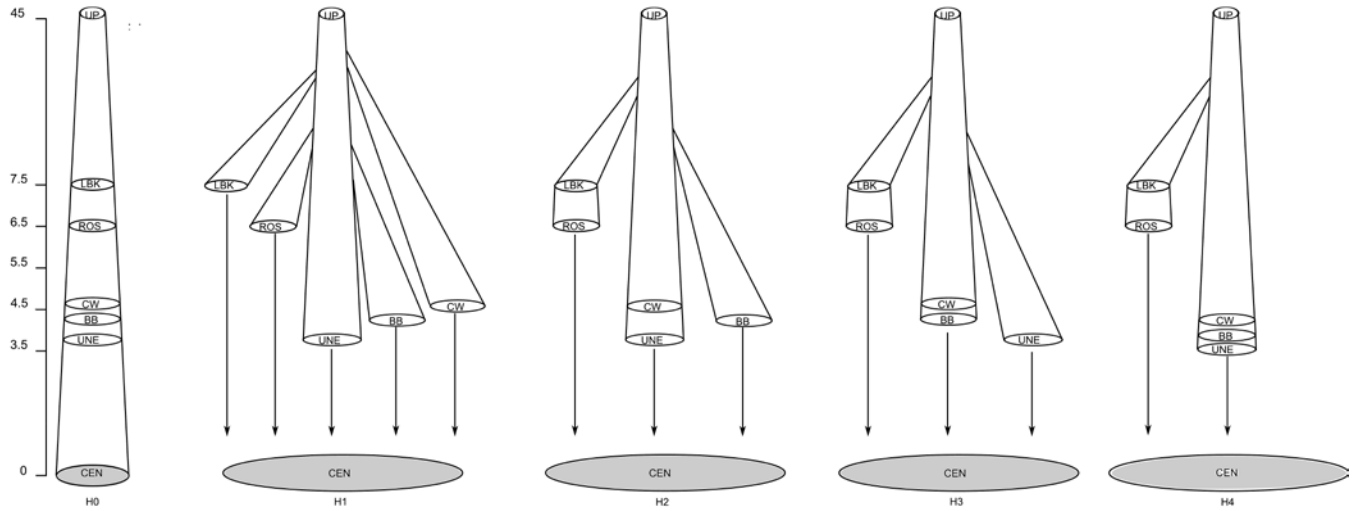

Figure 2. Diagrammatic representation of the five hypotheses simulated in BayeSSC, to investigate the influence of population dynamics during the Late Neolithic/Early Bronze Age on the population structure of modern Europeans. The null hypothesis (H_0) was that a single, continuous population has existed from the earliest Neolithic culture (LBK) until today. The opposite was also investigated, with a five-population model (H_1), where each Neolithic culture was considered a separate population. Two three-population models were tested. The first (H_2) was based on archaeological evidence and grouped the early farmers and cultures of potential Kurgan origin. A second, three population model (H_3) was tested, in which cultures were grouped by temporal intervals, into the early, middle to late Neolithic and late to Bronze Age. The two population model (H_4) investigated, separated cultures into Early Neolithic and Late Neolithic/Early Bronze Age groups. For all multi-population models (H_1, H_2, H_3, H_4), the likelihood of a population being the closest ancestral group to contemporary Central Europeans (CEN) was evaluated. Additionally for the multi-population hypotheses, all populations were modelled to coalesce at some point between the Upper Palaeolithic (UP, 45 000 years) and the onset of the Neolithic (7 500 years) in Central Europe. The y-axis is presented in thousands of years. Note the size of the circles, which represent the groups, is not to scale for effective population size; they are merely to represent an increase in effective population size from a small Palaeolithic population to a large effective population size today, which has increased exponentially.

Culture	Period (BC)	Distribution	Potential ancestral population/location	Pottery style	Burial practise	Grave goods	Reference
Linear Pottery Culture (LBK)	5450–4775	Central Europe	Starcevo-Koros-Cris complex and Vinca Culture (Carpathian Basin)	Broad incised lines on vessels, commonly conical bowls	Inhumation, orientation NNW-SSE and flexed limbs, rare occurrences of cremation	Grinding stones, pottery, polished adzes, objects made of spondylus (armlets, buckles and beads), ochre deposits and lithic artefacts	(Piontek, Marciniak 1992; Bogucki, Grygiel 1993; Gronenborn 1999)
Rössen	4625–4250	Germany and Eastern France	LBK	Stab and drag, punctuate ornament, with a common 'hanging chevron' motif	Inhumation, NNW-SSE and flexed limbs, some are in an extended position	Similar to the LBK, becoming more diversified	(Starling 1983; Bogucki, Grygiel 1993)
Corded Ware (CW)	2700-2000	Central and North Eastern Europe	Kurgan Culture (Eastern Steppe)	Cord impressed pottery and battle axes	Single burials, on side males facing W-E and females E-W and flexed limbs.	Stone axes, ceramics, sheep and goat bone	(Gimbutas 1963; Piontek, Marciniak 1992),
Bell Beaker (BB)	2500–2050	Western and Central Europe (Pan-Europe)	Iberian Peninsula	Inverted bell shaped ceramics	Single burials	Flint arrowheads, slate wrist guards, ornaments of jet, gold, amber, obsidian, human and horse bone	(Sherratt 1982; Price, Grupe, Schroter 1994; Vander Linden 2007; Fokkens et al. 2008)
Unetice	2050-1800	Central and Eastern Europe	Kurgan Culture	Stamped or cord impressed pottery	Pit graves, skeletons lying on the back with legs contracted upwards	Horse shaped figurines, stone maceheads, battle axes, pottery and ochre deposits	(Gimbutas 1963; Sherratt 1982; Starling 1983)

Table 1. Summary of the archaeological and anthropological information for the five Neolithic and Bronze Age cultures investigated in this study.

Individual	ACAD#	Skeletal element	Grave	Age/Sex	Location	Culture	HVS I Haplotype	Hg HVSI	Hg Geno CoRe22
UWS 3	4343/4344	C43/I41	5	18-25/M	Oberwiederstedt 1, Unterwiederstedt	LBK	T16224C, T16311C	K	K
UWS 8b	4345/4346	I41/m74	11	8-10/U	Oberwiederstedt 1, Unterwiederstedt	LBK	C16069T, T16126C	J	J
UWS 5.2	4347/4348	P24/M38	7	Adult/F	Oberwiederstedt 1, Unterwiederstedt	LBK	A16051G, T16092C, G16129A, C16147A, T16154C, T16172C, C16223T, C16248T, C16320T, C16355T	N1a	N1
UWS 6	4349/4350	m65/I11	9	9-12/U	Oberwiederstedt 1, Unterwiederstedt	LBK	G16129A, C16147A, T16154C, T16172C, C16223T, C16248T, C16320T, C16355T	N1a	N1
UWS 7	4351/4352	c83/m74	10	7-9/U	Oberwiederstedt 1, Unterwiederstedt	LBK	T16126C, C16147T, C16294T, C16296T, T16297C, T16304C	T2	T
UWS 11	4353/4354	c73/c63	13	3-4/U	Oberwiederstedt 1, Unterwiederstedt	LBK	T16126C, T16189C, C16294T, C16296T	T	T
UWS 4	4355/4356	Petrous temporal/C43	6	8-11/U	Oberwiederstedt 1, Unterwiederstedt	LBK	C16069T, T16126C	J	J
OSH 1	4359/4360	C33/C13	225	25-35/M	Oberwiederstedt 3, Schrammhöhe	Rossen	rCRS	H	H
OSH 2	4361/4362	C13/M46	195	30-38/F	Oberwiederstedt 3, Schrammhöhe	Rossen	rCRS	H	H
OSH 3	4363/4364	C13/M47	206	35-45/M	Oberwiederstedt 3, Schrammhöhe	Rossen	rCRS	H	H
OSH 5	4365/4366	C33/C13	130	12-14/U	Oberwiederstedt 3, Schrammhöhe	Rossen	C16179T, T16189C, C16223T, G16255A, C16278T, T16297C, T16362C	X2	X
OSH 6	4367/4368	Petrous temporal/M26	220	Infant/U	Oberwiederstedt 3, Schrammhöhe	Rossen	T16224C, T16311C	K	K
OSH 7	4369/4370	P35/P44	95	38-48/M	Oberwiederstedt 3, Schrammhöhe	Rossen	T16304C	H5	H

Table 2. Anatomical, archaeological and genetic information for all samples; HVS1 (np 15997-16409), male (M), female (F) and unknown (U).

Individual	ACAD#	Skeletal element	Grave	Age/Sex	Location	Culture	HVS I Haplotype	Hg HVSI	Hg Geno CoRe22
OSH 8	4371/4372	M28/C13	90	38-47/F	Oberwiederstedt 3, Schrammhöhe	Rossen	T16126C, T16189C, C16294T, C16296T	T2	T
OSH 10	4357/4358	M16/M55	64	7-8/M?	Oberwiederstedt 3, Schrammhöhe	Rossen	T16298C	HV	HV
ESP 13	4403/4404	M27/M17	4229	17-20/F	Esperstedt	Rossen	16126C, 16153A, 16294T, 16296T	T2e	T
OAO 1	4298/4299	M37?/M36	3	30-40/U	Oberwiederstedt, Arschkerbe Ost	Rossen	C16147A, T16172C, 16223T, C16248T, C16320T, C16355T	N1a	N1
QUEXII 1	4325/4326	M36/P35	1266	20-24/F	Quedlinburg XII	Corded Ware	rCRS	H	H
ALB 3	4375/4376	M16/M37	7144	50/F	Alberstedt	Corded Ware	16311C	HV	HV
ESP 5	4383/4384	m65/I41	5052	8/U	Esperstedt	Corded Ware	16192T, 16256T, 16270T	U5a	U
ESP 11	4399/4400	P34/M36	6216	25-35/M	Esperstedt	Corded Ware	16356C, 16362C	U4	U
ESP 14	4405/4406	P45/I21	6141	19-21/M	Esperstedt	Corded Ware	16224C, 16311C	K	K
ESP 15	4407/4408	P15/M36	6	30-40/M	Esperstedt	Corded Ware	16362C	H	H
ESP 16	4409/4410	C43/C23	6236	18-22/F?	Esperstedt	Corded Ware	16192T, 16223T, 16292T, 16325C	W6	W
ESP 17	4411/4412	M46/M47	4098	35-55/M	Esperstedt	Corded Ware	16256T, 16270T, 16399G	U5a1	U
ESP 19	4415/4416	M16/M26	5082	8/U	Esperstedt	Corded Ware	16192T, 16256T, 16270T	U5a	U
ESP 20	4417/4418	m75/m85	2200	7-8/M?	Esperstedt	Corded Ware	16069T, 16126C	J	J
ESP 22	4423/4424	M16/M36	6140	12-16/U	Esperstedt	Corded Ware	16189C, 16223T, 16278T	X	X/N
ESP 25	4429/4430	M18/M26	4179	15-18/U	Esperstedt	Corded Ware	16069T, 16126C	J	J
ESP 26	4431/4432	M36/C33	6233	40-60/F	Esperstedt	Corded Ware	16126C, 16294T, 16324C	T2	T
QUEXII 2	4327/4328	R. Tibia/L. Tibia	1265	19-22/F	Quedlinburg XII	Bell Beaker	rCRS	H	H
QUEXII 3	4329/4330	M36?/C23	6256	19-22/M?	Quedlinburg XII	Bell Beaker	rCRS	H	H
QUEXII 4	4331/4332	C23/C33	6255.1	55-65/F	Quedlinburg XII	Bell Beaker	C16069T, T16126C	J	J
QUEXII 6	4335/4336	L. Humerus/ R. Tibia	6256	19-22/M?	Quedlinburg XII	Bell Beaker	rCRS	H	H

Table 2, Continued. Anatomical, archaeological and genetic information for all samples; HVS1 (np 15997-16409), male (M), female (F) and unknown (U).

Individual	ACAD#	Skeletal element	Grave	Age/Sex	Location	Culture	HVS I Haplotype	Hg HVSI	Hg Geno CoRe22
ROT 6	4387/4388	M(27/26?)/M28	10044	45-60/F	Rothenschirmbach	Bell Beaker	16304C	H5	H
ALB 2	4444/4445	M46/C33	7137	17-20/M	Alberstedt	Bell Beaker	rCRS	H	H
ROT 1	4446/4447	C23/C33	10294	12-15/U	Rothenschirmbach	Bell Beaker	16256T	H3	H
ROT 2	4448/4449	P45/P44	10293	20-24/F?	Rothenschirmbach	Bell Beaker	16304C	H5	H
QUEVIII 1	4302/4303	M37/C33	3644	35-45/F?	Quedlinburg VIII	Unetice	T16189C, C16256T, C16270T, T16311C	U5a1a	U
QUEVIII 2	4304/4305	P34?/I21	3650	28-34/F	Quedlinburg VIII	Unetice	A16051G, T16092C, G16129c, A16183c, T16189C, T16362C	U2	U
QUEVIII 3	4306/4307	M26/M37	3647	28-38/M	Quedlinburg VIII	Unetice	T16126C, A16163G, C16186T, T16189C, C16294T	T1	T
QUEVIII 4	4308/4309	C13/M26	3646	25-35/F	Quedlinburg VIII	Unetice	G16213a	H7a	H
QUEVIII 5	4310/4311	M38/P34	3648	40-50/M?	Quedlinburg VIII	Unetice	C16256T, C16270T, a16399G	U5a1a	U
QUEVIII 6	4312/4313	C43/M48	3580	40-60/F	Quedlinburg VIII	Unetice	T16189C	U	U
QUEXIV 1	4337/4338	C12/I12	31043	18-22/F	Quedlinburg XIV	Unetice	T16126C, C16294T, C16296T	T2	T
QUEXII 7	4339/4340	M17/C23	7063	20-24/F	Quedlinburg XII	Unetice	C16192T, C16222T, C16256T, C16270T, a16399G	U5a1a	U
ESP 2	4377/4378	C23/C13	3340	U/M?	Esperstedt	Unetice	16086C, 16129A, 16223T, 16391A	I	I
ESP 3	4379/4380	C23/M16	1631	35-45/F?	Esperstedt	Unetice	16192T, 16256T, 16270T, 16399G	U5a1	U
ESP 4	4381/4382	M48/C43	3323	40-50/M	Esperstedt	Unetice	16147G, 16223T, 16292T	W	W
ESP 6	4385/4386	M16/C33	5	25-35/F	Esperstedt	Unetice	16129A, 16223T, 16278T, 16311C, 16391A	I1	I
ESP 9	4393/4394	C13/P15	6593	30-35/M	Esperstedt	Unetice	16189C, 16192T, 16270T	U5b	U
ESP 12	4401/4402	C13/C23	3310	26-30/F	Esperstedt	Unetice	16189C, 16223T, 16278T	X	X/N
ESP 18	4413/4414	P45/C33	1580/1579	50-60/M	Esperstedt	Unetice	16126C, 16294T, 16296T, 16304C	T2b	T
ESP 27	4433/4434	M38/M47	3326/3327	40-50/F	Esperstedt	Unetice	16126C, 16292T, 16294T	T2	T
ESP 29	4437/4438	L. Pars petrosa/m74	3332/3333	9/U	Esperstedt	Unetice	16086C, 16129A, 16223T, 16391A	I	I

Table 2, Continued. Anatomical, archaeological and genetic information for all samples; HVS1 (np 15997-16409), male (M), female (F) and unknown (U).

Primer name	Fragment length (bp)	Primer Sequence 5'-3'	T_M (^oC)	R^2	Efficiency (%)
L16117 H16218	141	TACATTACTGCCAGCCACCAT TGTGTGATAGTTGAGGGTTG	80	0.98	98
L16209 H16348	179	CCCCATGCTTACAAGCAAGT ATGGGGACGAGAAGGGATTTG	82	0.99	106

Table 3. QPCR primers; HVS1 position, fragment length, primer sequence, T_M (dissociation temperature), R^2 (coefficient of determination) and efficiency.

Individual	Average 141bp mtDNA copies/uL	Standard deviation 141bp mtDNA copies/uL	Average 179bp mtDNA copies/uL	Standard deviation 179bp mtDNA copies/uL
UWS 3	8975	5692	362	279
UWS 8b	37600	9051	212	141
UWS 5.2	16450	4313	92	18
UWS 6	89850	63852	66	68
UWS 7	144750	134704	159	178
UWS 11	605500	352846	830	354
UWS 4	157550	156200	352	484
OSH 1	1497500	1417749	1457	1574
OSH 2	163688	79638	82	73
OSH 3	16625	6116	35	35
OSH 5	111725	122011	893	713
OSH 6	942500	321734	188	236
OSH 7	236000	118794	238	176
OSH 8	33563	25120	18	23
OSH 10	1850000	183848	7548	2853
QUEVIII 1	220875	126042	30	32
QUEVIII 2	105050	NA	264	NA
QUEVIII 3	81000	83439	886	1250
QUEVIII 4	292600	389474	99	138
QUEVIII 5	1087500	583363	106	102
QUEVIII 6	63600	16546	198	163
QUEXII 1	1483250	1777313	2741	3393
QUEXII 2	215500	219910	1657	2296
QUEXII 3	197300	241406	118	88
QUEXII 4	1074500	1082580	813	859
QUEXII 6	4294500	4985810	22184	29721
QUEXIV 1	926000	811759	4510	4540
QUEXII 7	351000	72125	1598	0

Table 4. The average quantity of mtDNA (copies/uL) at two different fragment lengths for a selection of samples extracted in this study. The quantity of mtDNA for the samples is the average of the two independent samples for each individual and all samples were quantified in triplicate.

Fragment length (bp)	Average mtDNA copies/uL	Wilcoxon signed ranks test 141-179bp (Z score)	P-value (2-tailed)
141	582320	-5.159	0.000
179	1705		

Table 5. Assessment of appropriate molecular behaviour for degraded mtDNA. The difference in quantity of mtDNA between the long (179bp) and short (141bp) fragment was assessed using a Wilcoxon signed ranks test, with a significant difference indicated by a $p<0.05$.

Components	Cronbach's Alpha	Variance Accounted For Total (Eigenvalue)	% of Variance
1	0.80	4.12	21.68
2	0.68	2.84	14.93
3	0.61	2.36	12.44
Total	0.94	9.32	49.04

Table 6. The amount of variance accounted for by the first three principal components. Eigenvalues for the components that were above 1.0 were assumed to represent true structure as opposed to noise in the data. The amount of variance (%) accounted for by each component is presented.

Population	Component 1	2	3
DEB22	-1.28	2.38	0.88
LBK20	-0.92	2.65	0.17
LBK42	-1.11	2.49	0.55
LBK34	-0.91	2.84	0.56
LVA	0.87	-0.33	0.74
LTU	0.91	-0.04	-0.35
EST	0.90	-0.06	-0.06
FIN	1.54	0.14	0.67
KAR	1.74	-0.18	-0.14
KOM	1.13	-0.72	-0.42
CHM	1.85	0.01	-0.80
MOR	1.59	-1.00	1.16
CZE	0.32	-0.08	0.13
POL	0.66	0.15	-0.03
BLR	0.51	-0.45	0.19
RUS	0.50	-0.09	-0.07
SVK	0.59	-0.07	-0.06
UKR	0.60	0.11	0.25
CHV	1.01	-0.59	0.07
TAT	0.56	-0.35	-0.24
ALB	0.26	-0.37	-0.58
GRC	-0.63	-0.11	-0.37
ARO	0.44	0.46	-0.11
ROU	0.05	0.15	0.04
SVN	0.81	0.30	-0.02
BOS	0.82	-0.27	-0.12
BUL	0.02	-0.36	-0.24
CRO	0.67	0.03	0.52
HUN	0.24	0.31	0.16
BAS	1.24	0.25	-1.40
IRL	0.42	0.66	-0.27
SCO	0.09	0.40	0.15
AUT	0.18	0.26	-0.45
ENG	0.49	0.21	-0.34
DEU	0.53	0.24	-0.21
ISL	-0.01	0.19	-0.12
NOR	0.67	0.17	-0.55
SWE	0.60	0.11	-0.29
CHE	0.54	0.46	-0.55
FRA	0.31	0.27	-0.27
ITA	-0.31	0.14	-0.08
PRT	0.20	-0.17	-0.52
SAR	0.69	0.31	-0.90
SIC	0.23	0.03	-0.63
ESP	0.69	0.17	-0.58

Table 7. PCA scores for populations based on haplogroup frequency. The country definitions are provided in Table 9.

	Component		
Population	1	2	3
TUR	-1.07	-0.67	0.33
ARM	-1.04	-0.65	0.22
GEO	-1.19	-0.82	1.36
KAB	-1.22	-1.24	-0.35
OSS	-0.91	-1.10	1.14
AZE	-0.18	-0.31	0.71
KUR	-1.33	-0.83	1.10
IRN	-1.69	-1.21	0.49
SYR	-1.69	-1.27	-0.63
IRQ	-1.96	-1.31	-0.62
JOR	-1.77	-1.89	-1.40
PSE	-1.23	-0.59	-0.55
SAU	-2.42	-1.81	-1.12
MAR	-0.59	-0.47	-2.16
LBKU	-1.99	3.46	0.15
ROS	-1.20	1.22	0.48
CW	0.00	-0.21	0.82
BB	0.66	0.30	-1.43
Unetice	0.50	-1.26	5.99

Table 7, continued. PCA scores for populations based on haplogroup frequencies. The country definitions are provided in Table 9.

	Component		
Vectors	1	2	3
EAS	0.21	-0.31	0.02
AFR	-0.41	-0.35	-0.40
R0	-0.54	-0.42	-0.14
H	0.70	-0.02	-0.48
HV	-0.60	-0.06	0.17
J	-0.41	0.11	-0.17
T	-0.28	0.58	0.48
I	0.17	-0.31	0.74
N1a	-0.38	0.78	0.11
K	-0.48	0.60	-0.05
V	0.64	0.33	-0.15
W	-0.09	0.26	0.55
X	-0.26	-0.21	0.42
U2	0.06	-0.43	0.64
U3	-0.53	-0.31	-0.12
U4	0.45	-0.24	-0.02
U5a	0.54	-0.14	0.40
U5b	0.69	-0.11	0.11
misc	-0.60	-0.67	-0.03

Table 8. Vector loading values on the haplogroup descriptors used in PCA, for the three components extracted in the PCA.

Population	Shortcut	N	Longitude	Latitude	References/Genbank Acc. No.
Derenburg	DEB22	22	11.03	51.88	[1]
LBK	LBK20	20			[2]
LBK	LBK42	42			[1,2]
LBK	LBKU	7			Extracted in this study
Albanians	ALB	281	19.75	41.38	[3,4], EBC*
Armenians	ARM	192	44.53	40.18	AJ233203-AJ233395
Aromuns	ARO	133	21.25	41.35	[4]
Austrians	AUT	117	11.37	47.35	[5,6]
Azeris	AZE	88	48	40	[7,8]
Basques	BAS	106	357.2	42.97	[9,10]
Bosnians	BOS	322	18.4	43.87	[11], AY005485-AY005664
Bulgarians	BUL	141	26.13	42.93	[8,12]
Byelorussians	BLR	352	29.2	53.12	[13], LHPG**
Chuvash	CHV	92	46.77	55.97	[8,14]
Croats	CRO	980	17.02	43.28	[15], EBC
Czechs	CZE	449	14.15	50.1	[8,16,17], EBC
English	ENG	1297	358	53	[8,18,19]
Estonians	EST	497	26.7	58.37	EBC
Finns	FIN	508	26	62	[20,21,22]
French	FRA	998	0.07	47.33	[23,24]
Georgians	GEO	158	44	42	[7], EBC
Germans	DEU	1406	9.6	52.57	[8,25,26,27,28]
Greeks	GRC	741	22.85	40.53	[4,8,29,30], EBC
Hungarians	HUN	190	19	47.5	[30], EBC
Icelanders	ISL	448	337.4	64.5	[18,31,32]
Iranians	IRN	517	51.4	35.68	[7,33]
Iraqis	IRQ	168	44.4	33.33	[8,34]
Irish	IRL	300	350.88	53.47	[35]
Italians	ITA	830	12.48	41.9	[8,15,36,37,38,39], DQ081608-DQ081668
Jordanians	JOR	146	35.82	31.78	[40]
Kabardinians	KAB	163	43	43	[8], EBC
Karelians	KAR	83	34.3	63.72	[31]
Komi	KOM	127	53	61	[14]
Kurds	KUR	73	43.07	37.55	[7,8]
Latvians	LVA	299	24	57	[41]
Lithuanians	LTU	180	24	55	[42], EBC
Mari	CHM	136	48.1	55.95	[14]
Moroccans	MAR	336	353.12	30.97	[43], EBC
Mordvinians	MOR	99	44.47	54.3	[14]
Norwegians	NOR	663	10.57	59.88	[8,44,45,46,47]
Ossetians	OSS	289	44.5	43	[8], EBC
Palestinians	PSE	117	35.1	31.82	[8]
Poles	POL	583	18.5	53	[8,48], EBC
Portuguese	PRT	848	351.97	39.48	[49,50]
Romanians	ROU	197	28.58	44.13	[4,8]
Russians	RUS	918	32.88	53.93	[8,13,51], LHPG, EBC AY959517-AY959667 AY959713-AY959798

Sardinians	SAR	115	9	40	[8]
Saudi Arabians	SAU	325	46.48	24.62	[52], EBC
Scots	SCO	1121	355	58	[44]
Sicilians	SIC	106	14.17	37.62	[53]
Slovaks	SVK	510	19.33	48.83	EBC
Slovenians	SVN	233	14.5	46.07	[11,54,55]
Spanish	ESP	704	356.17	40.63	[9,49,55,56,57,58]
Swedes	SWE	471	17.73	59.33	EBC
Swiss	CHE	230	6.58	46.68	[59,60]
Syrians	SYR	169	36.2	33.58	[8], EBC
Tatars	TAT	225	52.38	55.3	[14]
Turkish	TUR	608	33	39	[7,8,12], EBC
Ukrainians	UKR	610	32.07	49.43	LHPG, EBC

*EBC - data of Estonian Biocentre, unpublished
**LHPG - data of Laboratory for Human Population Genetics (Research Centre for Medical Genetics RAMS), unpublished

Table 9. Details of Neolithic and modern-day populations used for comparison.

References for Table 9

1. 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 (2010). Ancient DNA from European Early Neolithic farmers reveals their near eastern affinities. PLoS Biol, 8**,** e1000536.
2. Haak, W, Forster, P, Bramanti, B, Matsumura, S, Brandt, G, Tanzer, 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**,** 1016-8.
3. Belledi M, Poloni ES, Casalotti R, Conterio F, Mikerezi I, et al. (2000) Maternal and paternal lineages in Albania and the genetic structure of Indo-European populations. Eur J Hum Genet 8: 480-486.
4. Bosch E, Calafell F, Gonzalez-Neira A, Flaiz C, Mateu E, et al. (2006) Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. Ann Hum Genet 70: 459-487.
5. Parson W, Parsons TJ, Scheithauer R, Holland MM (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. Int J Legal Med 111: 124-132.
6. Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, et al. (1994) Molecular genetic analyses of the Tyrolean Ice Man. Science 264: 1775-1778.
7. Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, et al. (2004) Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. AmJHumGenet 74: 827-845.
8. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67: 1251-1276.
9. Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, et al. (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum Genet 60: 331-350.
10. Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P, et al. (1995) Human mitochondrial DNA variation and the origin of Basques. AnnHumGenet 59 (Pt 1): 63-81.
11. Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Drobnic K, et al. (2003) Mitochondrial DNA variability in Bosnians and Slovenians. AnnHumGenet 67: 412-425.
12. Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjieva L (1996) From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. AnnHumGenet 60 (Pt 1): 35-49.
13. Belyaeva O, Bermisheva M, Khrunin A, Slominsky P, Bebyakova N, et al. (2003) Mitochondrial DNA Variations in Russian and Belorussian. Human Biology 75: 647-660.
14. Bermisheva M, Tambets K, Villems R, Khusnutdinova E (2002) [Diversity of mitochondrial DNA haplotypes in ethnic populations of the Volga-Ural region of Russia]. MolBiol(Mosk) 36: 990-1001.
15. Babalini C, Martinez-Labarga C, Tolk HV, Kivisild T, Giampaolo R, et al. (2005) The population history of the Croatian linguistic minority of Molise (southern Italy): a maternal view. EurJHumGenet 13: 902-912.
16. Vanecek T, Vorel F, Sip M (2004) Mitochondrial DNA D-loop hypervariable regions: Czech population data. Int J Legal Med 118: 14-18.

17. Malyarchuk BA, Vanecek T, Perkova MA, Derenko MV, Sip M (2006) Mitochondrial DNA variability in the Czech population, with application to the ethnic history of Slavs. Hum Biol 78: 681-696.
18. Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, et al. (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59: 185-203.
19. Sykes B (2006) Blood of the Isles: Exploring the Genetic Roots of Our Tribal History: Bantam. 306 p.
20. Meinila M, Finnila S, Majamaa K (2001) Evidence for mtDNA admixture between the Finns and the Saami. HumHered 52: 160-170.
21. Lahermo P, Sajantila A, Sistonen P, Lukka M, Aula P, et al. (1996) The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. AmJHumGenet 58: 1309-1322.
22. Kittles RA, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, et al. (1999) Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. Am J Phys Anthropol 108: 381-399.
23. Richard C, Pennarun E, Kivisild T, Tambets K, Tolk HV, et al. (2007) An mtDNA perspective of French genetic variation. Annals of Human Biology 34: 68-79.
24. Dubut V, Chollet L, Murail P, Cartault F, Beraud-Colomb E, et al. (2004) mtDNA polymorphisms in five French groups: importance of regional sampling. Eur J Hum Genet 12: 293-300.
25. Poetsch M, Wittig H, Krause D, Lignitz E (2003) Mitochondrial diversity of a northeast German population sample. Forensic SciInt 137: 125-132.
26. Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, et al. (1999) Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. IntJLegal Med 112: 291-298.
27. Lutz S, Weisser HJ, Heizmann J, Pollak S (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. IntJLegal Med 111: 67-77.
28. Baasner A, Schafer C, Junge A, Madea B (1998) Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. Forensic Sci Int 98: 169-178.
29. Irwin J, Saunier J, Strouss K, Paintner C, Diegoli T, et al. (2008) Mitochondrial control region sequences from northern Greece and Greek Cypriots. Int J Legal Med 122: 87-89.
30. Bogacsi-Szabo E, Kalmar T, Csanyi B, Tomory G, Czibula A, et al. (2005) Mitochondrial DNA of ancient Cumanians: culturally Asian steppe nomadic immigrants with substantially more western Eurasian mitochondrial DNA lineages. Hum Biol 77: 639-662.
31. Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, et al. (1995) Genes and languages in Europe: an analysis of mitochondrial lineages. Genome Res 5: 42-52.
32. Helgason A, Sigurdardottir S, Gulcher JR, Stefansson K, Ward R, et al. (2000) Sampling Saturation and the European MtDNA Pool: Implications for Detecting Genetic Relationships among Populations. Archaeogenetics: DNA and the population prehistory of Europe. Cambridge: McDonald Institute for Archaeological Research. pp. 285-294.
33. Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, et al. (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 Genetics 5: -.
34. Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, et al. (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.
35. McEvoy B, Richards M, Forster P, Bradley DG (2004) The Longue Duree of genetic ancestry: multiple genetic marker systems and Celtic origins on the Atlantic facade of Europe. AmJHumGenet 75: 693-702.
36. Tagliabracci A, Turchi C, Buscemi L, Sassaroli C (2001) Polymorphism of the mitochondrial DNA control region in Italians. Int J Legal Med 114: 224-228.
37. Achilli A, Olivieri A, Pala M, Metspalu E, Fornarino S, et al. (2007) Mitochondrial DNA variation of modern Tuscans supports the near eastern origin of Etruscans. Am J Hum Genet 80: 759-768.
38. Francalacci P, Bertranpetit J, Calafell F, Underhill PA (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. Am J Phys Anthropol 100: 443-460.
39. Mogentale-Profizi N, Chollet L, Stevanovitch A, Dubut V, Poggi C, et al. (2001) Mitochondrial DNA sequence diversity in two groups of Italian Veneto speakers from Veneto. Ann Hum Genet 65: 153-166.
40. Gonzalez AM, Karadsheh N, Maca-Meyer N, Flores C, Cabrera VM, et al. (2008) Mitochondrial DNA variation in Jordanians and their genetic relationship to other Middle East populations. Annals of Human Biology 35: 212-231.
41. Pliss L, Tambets K, Loogvali EL, Pronina N, Lazdins M, et al. (2006) Mitochondrial DNA portrait of Latvians: towards the understanding of the genetic structure of Baltic-speaking populations. Ann Hum Genet 70: 439-458.

42. Kasperaviciute D, Kucinskas V, Stoneking M (2004) Y chromosome and mitochondrial DNA variation in Lithuanians. AnnHumGenet 68: 438-452.
43. Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, et al. (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. AnnHumGenet 62 (Pt 6): 531-550.
44. Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, et al. (2001) mtDNA and the Islands of the North Atlantic: Estimating the Proportions of Norse and Gaelic Ancestry. AmJHumGenet 68: 723-737.
45. Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Borresen-Dale AL, et al. (2002) Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. EurJHumGenet 10: 521-529.
46. Opdal SH, Rognum TO, Vege A, Stave AK, Dupuy BM, et al. (1998) Increased number of substitutions in the D-loop of mitochondrial DNA in the sudden infant death syndrome. Acta Paediatr 87: 1039-1044.
47. Dupuy BM, Olaisen B (1996) mtDNA sequences in the Norwegian Saami and main populations. In: Carracedo A, Brinkmann B, Bär W, editors. Advances in forensic haemogenetics. Berlin: Springer. pp. 23-25.
48. Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Wozniak M, et al. (2002) Mitochondrial DNA variability in Poles and Russians. AnnHumGenet 66: 261-283.
49. Gonzalez AM, Brehm A, Perez JA, Maca-Meyer N, Flores C, et al. (2003) Mitochondrial DNA affinities at the Atlantic fringe of Europe. Am J Phys Anthropol 120: 391-404.
50. Pereira L, Cunha C, Amorim A (2004) Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database. Int J Legal Med 118: 132-136.
51. Kornienko IV, Vodolazhskii DI, Afanas'eva GV, Ivanov PL (2004) [Polymorphism of the central region of D-loop of mitochondrial DNA and personality identification by forensic medicine methods]. Sud Med Ekspert 47: 27-32.
52. Abu-Amero KK, Larruga JM, Cabrera VM, Gonzalez AM (2008) Mitochondrial DNA structure in the Arabian Peninsula. BMC Evol Biol 8: 45.
53. Cali F, Le Roux MG, D'Anna R, Flugy A, De Leo G, et al. (2001) MtDNA control region and RFLP data for Sicily and France. Int J Legal Med 114: 229-231.
54. Zupanic Pajnic I, Balazic J, Komel R (2004) Sequence polymorphism of the mitochondrial DNA control region in the Slovenian population. Int J Legal Med 118: 1-4.
55. Larruga JM, Diez F, Pinto FM, Flores C, Gonzalez AM (2001) Mitochondrial DNA characterisation of European isolates: the Maragatos from Spain. Eur J Hum Genet 9: 708-716.
56. Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo A (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. Eur J Hum Genet 6: 365-375.
57. Crespillo M, Luque JA, Paredes M, Fernandez R, Ramirez E, et al. (2000) Mitochondrial DNA sequences for 118 individuals from northeastern Spain. IntJLegal Med 114: 130-132.
58. Maca-Meyer N, Sanchez-Velasco P, Flores C, Larruga JM, Gonzalez AM, et al. (2003) Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). Ann Hum Genet 67: 329-339.
59. Dimo-Simonin N, Grange F, Taroni F, Brandt-Casadevall C, Mangin P (2000) Forensic evaluation of mtDNA in a population from south west Switzerland. Int J Legal Med 113: 89-97.
60. Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, et al. (1994) Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. Biol Chem Hoppe Seyler 375: 837-840.

Culture	n	no. of haplotype	Nucleotide Diversity +-SD	Mean number of pairwise differences	Segregating sites	Transition/Transversion ratio
LBK	49	19	0.020118	5.552479	26	0.96153846
Rössen	10	6	0.019876	5.485804	20	0.94736842
Corded Ware	24	7	0.017374	4.795224	24	1
Bell Beaker	8	2	0.005392	1.48814	5	1
Unetice	17	8	0.021039	5.806866	22	0.84210526
Central Europe	1757	636	0.01577	4.3526	156	0.78947368
F_{ST}s	CEN	LBK	Rössen	Corded Ware	Bell Beaker	Unetice
Central Europe	0.00000	0.04667	-0.00701	0.00841	-0.0435	0.03874

Table 10. Intra and inter-population statistics calculated from the Neolithic (*n*=108) and modern Central European (*n*= 1757) mtDNA sequences (np16069 - 16365), detailed in Table 9. A Kimura-2-P model was used with a gamma distribution and shape parameter of 0.22 applied.

Genetic structure	FCT	FSC	FST	% of Variation Among groups	Among populations within groups	Within populations	P value (among groups)
H_1							
CEN, LBK vs. Ros vs. CW vs. BB vs. UNE	0.007	0.003	0.01	0.69	0.33	98.98	0.095
CEN, Ros vs. LBK vs. CW vs. BB vs. UNE	0.029	0	0.029	2.85	0.02	97.13	0
CEN, CW vs. LBK vs. Ros vs. BB vs. UNE	0.03	0.001	0.03	2.96	0.06	96.98	0
CEN, BB vs. LBK vs. Ros vs. CW vs. UNE	0	0.031	0.031	3.1	0	96.9	0
CEN, UNE vs. LBK vs. Ros vs. CW vs. BB	0.022	0.001	0.023	2.22	0.12	97.66	0.001
H_2							
CEN, LBK, Ros vs. BB vs. CW, UNE	0.007	0.003	0.011	0.72	0.34	98.94	0.123
CEN, BB vs. LBK, Ros vs. CW, UNE	0	0.03	0.03	3.03	0	96.97	0
CEN, CW, UNE vs. LBK, Ros vs. BB	0.001	0.029	0.03	2.87	0.13	97	0.001
H_3							
CEN, LBK, Ros vs. CW, BB vs. UNE	0.012	0.003	0.015	1.21	0.3	98.49	0.071
CEN, CW, BB vs. LBK, Ros vs. UNE	0.039	0	0.039	3.93	0.01	96.06	0
CEN, UNE vs. LBK, Ros vs. CW, BB	0.024	0.001	0.025	2.45	0.1	97.46	0
H4							
CEN, LBK, Ros vs. CW, BB, UNE	0.039	0.001	0.040	3.9	0.1	96.00	0.00196
CEN, CW, BB, UNE vs. LBK, Ros	0.006	0.003	0.010	0.65	0.34	99.01	0.10264

Table 11. AMOVA hypothesis testing to assess the relationship among Early Neolithic culture and Late Neolithic/Early Bronze Age cultures, and modern Central European populations. Abbreviations are as follows; modern Central European data (CEN, n=1757), Linear Pottery Culture (LBK), Rössen (Ros), Corded Ware (CW), Bell Beaker (BB) and Unetice (UNE). A null distribution was calculated by re-sampling the data 1023 times. A gamma distribution was assumed, with shape parameter of 0.22 applied.

Hypothesis Testing	No. of Parameters	AIC	AICmodel - AIClowest
One population model Exponential model (H_o)	4	61.90	11.47
Five population model (H_1)			
H_1 - LBK most likely ancestors to CEN	15	58.62	8.18
H_1- Rössen most likely ancestors to CEN	15	55.68	5.24
H_1 - Corded Ware most likely ancestors to CEN	15	56.49	6.05
H_1 - Bell Beaker most likely ancestors to CEN	15	50.92	0.49
H_1 - Late Neolithic Culture (AK) 5	15	57.46	7.03
Three population model (H_2) based on archaeological evidence : Early Farmers, potentially Kurgan Cultures and Western European cultures			
H_2 - Early Farmers (LBK, Rössen) most likely ancestors to CEN	9	62.65	12.22
H_2 - Kurgan cultures (Corded Ware and Unetice) most likely ancestors to CEN	9	60.89	10.45
H_2 - Bell Beaker most likely ancestors to CEN	9	50.44	0.00
Three population model (H_3) based on temporal intervals: Early Neolithic, Late Neolithic and Early Bronze Age			
H_3 - Early farmers (LBK, Rössen) most likely ancestors to CEN	9	61.16	10.72
H_3 - Late Neolithic cultures (Corded Ware, Bell Beaker) most likely ancestors to CEN	9	53.06	2.63
H_3 - Bronze Age (Unetice) most likely ancestors to CEN	9	59.86	9.42
Two population model (H_4) Early Neolithic versus Late Neolithic/Early Bronze Age			
H_4 Corded Ware, Bell Beaker and Unetice most likely ancestors to CEN	6	54.77	4.33
H_4 LBK, Rössen most likely ancestors to CEN	6	60.06	9.63

Table 12. Model comparison results from the simulated coalescent analyses for the four hypotheses tested (14 models in total). All analyses were run for a minimum of 1, 000, 000 generations in BayeSSC. Approximate Bayesian Computation (ABC) was used to compare the top 1% of simulated and observed data using five F_{ST}s between contemporary Central Europeans (CEN) and each of the five Neolithic cultures. Models were re-run for 1 000 generations in BayesSSC with the maximum likelihood value for all parameters with a prior to determine the likelihood (L) of the model. This was used to calculate the Akaike's Information Criterion (AIC) to compare the models, and which also accounts for the model complexity.

CHAPTER FOUR

Ancient DNA from oral microbiota records the impact of dietary change on the evolution of human oral diseases

Ancient DNA from oral microbiota records the impact of dietary change on the evolution of human oral diseases

Summary

Human evolution has been punctuated by major changes in diet and culture, with important impacts on our biology. Two of the biggest dietary shifts involved increasingly carbohydrate-rich diets[1,2] in the Neolithic (~ 10,000 years BP in the Near East)[1,3] and the industrial processing of staples such as flour and sugar in the Industrial Revolution (~ 1800 AD)[4]. Increased signs of physiological stress in the Neolithic skeletal records suggest that these changes directly underpin many diseases associated with modern lifestyles[5,6]. Commensal bacteria comprise the majority of cells in the human body[7], and the increasing recognition of their role in human disease[8-12] questions how shifts in human diet might have disrupted co-evolved host-microbiota mutualisms. However, the evolutionary history of microbiota is poorly quantified, as preserved commensal bacterial DNA has yet to be recovered from fossil remains. Here we show that bacterial DNA within dental calculus (calcified plaque) on ancient human teeth provides a new record of past health changes. Oral bacteria form part of the human microbiota and in addition to causing oral disease[8,9], are associated with systemic diseases including arthritis[12], cardiovascular disease[10] and diabetes[11]. We found that dental calculus from early European farming communities (from ~ 7,500 years BP) have higher oral microbe diversity compared to modern populations[8,9,13,14], with a dominance of periodontal disease-associated bacteria. The composition of oral microbiota remained near-constant between Neolithic and Medieval times, before the now ubiquitous caries-forming bacteria became dominant, probably during the Industrial Revolution. Our data show that the ecosystem of the human mouth has altered over time, possibly due to changes in diet and oral hygiene. They also provide insight into the relationship between oral and systemic diseases. In particular, our finding that periodontal disease-associated bacteria have been present since the introduction of agriculture suggests that these bacteria are not contributing to the recent rise in prevalence of heart disease and diabetes[15]. Archaeological dental calculus provides a unique means to track the evolution of human disease, and reconstruct the past diversity and composition of 'indigenous' human microbiota.

Main Text

Dental plaque is a biofilm which is extremely dense in bacteria (2×10^8 mg^{-1})[7] that occurs both above and below the gum (supra- and subgingivally[16]). The development of dental calculus occurs when plaque becomes mineralised with calcium phosphate, which accretes both within and between the bacterial cells[17]. Bacteria in calculus are locked in a crystalline structure similar to bone[17] (Supplementary Information [SI] Figure S1.1), which is capable of preserving the morphology of bacteria even in 60,000 year-old Neanderthal samples[18]. Oral bacteria are predominantly transferred vertically from the primary caregiver(s) in early childhood[19], making archaeological dental calculus a potentially unique means of tracing admixture between ancient cultures, including population movement and disease spread.

The increased consumption of domesticated cereals (wheat and barley in the Near East) during the Neolithic is associated with an increased prevalence of dental calculus and oral pathology observable in human skeletons[6]. Specifically, the diseases include dental caries (tooth decay)[8] and periodontal disease[14], both of which although observed in the teeth of hunter-gatherers[6] and early hominids, such as *Australopithecus*[20], were rare. Caries is now an endemic disease, affecting 60-90% of school-aged children in industrialised countries, whilst periodontal disease occurs in 5-20% of the adult population worldwide[21]. Caries and periodontal disease are thought to result from perturbation of a formerly ecologically balanced oral biofilm through the increased consumption of fermentable carbohydrates[7,15]; however, it is unknown when or how this occurred.

We collected a mixture of supra- and subgingival calculus samples from seven ancient European agriculturist groups, spanning from the Neolithic to Medieval period, from 28 individuals, ranging in age from under 20 to over 60 years (36% female, 28% male and 36% unknown, SI Table S1.1). The samples included individuals from the earliest farming culture in Central Europe, the Linear Pottery Culture (LBK), as well as the late Neolithic (Bell Beaker), later Bronze Age and medieval rural and urban populations. We carried out all ancient DNA work in a laboratory dedicated to ancient environmental and bacterial DNA research, within the physically isolated, specialist Australian Centre for Ancient DNA facility, using strict decontamination and authentication protocols (SI, Section 1). We extracted bacterial DNA from sterilised calculus samples (n = 28) and used it to generate PCR amplicon libraries of the 16S rRNA gene, targeting three hypervariable regions (V1, V3 and V6, SI Tables S1.2 and S1.3) with barcoded primers. We sequenced the amplicons plus multiple

extraction blanks using both conventional Sanger and 454 pyrosequencing (GS-FLX Titanium) technology. Some 916,147 sequences were generated, and we discarded around 50% of these following quality filtering and denoising to remove sequences containing PCR and sequencing errors[22], leaving 410,427 sequences (SI Table S1.4).

The bacterial composition of the ancient calculus material was similar to modern oral samples from the Human Oral Microbiome Database (HOMD)[23] at the phylum level, and clearly distinct from laboratory reagents (extraction blanks), irrespective of which region of 16S rRNA gene was analysed (Figure 1, SI, section 2). For the V3 region, a dominance of Firmicutes in archaeological calculus (39%) was comparable to both the HOMD (37%) and pyrosequencing studies of modern oral samples[24,25]. The sequences we retrieved from the dental calculus samples were clearly distinguished from the bacterial sequences obtained from extraction blanks (6% Firmicutes), which were dominated by Proteobacteria (70%), typical of bacterial communities found in clean-room environments[26]. The ancient dental calculus samples contained all 15 phyla commonly found in the modern human oral cavity[23], with high proportional abundances of Actinobacteria (19%) matching those observable in modern mature plaque deposits. Interestingly, archaeological calculus displayed similarities to the oral mucosa of Amazon Amerindians, in that both contained higher frequencies of the *incertae* classes of Bacteroidetes and Firmicutes (in addition to Acidobacteria), compared to modern-day Western populations[27].

Since the archaeological specimens were buried, we compared the ancient sample data with modern oral and environmental microbial sequences to test for potential contamination from soil micro-organisms. Phylogenetic analyses of β-diversity, which is a measure of the number of operational taxonomic units (OTUs) that are unique between groups, confirmed that bacterial sequences from archaeological calculus appear to be endogenous, being more similar to modern plaque[8,9,13,14] than environmental samples[28-34] (SI Section 3). A plot of the Principal Co-ordinates Analysis (PCoA) scores generated from the phylogeny-based metric UniFrac for V3 (Figure 2) shows that the ancient calculus samples cluster with modern plaque and saliva, and that this group significantly differs (SI Table S3.1) from the environmental samples ($p < 0.001$) and extraction blanks ($p = 0.025$). The environmental and extraction blank sequences clustered together strongly ($p = 0.125$, non-parametric MANOVA). We observed the same pattern for sequences from V1 and V6 (SI Figure S3.1).

While we observed Gram-positive bacteria at similar relative abundances in the ancient calculus and modern oral samples, Gram-negative bacteria were less frequent in the ancient samples, suggesting a potential taphonomic bias related to cell wall structure (SI, Section 3). For instance, from linear regression analysis we found a decrease in relative abundance of Bacteroidetes ($p = 0.006$), Spirochaetes ($p = 0.09$) and Fusobacteria ($p = 0.1$) with increasing 'time depth'. To minimise the influence of this possible bias, we used unweighted-UniFrac distances (i.e. which takes into account only the presence or absence of lineages, and not their relative abundances) for all β-diversity analyses. Additionally, we applied all phylogeny-based analyses to both the full dataset and a subset of Gram-positive lineages only. The exclusion of Gram-negative taxa did not appear to affect the results (SI, Section 3).

The temporal transect of ancient dental calculus samples allows the timing and nature of shifts in human oral bacterial composition and diversity to be reconstructed, including the origins of key pathogens. A surprising finding is that early European farming communities have higher oral microbial diversity ($p < 0.0001$, SI Table S4.1) than modern Europeans, including many bacteria associated with periodontal disease (e.g. *Porphyromonas gingivalis,* Clostridiaceae, Synergistaceae, TM7, Lachnospiraceae, Parvimonas), but with low abundance of now ubiquitous pathogens such as S*treptococcus mutans* (Figure 3, SI Section, 4). The dominant bacteria detected from the Neolithic (~ 7, 500 years) through to the Medieval (~ 400 years) periods are notably consistent, in parallel with the broad similarity of food processing technology during this time[4]. The earliest farming groups we analysed (LBK and Bell Beaker) also harboured *Methanobrevibacter* which has been primarily found in individuals with severe periodontal disease[35]. While periodontal disease-associated bacteria would be expected to be well represented in dental calculus[36], comparison of the ancient dental calculus samples with plaque from modern individuals with periodontal disease revealed the ancient samples still had a lower abundance of bacteria associated with other oral diseases, such as caries (SI Section 4). The prevalence of periodontal disease-associated bacteria in early Neolithic samples supports current hypotheses that this disease is associated with an agricultural diet, due to increased amounts of soft carbohydrate foods compared with hunter-gatherer diets[7]. This finding is consistent with increases in periodontal disease apparent in the human skeletal record[37] and domesticated animals[38] following the transition to an agricultural diet and suggests that diet-mediated changes have substantially altered the ecosystem composition of the human mouth. A further consideration is the association between periodontal disease and old age[39]. Periodontal disease-associated pathogens, such as *Tannerella*, *Treponema* and *Methanobrevibacter*, were found even in the youngest (3-4 years) individual sampled (ID

8247). Our finding of periodontal disease-associated pathogens in individuals from a range of ages indicates that the results are not primarily related to age. One underlying assumption in this analysis is that the microbiota isolated from ancient calculus is representative of the plaque microbiota in those individuals, and thus comparable to modern plaque samples. We have preliminary evidence in modern samples to support this assumption (SI Section 3).

In comparison to the dominance of periodontal disease-associated bacteria between the Neolithic and Medieval period, today's oral environment is less biodiverse and dominated by potentially cariogenic bacteria (Figure 3, e.g. *S. mutans*). The dominance of caries-associated bacteria occurred after the Medieval period, and is most likely associated with the onset of the Industrial Revolution that began some 200 years ago and represented the largest change in food production and processing technology since the origins of farming in the Neolithic[4]. The Industrial Revolution saw the production of refined grain and concentrated sugar from processed sugar beet[4], generating mono- and disaccharides (the main substrates for microbial fermentation) which lowers plaque pH and causes enamel demineralisation[7]. Perhaps more importantly, the decline in overall oral microbial diversity ($p < 0.001$, SI Table S4.1) indicates that over the past few hundred years, the human mouth has become a substantially less biodiverse ecosystem, and therefore probably an environment less resilient to perturbations[40] in the form of diet imbalances and invasion[41] by pathogenic bacteria species. Aside from dietary changes, the recent development of oral hygiene may have contributed to the observed decline in oral microbiota diversity within calcified plaque. The frequent disruption of biofilms by brushing teeth[42], flossing[43] and through the use of mouth washes which contain antimicrobial agents[44] has been found to prevent the establishment plaque and hence the development of complex communities in oral biofilms[45].

It appears that variation in intake of carbohydrates, and potentially the development of oral hygiene, has altered the ecosystem structure of the mouth, opening up pathogenic niches for caries and periodontal disease at different times in human history. By determining the past state of oral microbiota, our data could be used to assist in understanding their current association with systemic diseases. For example, others have proposed that both diabetes and heart disease[15] are partially attributable to periodontal disease. However, while these systemic diseases have risen in prevalence over the last few decades[15], our ancient calculus data reveal that the abundance of periodontal disease-associated bacteria has been relatively stable since the introduction of farming. This indicates that while periodontal disease has been found to contribute to the pathogenesis of diabetes and heart disease by producing a prolonged

inflammatory state[10], it is probably not contributing to the rising incidence of these systemic diseases.

In addition to recording the changes in composition of microbiota over time, the bacteria in ancient calculus also provide a powerful new means to reconstruct dietary differences and cultural variation between ancient human groups. A discriminant analysis of the principal components generated from the UniFrac metric for V3, demonstrates non-random variation between each of the archaeological groups investigated, with the first two functions accounting for 95.3% of the variance between groups ($p < 0.001$, Figure 4). The differentiation and patterns of group relationships were consistent for all 16S rRNA gene regions, regardless of whether or not the analysis was restricted to Gram-positive bacteria (SI, section 5).

We have opened a powerful new avenue of research in bio-anthropology, one that will elucidate the response of human microbiota to a variety of environmental changes around the world. There is now scope to compare the composition and diversity of the oral microbiota from dental calculus in early agricultural and hunter-gatherer societies, to examine directly the effects of this and other nutritional transition(s) on oral and systemic health through time. There is also the potential to recover distinct evolutionary, phylogeographic and even socio-economic and cultural signatures from these ancient commensal microbes.

Methods

The ancient dental calculus sample details, including archaeological information, preparation methodology and authentication criteria, are described in SI section 1

DNA extraction, PCR, cloning and sequencing. We used between 0.05–0.2 g of powdered dental calculus for DNA extraction. A non-template control was included every two to three extractions. The dental calculus samples and non-template controls were lysed in 1 ml of lysis buffer, which contained 0.5 M EDTA (pH 8), SDS (10%) and Proteinase K (20 mg/ml). The samples and lysis buffer were rotated for 24 hours at 55 °C. Following the lysis of samples, DNA was isolated using the QIAamp DNA Investigator Kit (Qiagen). The final extraction volume was 100 µl, and extracts were stored at 4 ^{o}C. We used specific and universal primers to identify microbial DNA in the dental calculus samples and non-template controls. Specific PCR primers were used to amplify oral pathogens, *Streptococcus mutans* (GtfB gene) and *Porphyromonas gingivalis* (16S rRNA gene) (SI Table S1.2). The universal 16S rRNA gene primers were used to amplify a broad range of microbes within the samples and non-template controls (SI Table S1.2). For all primer sets, the PCR conditions were as follows; Amplitaq Gold (Applied Biosystems) at 2 U in 25µl volumes using 1× Buffer Gold, 2.5 mM $MgCl_2$, 0.25 mM of each dNTP (Fermentas), 400 µM of each primer, 1 mg/ml RSA (Sigma-Aldrich), ShrimpDNAse (Affymetrix) at 0.002 U/ul and 2 µl of DNA extract. ShrimpDNase was used to remove microbial contamination from PCR reagents prior to the amplification reaction. This enzyme works by cleaving double stranded DNA. ShrimpDNase was added to the PCR mixture (minus the extract) and incubated at 37 °C for 15 minutes, then inactivated by heating the mixture to 65 °C for 15 minutes. For the specific primers, the thermocycling conditions consisted of an initial enzyme activation at 95 °C for 6 minutes, followed by 45 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s and elongation at 72 °C for 30 s and a single final extension step at 60 °C for 10 minutes. The 16S rRNA gene universal primers were amplified using the following conditions: initial enzyme activation at 95 °C for 6 minutes, followed by 40 cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 30 s and elongation at 72 °C for 30 s with a single final extension step at 60 °C for 10 minutes. Each PCR reaction included multiple extraction and PCR blanks. All PCRs were visually examined by electrophoresis on 3.5% agarose TBE gels. Specific PCR products were purified using 5 µl of amplified product, exonuclease I (0.8 U/µl) and shrimp alkaline phosphatase (1 U/µl). The mixture was heated to 37 °C for 40 minutes and then heat inactivated at 80 °C for 10 minutes. The purified amplicons were directly sequenced using a BigDye Terminator 3.1 Kit (Applied Biosystems) according to the manufacturer's instructions. Sequencing products

were purified using a Multiscreen$_{HTS}$ Vacuum Manifold (Millipore), according to the manufacturer's protocol. Sequencing products were separated on the 3130xl Genetic Analyzer (Applied Biosystems) and the resulting sequences edited using Sequencher (version 4.7).

We cloned the 16S rRNA gene universal amplicons to monitor contamination within samples and non-template controls, and to assess the suitability of calculus samples for 454 sequencing. The PCR products were purified using Agencourt Ampure (Beckman Coulter) according to the manufacturer's instructions. The purified amplification products were cloned using a StrataClone PCR cloning Kit (Stratagene). Clones were directly added to the colony PCR mix, containing 10× Hotmaster Buffer (Eppendorf), Hotmaster Taq 0.5 U/µl (5Prime), forward and reverse M13 primers (10 µM) (SI Table S1.2) in a 25 µl reaction. The thermocycling conditions consisted of 94 ^{o}C for 10 minutes and 35 cycles of 94 ^{o}C for 20 s, 55 ^{o}C for 10 s and 65 ^{o}C for 45 s, with a single extension of 65 ^{o}C for 10 minutes. Colony PCRs were visually inspected on 2% agarose TBE gels. Colony PCR products were purified and sequenced using the same protocols as described for the products of the specific primers.

454 GS FLX Titanium Sequencing. Pyrosequencing was applied to 28 ancient dental calculus samples and two extraction blanks. For these samples, three hypervariable regions of the 16S rRNA gene (V1, V3 and V6) were amplified, using the conditions described above. The forward and reverse primers contained 454 Lib-L kit A- and B-adaptors, respectively, at the 5' end. The forward primer also contained a sample-specific barcode (SI Table S1.3). The barcodes were developed by the Human Microbiome Project (http://www.hmpdacc.org/tools_protocols/tools_protocols.php) and had not previously been used in either the Australian Centre for Ancient DNA or the Sanger Institute, where the 454 sequencing was performed. Hence, all sequences retrieved which did not contain a barcode were assumed to be contaminants and discarded. Each region of the 16S rRNA gene was amplified twice (on different days) and the duplicates were pooled for 454 sequencing. The combining of duplicate PCR reactions was performed to minimise the potential impact of preferential sequence amplification.

Filtering, OTU picking, alignment and taxonomic assignment of 454 sequences. The sequences produced from the GS FLX Titanium analysis were processed using the QIIME (version 1.2.0) software package[46]. Quality filtering was done to remove sequences which were either under 60 bp (potential primer dimers), contained ambiguous bases, had primer or barcode mismatches, contained homopolymers which exceeded six bases or had an average quality score below 25. The remaining sequences were between 60 and 210 bp in length. The quality-filtered sequences were denoised[22] to remove sequences containing errors produced

during the PCR and sequence reaction. Denoising resulted in the removal of ~ 50% of sequences, which were identified as having ambiguous flow data (SI Table S1.4). However, we found that sequence classification from the non-denoised data was comparable to the denoised dataset. Similar sequences were binned into OTUs using optimal UCLUST[47] at a 95% likeness. Clustering is more commonly performed at 97%; however, a 95% cut-off has been found to more accurately classify OTUs for closely related, short sequences[48]. Representative sequences from each OTU were aligned using PyNAST[46] against the GreenGenes core set, with a minimum length of 60 bp and identity of 75%. PyNAST aligns the short GS FLX generated sequences (60-210 bp) against the full 16S rRNA gene gene. We removed those columns from the alignment which solely contained gaps prior to building phylogenetic trees. To overcome the difficulty in aligning highly variable 16S rRNA gene sequences, it is common to hide or lane-mask regions where at least 50% of the base composition is not conserved[49]. We did not hide variable regions because lane-masked alignments can 'mute' the phylogenetic diversity observed[48]. The gap-filtered sequences were taxonomically assigned using the RDP classifier and nomenclature [50].

We provide descriptions of the analyses done on the ancient dental calculus sequences in the SI; β-diversity (Section 3), alpha-diversity (Section 4) and discriminant analyses (Section 5).

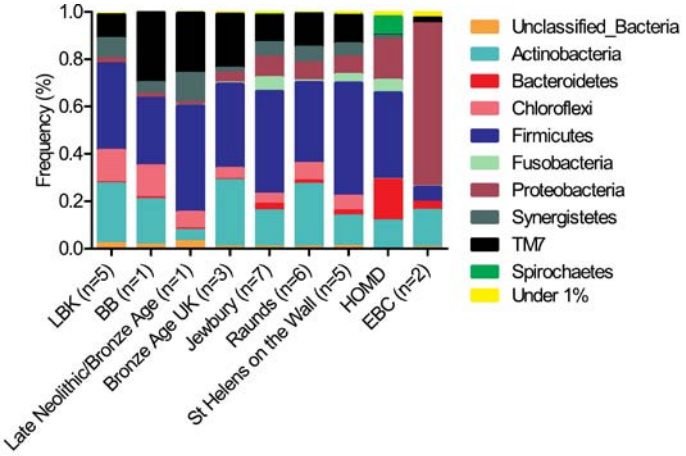

Figure 1. Microbial composition of the ancient dental calculus deposits was similar at the phylum level to modern oral samples from the Human Oral Microbiome Database (HOMD)[23], and distinct from non-template controls (EBC). For the V3 region, the phylum frequencies are presented for the ancient calculus samples (LBK; Linear Pottery Culture, BB; Bell Beaker) and non-template controls (or extraction blanks). We determined the frequencies of phyla in HOMD from partial and full-length 16S rRNA gene. The phyla with a frequency below 1% include; Archaea, Acidobacteria, Chlamydiae, Cyanobacteria, Deinococcus-Thermus, Planctomycetes, SR1, Tenericutes, Verrucomicrobia and Other.

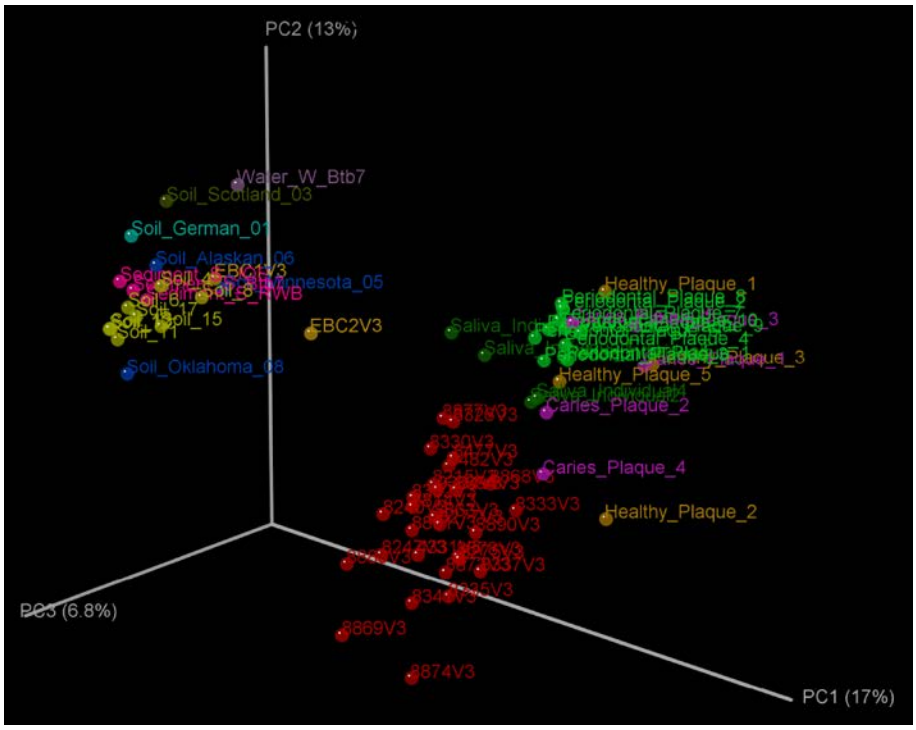

Figure 2. PCoA plot of *β*-diversity reveals a close phylogenetic relationship between ancient dental calculus and modern oral samples, both of which are distinct from the non-template controls and environmental samples. *β*-diversity was calculated using the UniFrac metric for the V3 region. PCoA was applied to the unweighted, UniFrac distances. Ancient calculus deposits (red circles) were compared to modern oral and environmental samples. We compiled a comparative modern dataset from clone- and pyro- sequence data (339,076 sequences in total). This included sequences from European oral samples (healthy[9,13], caries[8] and periodontal disease[14] sites), in addition to soils[28-34], sediments and freshwater[28]. The clone sequences partially covered the 16S rRNA gene and were trimmed to the V3 region prior to analysis.

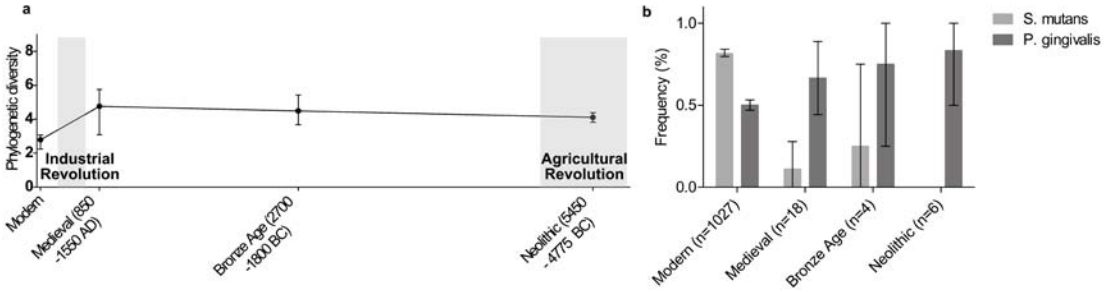

Figure 3. Changes in the diversity (a) and composition (b) of oral microbiota since the introduction of agriculture during the Neolithic in Europe. (a) Following the Medieval period there is a decrease in the phylogenetic measure of α-diversity[51] that coincides with the onset of the Industrial Revolution. For the V3 region, the phylogenetic diversity of the archaeological dental calculus samples (n = 28) was estimated from only classified bacterial sequences and compared to modern plaque (n = 20)[8,9,14] and saliva (n = 5)[13]. Phylogenetic diversity was calculated at a depth of 45 sequences and bootstrapped to assess the robustness of the pattern. In conjunction with reduced diversity, we observed a change in the composition of oral microbiota. (b) During the early farming period compared to today, we observed from the specific primer generated data, a higher frequency of periodontal disease-associated bacteria (*P. gingivalis*) and lower frequency of caries-associated bacteria (*S. mutans*). The ancient dental calculus samples were compared to modern plaque and saliva samples, which had also been sequenced using specific primers for *P. gingivalis* and *S. mutans*[52-57]. Individuals from these studies covered a wide range of health states.

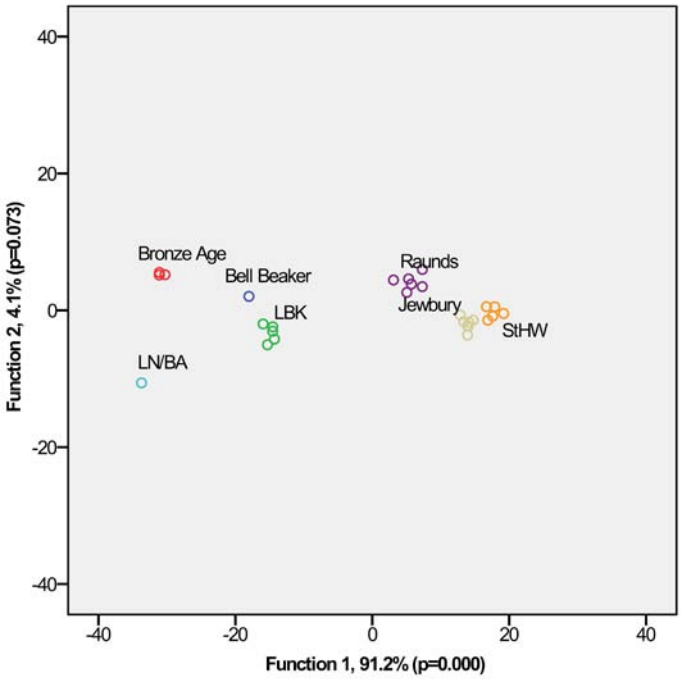

Figure 4. Discriminant analysis of *β*-diversity shows clustering of individuals from the same archaeological groups based on the microbial makeup of dental calculus. We applied discriminant analysis to the principal components generated from the unweighted UniFrac distances, which were calculated from the V3 region sequences. Each individual is represented by a circle and coloured according to archaeological grouping (LBK, green; Bell Beaker, blue; Late Neolithic/Bronze Age (LN/BA), aqua; Bronze Age, red; Raunds, purple; Jewbury, brown and St Helen on the Walls (StHW), orange). The majority of phylogenetic variation (91.2%) was described by the first discriminant function.

REFERENCES

1 Braidwood, R. J., Howe, B. & Reed, C. A. The Iranian Prehistoric Project: New problems arise as more is learned of the first attempts at food production and settled village life. *Science* **133**, 2008-2010, (1961).

2 Oelzea, V. M. *et al.* Early Neolithic diet and animal husbandry: stable isotope evidence from three Linearbandkeramik (LBK) sites in Central Germany *Journal of Archaeological Science* **38**, 270-279, (2011).

3 Childe, V. G. *The dawn of European civilisation.*, (Kegan Paul, 1925).

4 Cordain, L. *et al.* Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr* **81**, 341-354, (2005).

5 Angel, J. L. Porotic hyperostosis, anemias, malarias, and marshes in the prehistoric Eastern Mediterranean. *Science* **153**, 760-763, (1966).

6 Aufderheide, A. C., Rodriguez-Martin, C. & Langsjoen, O. *The Cambridge encyclopedia of human paleopathology* (Cambridge University Press, 1998).

7 Marsh, P. D. Microbiology of dental plaque biofilms and their role in oral health and caries. *Dent Clin North Am* **54**, 441-454, (2010).

8 Aas, J. A. *et al.* Bacteria of dental caries in primary and permanent teeth in children and young adults. *J Clin Microbiol* **46**, 1407-1417, (2008).

9 Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I. & Dewhirst, F. E. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* **43**, 5721-5732, (2005).

10 Dave, S. & Van Dyke, T. The link between periodontal disease and cardiovascular disease is probably inflammation. *Oral Dis* **14**, 95-101, (2008).

11 Grossi, S. G. & Genco, R. J. Periodontal disease and diabetes mellitus: a two-way relationship. *Ann Periodontol* **3**, 51-61, (1998).

12 Mercado, F. B., Marshall, R. I., Klestov, A. C. & Bartold, P. M. Relationship between rheumatoid arthritis and periodontitis. *J Periodontol* **72**, 779-787, (2001).

13 Lazarevic, V., Whiteson, K., Hernandez, D., Francois, P. & Schrenzel, J. Study of inter- and intra-individual variations in the salivary microbiota. *BMC Genomics* **11**, 523, (2010).

14 Faveri, M. *et al.* Microbiological diversity of generalized aggressive periodontitis by 16S rRNA clonal analysis. *Oral Microbiol Immunol* **23**, 112-118, (2008).

15 Hujoel, P. Dietary carbohydrates and dental-systemic diseases. *J Dent Res* **88**, 490-502, (2009).

16 Lieverse, A. R. Diet and the aetiology of dental calculus *International Journal of Osteoarchaeology* **9**, 219 - 232, (1999).

17 Jin, Y. & Yip, H. K. Supragingival calculus: formation and control. *Crit Rev Oral Biol Med* **13**, 426-441, (2002).

18 Vandermeersch, B. *et al.* Middle Palaeolithic dental Bacteria from Kebara, Israel. *C.R. Acad Sci. Paris* **319**, 727-731, (1994).

19 Suchett-Kaye, G., Decoret, D. & Barsotti, O. Intra-familial distribution of Fusobacterium nucleatum strains in healthy families with optimal plaque control. *J Clin Periodontol* **26**, 401-404, (1999).

20 Grine, F. E., Gwinnett, A. J. & Oaks, J. H. Early hominid dental pathology: interproximal caries in 1.5 million-year-old Paranthropus robustus from Swartkrans. *Arch Oral Biol* **35**, 381-386, (1990).

21 Petersen, P. E., Bourgeois, D., Ogawa, H., Estupinan-Day, S. & Ndiaye, C. The global burden of oral diseases and risks to oral health. *Bull World Health Organ* **83**, 661-669, (2005).

22 Quince, C. *et al.* Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* **6**, 639-641, (2009).

23 Dewhirst, F. E. *et al.* The human oral microbiome. *J Bacteriol* **192**, 5002-5017, (2010).

24 Lazarevic, V. *et al.* Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *J Microbiol Methods* **79**, 266-271, (2009).

25 Keijser, B. J. *et al.* Pyrosequencing analysis of the oral microflora of healthy adults. *J Dent Res* **87**, 1016-1020, (2008).

26 La Duc, M. T., Kern, R. & Venkateswaran, K. Microbial monitoring of spacecraft and associated environments. *Microb Ecol* **47**, 150-158, (2004).

27 Contreras, M. *et al.* The bacterial microbiota in the oral mucosa of rural Amerindians. *Microbiology* **156**, 3282-3287, (2010).

28 Beier, S., Witzel, K. P. & Marxsen, J. Bacterial community composition in Central European running waters examined by temperature gradient gel electrophoresis and sequence analysis of 16S rRNA genes. *Appl Environ Microbiol* **74**, 188-199, (2008).

29 Schloss, P. D. & Handelsman, J. Toward a census of bacteria in soil. *PLoS Comput Biol* **2**, e92, (2006).

30 Ellis, R. J., Morgan, P., Weightman, A. J. & Fry, J. C. Cultivation-dependent and -independent approaches for determining bacterial diversity in heavy-metal-contaminated soil. *Appl Environ Microbiol* **69**, 3223-3230, (2003).

31 Nogales, B. *et al.* Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. *Appl Environ Microbiol* **67**, 1874-1884, (2001).

32 Elshahed, M. S. *et al.* Novelty and uniqueness patterns of rare members of the soil biosphere. *Appl Environ Microbiol* **74**, 5422-5428, (2008).

33 Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308**, 554-557, (2005).

34 Will, C. *et al.* Horizon-Specific Bacterial Community Composition of German Grassland Soils, as Revealed by Pyrosequencing-Based Analysis of 16S rRNA Genes. *Applied and Environmental Microbiology* **76**, 6751-6759, (2010).

35 Lepp, P. W. *et al.* Methanogenic Archaea and human periodontal disease. *Proc Natl Acad Sci U S A* **101**, 6176-6181, (2004).

36 Dahlen, G., Naucler, C., Nordwall, S. & Suksu-art, N. Oral microflora in betel-chewing adults of the Karen tribe in Thailand. *Anaerobe* **16**, 331-336, (2010).

37 Kerr, N. W. Prevalence and natural history of periodontal disease in prehistoric Scots (pre-900 AD). *J Periodontal Res* **33**, 131-137, (1998).

38 Logan, E. I. Dietary influences on periodontal health in dogs and cats. *Vet Clin North Am Small Anim Pract* **36**, 1385-1401, ix, (2006).

39 Albandar, J. M., Brunelle, J. A. & Kingman, A. Destructive periodontal disease in adults 30 years of age and older in the United States, 1988-1994. *J Periodontol* **70**, 13-29, (1999).

40 Petchey, O. & Gaston, K. Effects on ecosystem resilience of biodiversity, extinctions, and the structure of regional species pools. *Theoretical Ecology* **2**, 177-187, (2009).

41 Loreau, M. *et al.* in *Biodiversity and Ecosystem Functioning. Synthesis and Perspectives* (eds M Loreau, S Naeem, & P Inchausti) 79-91 (Oxford University Press, 2002).

42 Addy, M., Griffiths, G., Dummer, P., Kingdom, A. & Shaw, W. C. The distribution of plaque and gingivitis and the influence of toothbrushing hand in a group of South Wales 11-12 year-old children. *J Clin Periodontol* **14**, 564-572, (1987).

43 Stewart, J. E. & Wolfe, G. R. The retention of newly-acquired brushing and flossing skills. *J Clin Periodontol* **16**, 331-332, (1989).

44 Loe, H. & Schiott, C. R. The effect of mouthrinses and topical application of chlorhexidine on the development of dental plaque and gingivitis in man. *J Periodontal Res* **5**, 79-83, (1970).

45 McBain, A. J. *et al.* Effects of a chlorhexidine gluconate-containing mouthwash on the vitality and antimicrobial susceptibility of in vitro oral bacterial ecosystems. *Appl Environ Microbiol* **69**, 4770-4776, (2003).

46 Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335-336, (2010).

47 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461, (2010).

48 Schloss, P. D. The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput Biol* **6**, e1000844, (2010).

49 Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173**, 697-703, (1991).

50 Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**, 5261-5267, (2007).

51 Faith, D. P. Conservation evaluation and phylogenetic diversity. *VBiological Conservation* **61**, 1-10, (1992).

52 Yoshida, A. *et al.* Development of a 5' nuclease-based real-time PCR assay for quantitative detection of cariogenic dental pathogens Streptococcus mutans and Streptococcus sobrinus. *J Clin Microbiol* **41**, 4438-4441, (2003).

53 Yoshida, A. *et al.* Development of a 5' fluorogenic nuclease-based real-time PCR assay for quantitative detection of Actinobacillus actinomycetemcomitans and Porphyromonas gingivalis. *J Clin Microbiol* **41**, 863-866, (2003).

54 Cortelli, S. C. *et al.* Clinical status and detection of periodontopathogens and Streptococcus mutans in children with high levels of supragingival biofilm. *Braz Oral Res* **23**, 313-318, (2009).

55 Choi, B. K. *et al.* Detection of major putative periodontopathogens in Korean advanced adult periodontitis patients using a nucleic acid-based approach. *J Periodontol* **71**, 1387-1394, (2000).

56 Mayanagi, G., Sato, T., Shimauchi, H. & Takahashi, N. Detection frequency of periodontitis-associated bacteria by polymerase chain reaction in subgingival and supragingival plaque of periodontitis and healthy subjects. *Oral Microbiol Immunol* **19**, 379-385, (2004).

57 Hintao, J., Teanpaisan, R., Chongsuvivatwong, V., Ratarasan, C. & Dahlen, G. The microbiological profiles of saliva, supragingival and subgingival plaque and dental

caries in adults with and without type 2 diabetes mellitus. *Oral Microbiol Immunol* **22**, 175-181, (2007).

SUPPLEMENTARY MATERIAL SECTION 1. Sample information, preparation and authentication

Sample Description

We analysed microbial DNA in dental calculus samples from 46 individuals and 70% contained oral microbes. This was determined by PCR amplification using oral specific and universal 16S rRNA gene primers (Table S1.2). Based on the PCR amplification results, 28 of the highest-quality samples were sequenced using 454 pyrosequencing technology.

The dental calculus samples that were sequenced ranged in age from the early Neolithic (5500 BC) to the high Medieval (1550 AD) (Table S1.1). This included samples from the earliest farming community in central Europe (Germany), the Linear Pottery Culture (LBK), as well as samples from late Neolithic (Bell Beaker) and early Bronze Age cultures. Furthermore, later farming groups from England were sampled, including Bronze Age material, and rural (Raunds) and urban (Jewbury and St Helen on the Walls) populations from the Medieval period. The use of dental calculus samples from two regions, England and Germany, was dictated by the availability of specimens. Although using samples from different regions was not ideal, this was not thought to have influenced the results, as all seven groups examined were from pre-Industrial, European populations that practiced similar forms of farming[1].

Genetic analysis was performed on different types of dental calculus deposits; supragingival (32%) and subgingival (11%) calculus, and samples which contained both supra- and subgingival calculus (36%) (Table S1.1). For the majority of groups investigated (five out of seven), both supra- and subgingival calculus deposits or a combination of both, were used for genetic analysis. While the composition of microbes in supra- and subgingival calculus has been found to be similar, the abundance of taxa does vary[2]. Therefore, abundances were not considered in phylogenetic analyses.

Ancient DNA work conditions

We performed all ancient DNA at the Australian Centre for Ancient DNA (ACAD) in a dedicated, physically isolated laboratory, which is approximately a 15 minute walk from the laboratory used for post-PCR analyses. Strict protocols were followed to minimise the amount of human DNA in the ancient DNA laboratory. These included the wearing of freshly laundered clothes, a full body suit, shoe covers, boots, facemask, face shield and triple gloving. Furthermore, there is no personnel movement from the post-PCR environment to the ancient DNA laboratory within a single day. All surfaces in the ancient DNA laboratory are

routinely triple-wiped with bleach, decon and isopropanol. The laboratory is irradiated with ultraviolet (UV) light (30–50 nm) overnight for approximately 4 hours. All consumables, disposables, tools and instruments are externally bleached and UV irradiated before entering the lab and then subjected to routine cleaning before, during and after use.

Authentication Criteria

Due to the high risk of contamination when dealing with human remains for ancient microbial genetic analysis, precautions were taken to prevent contamination with modern DNA. (*i*) The samples were either collected under DNA-free conditions during excavation, which included no washing, treating or examining the samples (German material) or at the least, the samples had been minimally handled (i.e. the English material had been excavated by author K.D). (*ii*) All preparation and analytical steps prior to DNA amplification were conducted in a clean-room area solely dedicated to bacterial and environmental work, located in a physically separated building without any modern DNA work (pre-PCR area). Amplification, cloning and sequencing were done in the post-PCR lab. (*iii*) A small number of samples were extracted on a single day (e.g. on average five per day). (*iv*) Contamination was monitored by running non-template controls in parallel, including extraction and PCR blanks. (*v*) Independent extractions were not possible due to the small size of samples, with commonly only one calculus deposit per individual available for DNA analysis. (*vi*) PCR reagents were enzymatically treated to remove microbial contamination prior to PCR amplification. (*vii*) A minimum of eight independent PCR reactions were done per dental calculus sample. Duplicate PCR amplifications were performed on separate days. (*viii*) Pyrosequencing was used to provid in-depth coverage of the microbial contents of dental calculus samples and non-template controls. (*ix*) Sequences were compared to known bacterial DNA sequences, using the Ribosomal Database Project (RDP) and Human Oral Microbiome Database (HOMD) databases, and assessed for sequence and PCR errors using the denoise[3] algorithm.

Sample Preparation

Dental calculus samples, if not already removed from teeth, were isolated using a tooth scalar. The dental calculus samples were decontaminated by exposing the calculus deposits to high-intensity UV irradiation for 20 minutes, followed by immersion in bleach (2% (w/v) sodium hypochlorite) for 5 minutes and then soaking in ethanol (100%) for 5 minutes. To assess whether mechanically removing the outer surface of calculus deposits assisted decontamination, we split a large calculus sample from one individual, and removed the surface from one half of the deposit, while the other half was immersed in bleach. Both forms of decontamination resulted in the recovery of equivalent suites of bacterial taxa. Therefore,

to decontaminate samples we immersed the calculus deposits in bleach and did not mechanically remove the outer surface. By only immersing the samples in bleach, we reduced the amount of sample handling and loss of material compared to if the outer surface had been removed. The sterilised dental calculus samples were crushed to produce a non-uniform powder for DNA extraction.

NOTE:
This figure is included on page 158 of the print copy of the thesis held in the University of Adelaide Library.

Figure S1.1. Calculus from the macroscopic to microscopic scale. a) The right maxilla of sample 8482 from the medieval UK population Jewbury, presenting a large buccal deposit of supra and subgingival calculus, which covers multiple teeth; b) another specimen from the same population, showing remarkable microscopic incremental growth lines in calculus, magnification ×50 (scanning electron micrograph); c) further specimen showing a large unidentified 'golf ball-shaped' colony of rod-shaped bacteria, magnification ×3000 (scanning electron micrograph). 1b) and c) are courtesy of K. Dobney (in Lilley et al. 1994).

Table S1.1. Archaeological, anatomical and anthropological sample description.

Sample ID	Museum#	Group/Culture	Period	Location
8215	HK2000:4 083a, 613.1	Linear Pottery Culture	Neolithic (5450-4775 cal BC)	Halberstadt-Sonntagsfeld, Germany
8240	HK2000:4 228a, 861	Linear Pottery Culture	Neolithic (5450-4775 cal BC)	Halberstadt-Sonntagsfeld, Germany
8247	HK2000:4 233a, 870	Linear Pottery Culture	Neolithic (5450-4775 cal BC)	Halberstadt-Sonntagsfeld, Germany
8275	HK2000:7 374a, 1324	Linear Pottery Culture	Neolithic (5450-4775 cal BC)	Halberstadt-Sonntagsfeld, Germany
8277	HK2000:4 014b, 413.1	Linear Pottery Culture	Neolithic (5450-4775 cal BC)	Halberstadt-Sonntagsfeld, Germany
4331	HK2004:9 463a, 6255.1	Bell Beaker	Neolithic (5450-4775 cal BC)	Quedlinburg XII, Germany
9436	HK, 43	Late Neolithic/early Bronze Age	Late Neolithic/early Bronze Age	Benzingerode-Heimburg, Germany
8890	T82GF	Yorkshire BronzeAge	Bronze Age	Yorkshire, England
8891	14Barrow 163	Yorkshire BronzeAge	Bronze Age	Yorkshire, England
8894	T98	Yorkshire BronzeAge	Bronze Age	Yorkshire, England
8326	2095	Jewbury	Late Medieval (1200-1300AD)	York, England
8330	2106	Jewbury	Late Medieval (1200-1300AD)	York, England
8332	4440	Jewbury	Late Medieval (1200-1300AD)	York, England
8477	2357	Jewbury	Late Medieval (1200-1300AD)	York, England
8482	2654	Jewbury	Late Medieval (1200-1300AD)	York, England
8814	4161	Jewbury	Late Medieval (1200-1300AD)	York, England
8863	4485	Jewbury	Late Medieval (1200-1300AD)	York, England
8333	R5287	Raunds Furnells	Early Medieval (850–1100AD)	Northamptonshire, England
8335	R5252	Raunds Furnells	Early Medieval (850–1100AD)	Northamptonshire, England
8337	R5136	Raunds Furnells	Early Medieval (850–1100AD)	Northamptonshire, England
8341	R5206	Raunds Furnells	Early Medieval (850–1100AD)	Northamptonshire, England
8868	R5157	Raunds Furnells	Early Medieval (850–1100AD)	Northamptonshire, England
8869	R5229	Raunds Furnells	Early Medieval (850–1100AD)	Northamptonshire, England
8873	5228	St Helens on the Wall, Aldwark	Late Medieval (950–1550AD)	York, England
8874	5241	St Helens on the Wall, Aldwark	Late Medieval (950–1550AD)	York, England
8877	5113	St Helens on the Wall, Aldwark	Late Medieval (950–1550AD)	York, England
8878	5203	St Helens on the Wall, Aldwark	Late Medieval (950–1550AD)	York, England
8883	5244	St Helens on the Wall, Aldwark	Late Medieval (950–1550AD)	York, England

Table S1.1 continued. Archaeological, anatomical and anthropological sample description.

Sample ID	Dental calculus location	Type of dental calculus	Sex	Age (years)	Skeletal pathologies
8215	M38	Unknown	M	50-65	Unknown
8240	M48	Unknown	M	35-45	Unknown
8247	M55	Unknown	F?	3-4	Unknown
8275	M36	Unknown	F	40-60	Unknown
8277	M18	Unknown	M	30-35	Unknown
4331	C33	buccal supra/subgingival	F	55-65	Unknown
9436	M38	Unknown	M	35-50	Unknown
8890	M1	supra/subgingival	U	U	Unknown
8891	M1	buccal supra/subgingival	U	U	Unknown
8894	C	buccal supra/subgingival	U	U	Unknown
8326	C33	lingual supragingival	F	35-40	Overall findings for the Jewbury population[4]; Anaemia present in 22% of the population, periostitis in 15.5%, enamel hypoplasia in 45.7%, caries in 59.5% and periodontal disease in over 80% of individuals.
8330	I41	buccal subgingival	F	30-35	
8332	I22	lingual supragingival	F	20-30	
8477	I31	lingual supra/subgingival	F	30-40	
8482	PM11	buccal subgingival	M	20-25	
8814	M3	supragingival	F	20-25	
8863	I41	lingual supragingival	U	9-14	
8333	I2	buccal supragingival	U	U	Overall findings for the population at Raunds[5]; Cribra orbitalia (associated with anaemia) present in 29.3%, osteoarthritis present nearly 100% of individuals over the age of 17 years old, caries present in 30% and periodontal disease in 49.7% of the population.
8335	M3	lingual supra/subgingival	U	U	
8337	M1	buccal supra/subgingival	U	U	
8341	M3	lingual supra/subgingival	U	U	
8868	I1	lingual and buccal supragingival	U	U	
8869	M2	buccal supragingival	U	U	
8873	M18	occlusal supragingival	F	>60	Widespread arthritis, ante-mortem tooth loss and three dental abscesses[6]
8874	M1	lingual supragingival	F	middle aged	Slight arthritis on the foot[6]
8877	I2	buccal supra/subgingival	M	young	Periostitis of left femur, fibulae, humerus and clavicle, and both tibiae[6]
8878	I1	lingual subgingival	M	>45	Arthritis of the neck and ante-mortem tooth loss[6]
8883	I1	buccal supra/subgingival	M	>60	Dental abscesses and caries[6]

Table S1.2. Specific and universal primer details.

Primer name/position	Primer Sequence 5'-3'	Target gene
Specific Primers		
S. mutans 3368F	GCCTACAGCTCAGAGATGCTATTCT	gtfB[7]
S. mutans 3481R	GCCATACACCACTCATGAATTGA	
P. gingivalis 1198F	TACCCATCGTCGCCTTGGT	16S rRNA[8]
P. gingivalis 1323R	CGGACTAAAACCGCATACACTTG	
Universal Primers		
V1 8F	AGAGTTTGATYMTGGCTCAG	16S rRNA
V1 120R	TTACTCACCCGTICGCCRCT	
V3 338F	ACTCCTACGGGAGGCAGCAGT	16S rRNA
V3 533R	TTACCGCGGCTGCTGGCAC	
V6 926F	AAACTYAAAKGAATTGACGGG	16S rRNA
V6 1046R	CGACARCCATGCASCACCT	
M13F	ACTGGCCGTCGTTTTACAA	StrataClone™ PCR Cloning Vector pSC-A
M13R	GGAAACAGCTATGACCATG	

Table S1.3. Sample specific barcodes for the three regions of 16S rRNA gene investigated. The barcode is located on the forward primer between the 454-A-adaptor and PCR primer, which is specific for the 16S rRNA gene. The barcodes were created by the Human Microbiome Project (http://www.hmpdacc.org/tools_protocols/tools_protocols.php).

Sample ID	Barcode Sequence for V1	Barcode Sequence for V3	Barcode Sequence for V6
8215	TTCGTGGC	TTGAACTC	TCGAGGAAC
8240	AACCTGGC	CCTTCCGC	AACTGTTC
8247	TCGTTGTC	TGGTTGGTC	AACGGAGTC
8275	ACGAAGTC	AGTCCGTC	TTCTCAAC
8277	ACGAGAAC	AACGAGGC	ACCGGAAGC
4331	TGGTGAAC	CCGTTCAC	ACGTTCCAC
9364	TTGTGTTC	TGTCCGGTC	TTCGTTATC
8890	CCGGCCAC	ACCTGAAC	TTGGAGGC
8891	AATGGTAC	AAGAGTTC	TTATCGGC
8894	TCTCCGTC	ATTCGTAC	AAGAAGAC
8326	AGTTGGC	AAGCCGC	TATCAAC
8330	AGCTTC	ACTCAC	CGACTC
8332	TCTTGGC	TCCGCTC	ACGGCTC
8477	CGGTATC	CAAGAAC	AGGCGGC
8482	TGACGAC	AGACCTC	ATCTTAC
8814	ACAAGGC	ATACCAC	AACCAGC
8863	AAGGTGC	TCGCGGC	TTCGAGC
8333	ACGCGC	CCTCTC	AGACAC
8335	TAATCTC	TATTGAC	ACACGGAC
8337	TCACCTC	AGTCGAC	TGCCGAAC
8341	CCAGGAC	AACAACTC	CGGTCTTC
8868	TGCGTTC	CGTCGTC	TTGACAAC
8869	TCTCGAC	AAGGCAC	TCCAGAAC
8873	ACTCCTC	AACACAAC	AAGGCCTC
8874	TTCCTGC	TTCTTGAC	ACTAATTC
8877	TTCATAC	TCCAAGTC	TGACCGTC
8878	TATTCGTC	TTCGCGAC	TGTCGGAC
8883	TAGGAATC	CCGGTCGC	AGGTTGTC
EBC1	CCACGGTC	AACCGTGTC	ACCGTAATC
EBC2	TTAAGATTC	CTTCCTTC	ACCTTGGTC

Table S1.4. Number of sequences per sample for the three regions of 16S rRNA gene amplified after quality filtering and denoising.

	V1		V3		V6	
Sample ID	Quality filtered	Quality filtered and denoised	Quality filtered	Quality filtered and denoised	Quality filtered	Quality filtered and denoised
8215	9913	5037	2533	1266	3940	2016
8240	16369	8178	3885	1988	6890	3479
8247	54351	27168	1609	792	4963	2430
8275	10935	5438	1905	961	5145	2657
8277	16370	8159	2790	1423	7474	3741
4331	55658	27643	1843	935	2750	1425
9364	15134	7755	2088	1061	7356	3760
8890	39099	19705	1938	947	3233	1689
8891	14947	7389	7724	3969	4491	2364
8894	9284	4667	1813	946	2495	1287
8326	14710	7359	2136	1115	6493	3290
8330	13513	6727	6438	3234	5151	2592
8332	5838	2904	2742	1396	5783	2869
8477	5350	2680	5734	2885	3790	1873
8482	3317	1696	3149	1628	6299	3180
8814	13987	7009	4165	2143	10175	5091
8863	8498	4295	4163	2138	18402	9492
8333	5527	2778	2974	1469	5365	2649
8335	36233	18221	4603	2380	2759	1401
8337	6646	3346	2810	1433	2172	1115
8341	20337	10248	9080	4628	3325	1669
8868	25137	12636	2634	1308	5461	2707
8869	30129	15289	5958	3036	6151	3084
8873	12005	5932	3284	1645	3129	1587
8874	18353	9496	6926	3613	3526	1799
8877	19570	9911	6243	3168	5863	2974
8878	14304	7183	2992	1533	8197	4208
8883	22556	11315	3197	1683	3855	1951
EBC1	9802	4911	1217	603	6569	3320
EBC2	13137	6696	1455	761	1742	870
Total	541009	271771	110028	56087	162944	82569

SUPPLEMENTARY MATERIAL SECTION 2. The variation in phyla between the three hypervariable regions of 16S rRNA gene examined.

The composition of phyla in ancient dental calculus was similar, independent of the region of 16S rRNA gene investigated (Figure S2.1). To assess differences in the microbial composition of the three hypervariable regions examined, the relative abundance of phyla in V1, V3 and V6 was compared using a nonparametric permutation based *t*-test in the proGram Metastats[9]. The frequency of unclassified bacteria varied non-randomly between the three regions of 16S rRNA gene. The shortest regions (V1 and V6) had more unclassified bacteria than the longest fragment, V3 (V1 vs. V3 $p = 0.0005$; V6 vs. V3 $p = 0.0005$). We expected to observe a higher portion of unclassified bacteria in the shorter sequences, as they have less phylogenetic information compared to the longer sequences. Aside from length, the sequences may have also been designated as unclassified due to post-mortem DNA damage. It is unlikely that the unclassified bacteria represent genuine 16S rRNA gene sequences which have altered over time. The accumulation of a large number of mutations in the 16S rRNA gene would be unlikely over the time frame examined in this study. A maximum of 0.006075 mutations could be expected over the longest fragment of DNA (180 bp) amplified from the oldest dental calculus sample (~7,500 years before present), assuming a within-lineage mutation rate of 0.0045 mutations per site per million years[10].

A general trend across all the regions of 16S rRNA gene examined was low abundance of Gram-negative taxa. Linear regression analyses were used to assess variation in abundance of Gram-negative bacteria with increasing ‘time depth’. Regression analyses revealed evidence of a trend ($p = 0.006$), with Gram-negative phyla decreasing in frequency as the ‘time since death’ of samples increased.

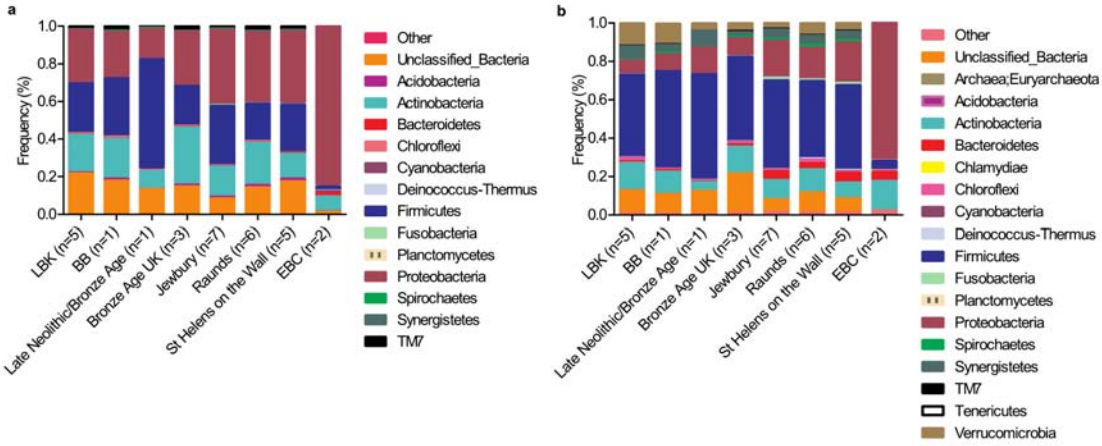

Figure S2.1. The microbial phyla in ancient dental calculus deposits and non-template controls (EBC) for the V1 (a) and V6 (b) regions of 16S rRNA gene investigated. Abbreviations: LBK (Linear Pottery Culture), BB (Bell Beaker) and EBC (Extraction blank controls).

SUPPLEMENTARY MATERIAL SECTION 3. -diversity, phylogeny-based analyses.

We used -diversity[11] analyses to compare the ancient dental calculus data with modern oral and environmental microbial sequences. This was performed to assess the relationship between past and present oral samples and test for potential contamination with soil microbes. The modern oral and environmental sequences were compiled from studies that sequenced by cloning[12-20] and pyrosequencing[21,22]. The modern data encompassed sequences from 30 oral samples, including saliva[21], and plaque, which was obtained from healthy[12], caries[13] and periodontal disease[14] sites, in addition to 16 environmental samples, including soils[15-19,22], sediments and freshwater[20]. We used clone sequences of near to the full-length of 16S rRNA gene because they covered all regions of 16S rRNA gene examined (V1-V6) in our study. This enabled the creation of a modern dataset which was directly comparable to each region of 16S rRNA gene investigated. The modern sequences were trimmed to V1, V3 and V6, and only sequences which completely covered each region were retained. Alignments of ancient and modern sequences for each hypervariable region were generated using PyNAST[23].

Phylogeny-based metrics were used to calculate -diversity. We opted to estimate -diversity using phylogeny based methods, as opposed to using operational taxonomic unit- (OTU) based calculations, because of the robustness of phylogenetic analyses to sequencing and PCR errors[24]. Phylogenetic trees for V1, V3 and V6 were produced from the alignments of ancient and modern sequences generated for each hypervariable region. The phylogenies were inferred using Maximum likelihood in RAxML (version 7.0.4, CIPRES webserver[25]). We computed the UniFrac distances[26] for each phylogeny and used these distances for the Principal Co-ordinates Analysis (PCoA) in the QIIME software package[23]. Due to an underrepresentation of Gram-negative taxa in the ancient samples, all UniFrac distances calculated were unweighted by abundance. Additionally, UniFrac distances were calculated using both the entire dataset and a restricted dataset containing only Gram-positive lineages. For each hypervariable region examined, samples were clustered based on their between sample distance using UPGMA and jackknifing of the UniFrac distances. Jackknifing involved re-sampling sequences 100 times with replacement at a depth of 10 sequences per sample. This low number of sequences per sample was due to the small number of sequences obtained from the published modern clone library data. We used a non-parametric, multivariate analysis of variance (NP-MANOVA) to assess if phylogenetic variation existed between clusters. This was performed on the unweighted UniFrac distances in the program PAST (version 2.07), by specifying a user generated distance matrix, on which 100,000 permutations were run. The reported values have been corrected for the number of pairwise

comparisons performed (Bonferroni-corrected) (Table S3.1). Whether all the sequence data or a restricted subset of only Gram positive lineages was used, the ancient dental calculus and modern oral samples still showed non-random differences compared to the environmental material and extraction blanks (Table S3.1).

We compared ancient dental calculus to modern oral samples, namely plaque and saliva. Analogous published data for modern dental calculus was not available. However, culture studies have shown there is a close microbial relationship between calculus and plaque[27]. To discover if the microbial makeup of calculus and plaque was similar when using genetic analysis, we genetically examined the composition of calcified and non-calcified forms of plaque in two modern European male volunteers. As described in the Materials and Methods, DNA was extracted from the calculus and plaque samples and amplified using the V3 region of the 16S rRNA gene (Table S1.2). The universal primer amplicons were cloned and a total of 90 clones for the four samples were Sanger sequenced. The calculus and plaque samples were dominated by Firmicutes (31%), particularly *Streptococcus* (17%). The microbial makeup of calculus and plaque was compared using phylogeny-based analysis of -diversity (UniFrac metric). The UniFrac analysis revealed no evidence of differences between the calcified and non-calcified forms of plaque ($p = 0.3324$, NP-MANOVA). While preliminary, this suggests that plaque provides a representative picture of calculus diversity.

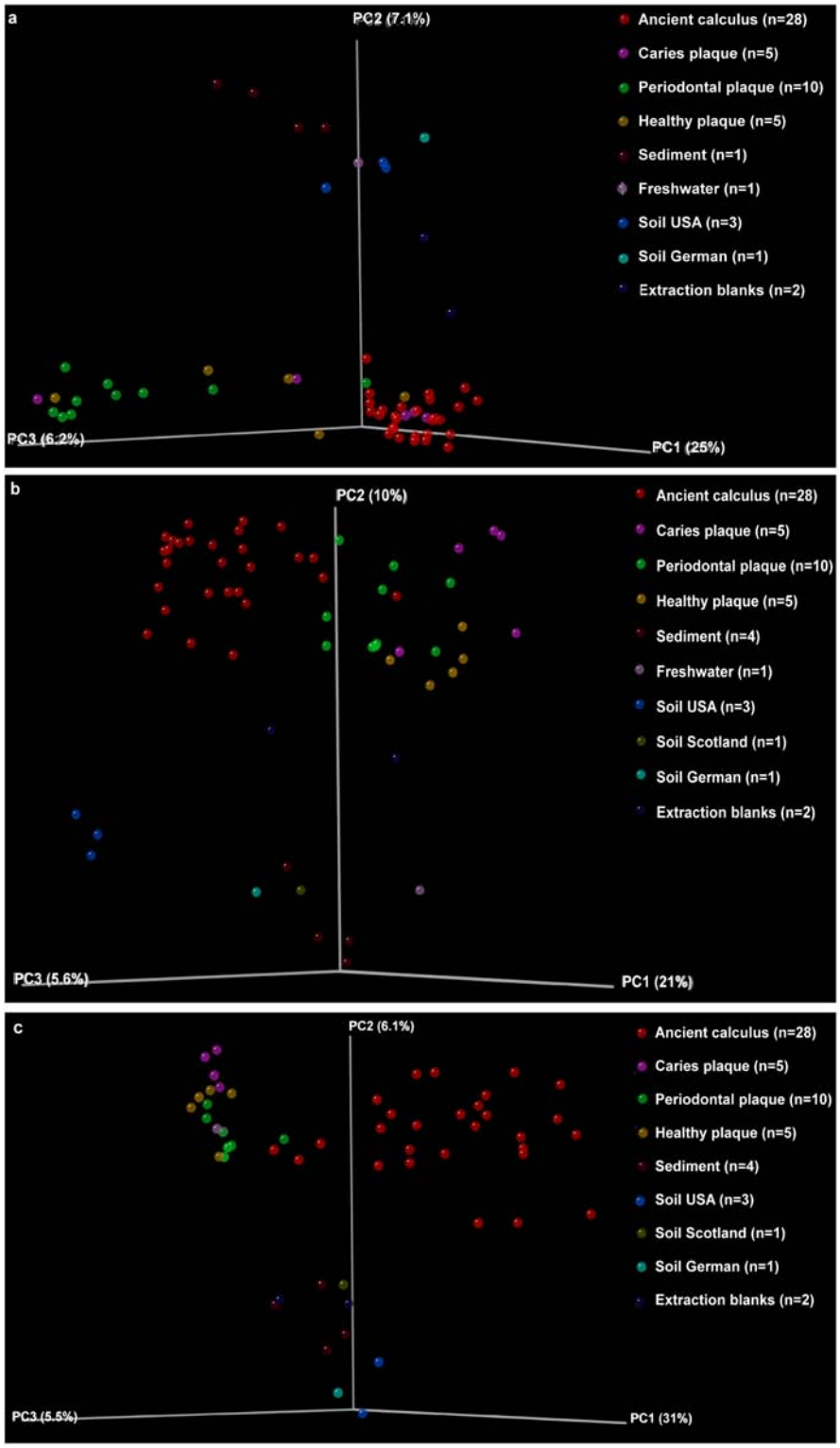

Figure S3.1. PCoA plot of -diversity reveals a close phylogenetic relationship between the ancient dental calculus and modern oral samples, both of which are distinct from the non-template controls and environmental samples. This pattern was replicated in each region of 16S rRNA gene investigated; (a) V1, (b) V3 and (c) V6. Ancient calculus samples were compared to modern oral[12-14,21] and environmental [15-19] clone data. The clone sequences partially covered the 16S rRNA gene and were trimmed to each region of interest prior to each analysis.

Table S3.1. Assessment of phylogenetic variation between ancient and modern microbial sequences, using a non-parametric permutation based test. The NP-MANOVA's were calculated using the unweighted UniFrac distances estimated from the V3 region for the whole dataset and a restricted subset of only Gram-positive taxa. We performed the NP-MANOVA's using 100,000 permutations and the grey-shaded values indicate the Bonferroni-corrected values.

V3 Beta diversity-complete dataset	Archaeological calculus/modern saliva&plaque	Soil/sediment/water	Extraction blanks
Archaeological calculus/modern saliva&plaque	0	0	0.00827
Soil/sediment/water	0	0	0.04149
Extraction blanks	0.02481	0.1245	0
V3 Beta diversity-Gram-positive lineages only	**Archaeological calculus/modern saliva&plaque**	**Soil/sediment/water**	**Extraction blanks**
Archaeological calculus/modern saliva&plaque	0	0	0.0009
Soil/sediment/water	0	0	0.3822
Extraction blanks	0.0027	1	0

SUPPLEMENTARY MATERIAL SECTION 4. The change in diversity and composition of oral microbiota since the introduction of agriculture in Europe.

Variation in oral microbiota over time was assessed by comparing the diversity and composition of bacteria over four broad-range periods: Neolithic (7500 – 7000 years BP), Bronze Age (4500 – 3500 years BP), Medieval (1100 – 400 years BP), and today. For each time period the -diversity (within group diversity) was measured from only classified bacterial sequences, using both OTU (Shannon Index) and phylogeny-based metrics. Both methods showed a decrease in diversity between the Medieval and the present day (Figure S4.1 and Table S4.1). The Shannon Index was used to estimate -diversity directly from OTU abundance data in the QIIME software package[23]. We compared the Shannon Index, diversity values between each time period, by converting the diversity values to a distance matrix (Bray-Curtis) and performing an NP-MANOVA (Table S4.1). The Shannon Index was calculated for both the entire dataset and only Gram-positive lineages. -diversity was also measured phylogenetically[28], in a parsimony-based fashion in QIIME[23]. Phylogenetic analyses of -diversity are less susceptible to rare variants in the sequences[28] which might be caused by DNA damage. As the ancient and modern data were sequenced at various depths, rarefaction was performed. This involved performing ten sampling repetitions without replacement, at each sequencing depth (number of sequences). As our dataset was small, we bootstrapped the -diversity values generated for each time point, at a range of sampling depths (15-45 sequences). A decrease in diversity over time was observed independent of the number of sequences examined (Figure S4.1).

We found that both the diversity and composition of oral microbiota varied over time (Figure S4.2). Rank abundance curves were generated for each time period (Neolithic, Bronze Age, Medieval and today) using the relative abundance of classified genera. These plots revealed a relatively stable composition of the most dominant bacteria in dental calculus between the Neolithic and Medieval period, with large changes observed between the Medieval and the present day (Figure S4.2). In particular, in modern populations compared to early farming groups, there is an increase in caries-associated bacteria and decrease in periodontal disease-associated microbes.

It was necessary to investigate whether the findings were biased by sampling dental calculus, the presence of which has been associated with the development of periodontal disease[29,30].

To assess whether the ancient results were biased, archaeological calculus was compared to only modern individuals with periodontal disease. The results do not appear to be biased, as the prehistoric samples still showed a higher diversity (Table S4.1) and lower frequency of *Streptococcus* (Figure S4.2) than modern individuals with periodontal disease.

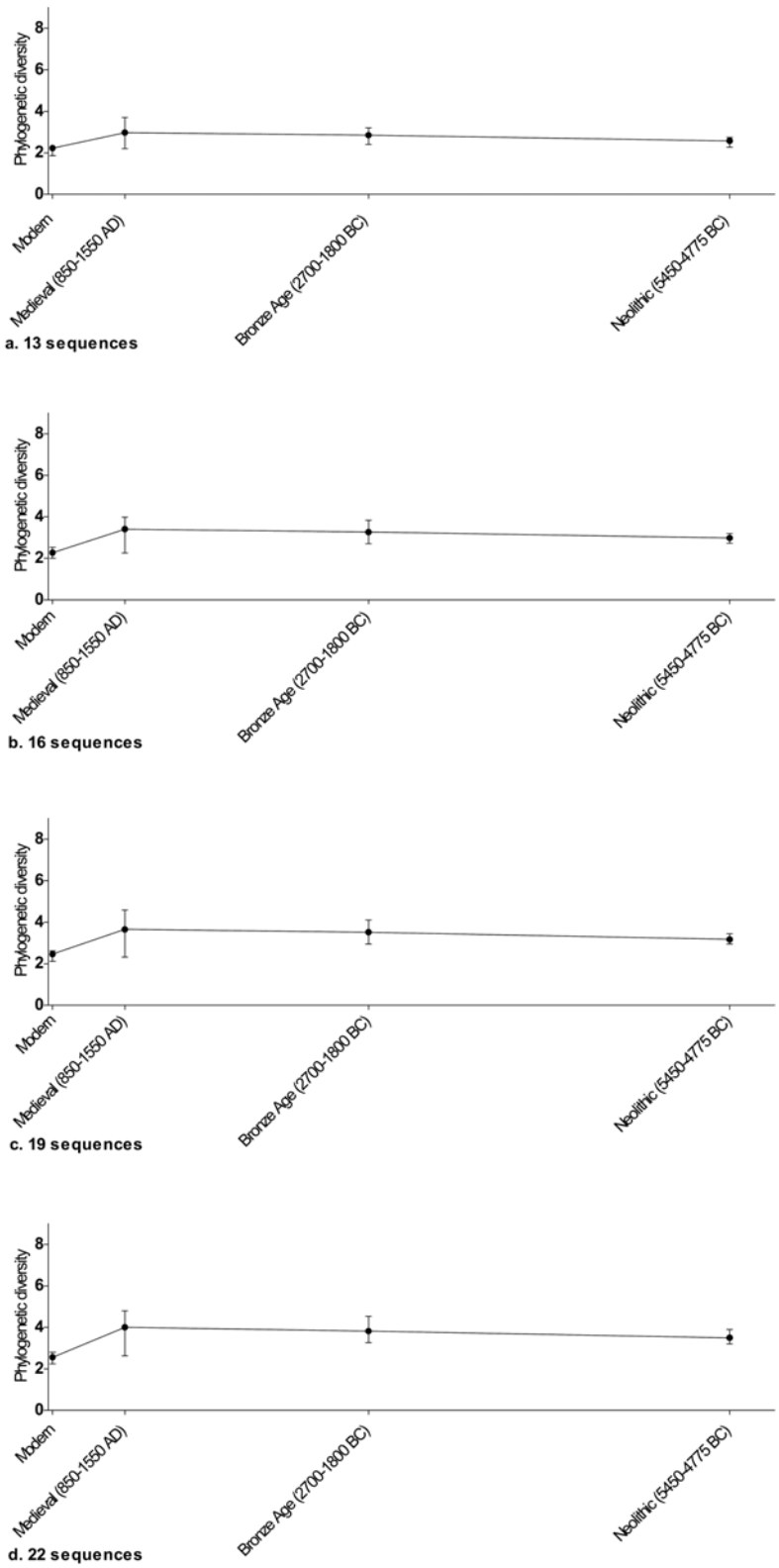

Figure S4.1. Decrease in -diversity over time, shown at a range of sequencing depths (a-d). -diversity for each time period was calculated from phylogenetic analysis[28] of the V3 region and the results were bootstrapped. The phylogenetic diversity of the archaeological dental calculus samples (n = 28) was compared to modern plaque (n = 20)[12-14] and saliva (n = 5)[21].

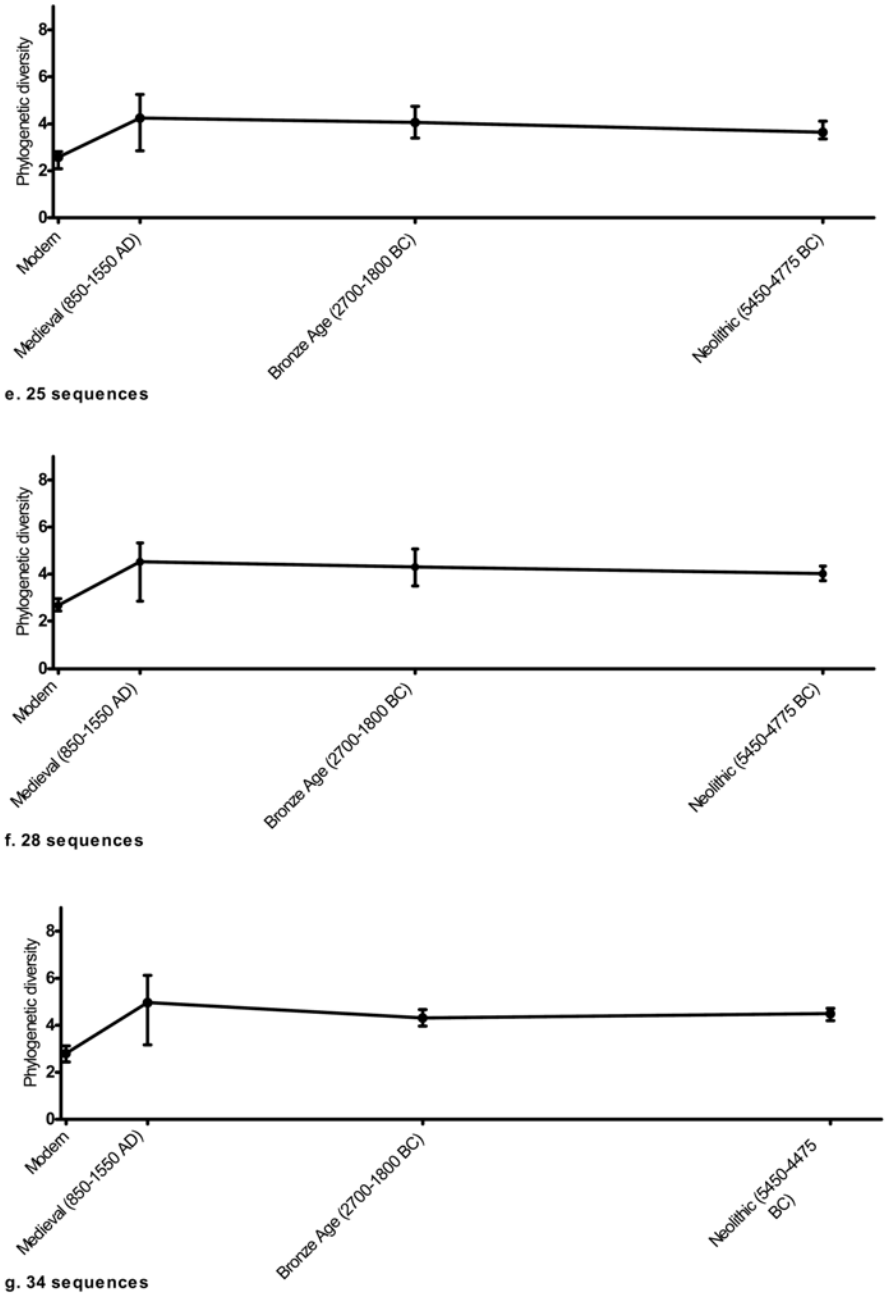

Figure S4.1, continued. Decrease in -diversity over time, shown at a range of sequencing depths (f-g). -diversity for each time period was calculated from phylogenetic analysis[28] of the V3 region and the results were bootstrapped. The phylogenetic diversity of the archaeological dental calculus samples (n = 28) was compared to modern plaque (n = 20)[12-14] and saliva (n = 5)[21].

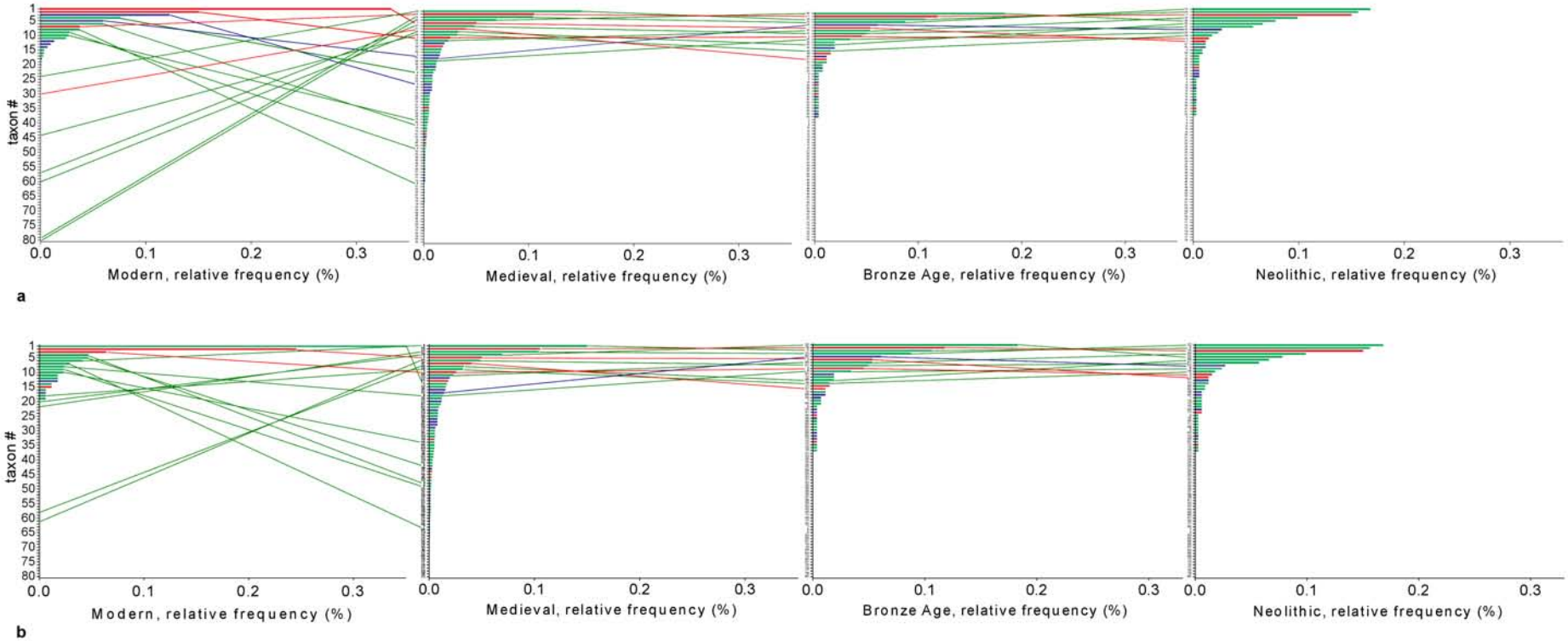

Figure S4.2. Rank abundance curves of 80 genera and the changing relative frequency of those genera over time. The relative frequency of the genera was determined from the V3 region and excluded unclassified bacteria. The bars are coloured according to disease association; red for caries, green for periodontal disease and blue for healthy lineages. For the ten most abundant genera, connecting lines indicate the position of that genus at adjacent time points. This highlights that between the Medieval and today there has been a major change in the composition of oral microbiota. (a) Compares the ancient samples (n = 28) to modern saliva and plaque (n = 25)[12-14,21] from individuals with caries, periodontal disease and in health. The corresponding taxa to the taxon numbers are presented in Table S4.1 (b) Restricts the comparison of ancient calculus samples to only modern individuals with periodontal disease (n = 10)[14]. The corresponding taxa to the taxon numbers are presented in Table S4.2

Table S4.1. Taxon numbers from the rank abundance curve (Figure S4.2a) of the complete ancient and modern data set.

Taxon Rank	Modern	Medieval	Bronze Age	Neolithic
1	*Streptococcus*	*Clostridiaceae*	*TM7_genera_incertae_sedis*	*Synergistaceae*
2	*Veillonella*	Actinobacteria	Actinobacteria	*Clostridiaceae*
3	*Haemophilus*	*Synergistaceae*	*Clostridiaceae*	Actinobacteria
4	*Prevotellaceae*	*TM7_genera_incertae_sedis*	*Synergistaceae*	*Clostridiales;Incertae Sedis XI*
5	*Neisseria*	*Streptococcus*	*Ruminococcaceae*	*TM7_genera_incertae_sedis*
6	*Granulicatella*	*Chloroflexi*	*Clostridiales;Incertae Sedis XI*	*Clostridiales;Incertae Sedis XIII*
7	Actinobacteria	*Leptotrichia*	*Streptococcus*	*Chloroflexi*
8	*Megasphaera*	*Clostridiales;Incertae Sedis XIII*	*Chloroflexi*	*Ruminococcaceae*
9	*Gemella*	*Lachnospiraceae*	*Veillonella*	*Eubacterium*
10	*Porphyromonas*	*Veillonella*	*Eubacterium*	*Lachnospiraceae*
11	*Selenomonas*	*Clostridiales;Incertae Sedis XI*	*Syntrophomonas*	*Veillonella*
12	*Pasteurellaceae*	*Capnocytophaga*	*Clostridiales;Incertae Sedis XIII*	*Syntrophomonas*
13	*Capnocytophaga*	*Propionibacterineae*	*Peptococcaceae*	Parvimonas
14	*Erysipelotrichaceae*	*Selenomonas*	*Propionibacterineae*	*Streptococcus*
15	*Treponema*	*Peptococcaceae*	*Lachnospiraceae*	*Clostridia*
16	*Dialister*	*Neisseria*	*Peptostreptococcus*	*Prevotellaceae*
17	*Porphyromonadaceae*	*Ruminococcaceae*	*Leptotrichia*	Proteobacteria
18	*Bacteroides*	*Eubacterium*	Bacilli	*Paenibacillaceae*
19	*Cardiobacteriaceae*	*Campylobacter*	*Johnsonella*	*Bacillales_incertae_sedis*
20	*Staphylococcus*	*Cardiobacteriaceae*	*Staphylococcus*	*Propionibacterineae*
21	*Tannerella*	*Porphyromonadaceae*	*Selenomonas*	Mogibacterium
22	*Kingella*	*Prevotellaceae*	*Paenibacillaceae*	*Peptostreptococcus*
23	*Lactobacillus*	*Peptostreptococcus*	*Sneathia*	*Peptococcaceae*
24	*Clostridiaceae*	*Butyrivibrio*	*Clostridia*	*Selenomonas*
25	*Lactococcus*	*Pasteurellaceae*	*Bacillales_incertae_sedis*	*Acidobacteria_Gp6*

Table S4.1, continued. Taxon numbers from the rank abundance curve (Figure S4.2a) of the complete ancient and modern data set.

Taxon Rank	Modern	Medieval	Bronze Age	Neolithic
26	*Butyrivibrio*	Firmicutes	*Acidobacteria_Gp7*	*Acidobacteria_Gp4*
27	Parvimonas	*Bacteroides*	*Acidobacteria_Gp6*	*Bifidobacteriaceae*
28	*Eubacterium*	*Haemophilus*	*Acidobacteria_Gp17*	*Campylobacter*
29	Bacilli	Proteobacteria	*Bacillaceae*	*Fusobacterium*
30	*Peptococcaceae*	*Pseudoramibacter*	*Bifidobacteriaceae*	*Bacteroides*
31	*Leptotrichia*	*Johnsonella*	*Fusobacterium*	*Porphyromonadaceae*
32	*Filifactor*	*Kingella*	*Butyrivibrio*	*Dialister*
33	*Peptostreptococcus*	*Clostridia*	*Lactobacillus*	*Treponema*
34	*Mycoplasma*	Parvimonas	*Bacteroides*	*Capnocytophaga*
35	*Fusobacterium*	*Lactobacillus*	*Gemella*	*Porphyromonas*
36	*Campylobacter*	*Tannerella*	*Neisseria*	*Gemella*
37	Firmicutes	Mogibacterium	*Prevotellaceae*	*Granulicatella*
38	*Brochothrix*	*Fusobacterium*	*SR1_genera_incertae_sedis*	*SR1_genera_incertae_sedis*
39	*Azospira*	*Porphyromonas*	*Euryarchaeota*	*Euryarchaeota*
40	*Propionivibrio*	*Aggregatibacter*	*Pseudoramibacter*	*Pseudoramibacter*
41	*Bifidobacteriaceae*	*Dialister*	Proteobacteria	*Vibrio*
42	*Bacillaceae*	*Granulicatella*	*Vibrio*	*Oceanospirillales*
43	*Mogibacterium*	*Paenibacillaceae*	*Oceanospirillales*	*Aggregatibacter*
44	*Lachnospiraceae*	*Bifidobacteriaceae*	*Aggregatibacter*	*Sorangiineae*
45	*Enterococcaceae*	*Oribacterium*	*Sorangiineae*	*Nitrosospira*
46	*Acidobacteria_Gp16*	*Enterococcaceae*	*Nitrosospira*	*Sneathia*
47	*Acidobacteria_Gp17*	*Treponema*	*Deferribacteraceae*	*Deferribacteraceae*
48	*Acidobacteria_Gp2*	*Catonella*	*Oribacterium*	*Oribacterium*
49	*Acidobacteria_Gp23;*	*Filifactor*	*Catonella*	*Johnsonella*
50	*Acidobacteria_Gp3*	*Gemella*	*Clostridiales;Incertae Sedis XIV*	*Catonella*

Table S4.1, continued. Taxon numbers from the rank abundance curve (Figure S4.2a) of the complete ancient and modern data set.

Taxon Rank	Modern	Medieval	Bronze Age	Neolithic
51	*Acidobacteria_Gp4*	*Oceanospirillales*	*Bacteroidales_incertae_sedis*	*Clostridiales;Incertae Sedis XIV*
52	*Acidobacteria_Gp6*	*Sorangiineae*	*Acidobacteria_Gp4*	*Bacteroidetes;Bacteroidales_incertae_sedis*
53	*Acidobacteria_Gp7*	*Syntrophomonas*	*Acidobacteria_Gp3*	*Acidobacteria_Gp7*
54	*Propionibacterineae*	*Clostridiales;Incertae Sedis XIV*	*Acidobacteria_Gp23*	*Acidobacteria_Gp3*
55	*Bacillales_incertae_sedis*	*Bacteroidetes;Bacteroidales_incertae_sedis*	*Acidobacteria_Gp2*	*Acidobacteria_Gp23;*
56	*Bacteroidetes;Bacteroidales_incertae_sedis*	*Acidobacteria_Gp17*	*Acidobacteria_Gp16*	*Acidobacteria_Gp2*
57	*Chloroflexi*	*Acidobacteria_Gp16*	*Enterococcaceae*	*Acidobacteria_Gp17*
58	*Clostridia*	*Bacillaceae*	Mogibacterium	*Acidobacteria_Gp16*
59	*Clostridiales;Incertae Sedis XI*	*Propionivibrio*	*Propionivibrio*	*Enterococcaceae*
60	*Clostridiales;Incertae Sedis XIII*	*SR1;SR1_genera_incertae_sedis*	*Azospira*	*Bacillaceae*
61	*Clostridiales;Incertae Sedis XIV*	*Nitrosospira*	*Listeriaceae;Brochothrix*	*Propionivibrio*
62	*Catonella*	*Deferribacteraceae*	Firmicutes	*Azospira*
63	*Johnsonella*	*Bacillales_incertae_sedis*	*Campylobacter*	*Brochothrix*
64	*Oribacterium*	*Acidobacteria_Gp6*	*Mycoplasma*	Firmicutes
65	*Ruminococcaceae*	*Acidobacteria_Gp4*	*Filifactor*	*Mycoplasma*
66	*Syntrophomonas*	*Megasphaera*	Parvimonas	*Filifactor*
67	*Deferribacteraceae*	*Euryarchaeota*	*Lactococcus*	*Leptotrichia*
68	*Sneathia*	*Vibrio*	*Kingella*	Bacilli
69	*Paenibacillaceae*	*Sneathia*	*Tannerella*	*Butyrivibrio*
70	*Nitrosospira*	*Acidobacteria_Gp3*	*Cardiobacteriaceae*	*Lactococcus*
71	*Sorangiineae*	*Acidobacteria_Gp23*	*Porphyromonadaceae*	*Lactobacillus*
72	*Aggregatibacter*	*Acidobacteria_Gp2*	*Dialister*	*Kingella*
73	*Oceanospirillales*	*Brochothrix*	*Treponema*	*Tannerella*
74	*Vibrio*	Bacilli	*Erysipelotrichaceae*	*Staphylococcus*
75	Proteobacteria	*Lactococcus*	*Capnocytophaga*	*Cardiobacteriaceae*
76	*Pseudoramibacter*	*Erysipelotrichaceae*	*Pasteurellaceae*	*Erysipelotrichaceae*
77	*Euryarchaeota*	*Acidobacteria_Gp7*	*Porphyromonas*	*Pasteurellaceae*
78	*SR1;SR1_genera_incertae_sedis*	*Azospira*	*Megasphaera*	*Megasphaera*

Table S4.1, continued. Taxon numbers from the rank abundance curve (Figure S4.2a) of the complete ancient and modern data set.

Taxon Rank	Modern	Medieval	Bronze Age	Neolithic
79	*Synergistaceae*	*Mycoplasma*	*Granulicatella*	*Neisseria*
80	*TM7_genera_incertae_sedis*	*Staphylococcus*	*Haemophilus*	*Haemophilus*

Table S4.2. Taxon numbers from the rank abundance curve (Figure S4.2b), comparing the ancient data to modern individuals with periodontal disease.

Taxon Rank	Modern-Plaque from individuals with periodontal disease	Medieval	Bronze Age	Neolithic
1	*Selenomonas*	*Clostridiaceae*	*TM7_genera_incertae_sedis*	*Synergistaceae*
2	*Streptococcus*	Actinobacteria	Actinobacteria	*Clostridiaceae*
3	*Veillonella*	*Synergistaceae*	*Clostridiaceae*	Actinobacteria
4	*Dialister*	*TM7_genera_incertae_sedis*	*Synergistaceae*	*Clostridiales;Incertae Sedis XI*
5	*Gemella*	*Streptococcus*	*Ruminococcaceae*	*TM7_genera_incertae_sedis*
6	*Clostridiaceae*	*Chloroflexi*	*Clostridiales;Incertae Sedis XI*	*Clostridiales;Incertae Sedis XIII*
7	*Megasphaera*	*Leptotrichia*	*Streptococcus*	*Chloroflexi*
8	Parvimonas	*Clostridiales;Incertae Sedis XIII*	*Chloroflexi*	*Ruminococcaceae*
9	*Filifactor*	*Lachnospiraceae*	*Veillonella*	*Eubacterium*
10	*Eubacterium*	*Veillonella*	*Eubacterium*	*Lachnospiraceae*
11	*Capnocytophaga*	*Clostridiales;Incertae Sedis XI*	*Clostridiales;Incertae Sedis XIII*	*Veillonella*
12	*Peptococcaceae*	*Capnocytophaga*	*Peptococcaceae*	Parvimonas
13	*Granulicatella*	*Propionibacterineae*	*Syntrophomonas*	*Syntrophomonas*
14	Actinobacteria	*Selenomonas*	*Propionibacterineae*	*Streptococcus*
15	*Peptostreptococcus*	*Peptococcaceae*	*Lachnospiraceae*	*Prevotellaceae*
16	Mogibacterium	*Neisseria*	Bacilli	*Clostridia*
17	*Lachnospiraceae*	*Ruminococcaceae*	*Peptostreptococcus*	*Propionibacterineae*
18	*Erysipelotrichaceae*	*Eubacterium*	*Leptotrichia*	*Bacillales_incertae_sedis*
19	*Treponema*	*Campylobacter*	*Johnsonella*	Mogibacterium
20	*Acidobacteria_Gp16*	*Cardiobacteriaceae*	*Selenomonas*	*Peptococcaceae*
21	*Acidobacteria_Gp17*	*Porphyromonadaceae*	*Staphylococcus*	*Peptostreptococcus*
22	*Acidobacteria_Gp2*	*Prevotellaceae*	*Acidobacteria_Gp17*	*Selenomonas*
23	*Acidobacteria_Gp23*	*Butyrivibrio*	*Acidobacteria_Gp6*	*Paenibacillaceae*
24	*Acidobacteria_Gp3*	*Peptostreptococcus*	*Acidobacteria_Gp7*	Proteobacteria
25	*Acidobacteria_Gp4*	*Pasteurellaceae*	*Bifidobacteriaceae*	*Acidobacteria_Gp4*

Table S4.2. Taxon numbers from the rank abundance curve (Figure S4.2b), comparing the ancient data to modern individuals with periodontal disease.

Taxon Rank	Modern-Plaque from individuals with periodontal disease	Medieval	Bronze Age	Neolithic
26	*Acidobacteria_Gp6*	*Bacteroides*	*Bacillaceae*	*Acidobacteria_Gp6*
27	*Acidobacteria_Gp7*	Firmicutes	*Bacillales_incertae_sedis*	*Bifidobacteriaceae*
28	*Bifidobacteriaceae*	*Haemophilus*	*Bacteroides*	*Porphyromonadaceae*
29	*Propionibacterineae*	Proteobacteria	*Prevotellaceae*	*Bacteroides*
30	*Bacillaceae*	*Johnsonella*	*Clostridia*	*Capnocytophaga*
31	*Bacillales_incertae_sedis*	*Kingella*	*Butyrivibrio*	*Porphyromonas*
32	Bacilli	*Pseudoramibacter*	*Fusobacterium*	*Dialister*
33	*Porphyromonadaceae*	*Tannerella*	*Sneathia*	*Fusobacterium*
34	*Bacteroidales_incertae_sedis*	*Clostridia*	*Lactobacillus*	*Granulicatella*
35	*Bacteroides*	Parvimonas	*Neisseria*	*Campylobacter*
36	*Tannerella*	*Lactobacillus*	*Paenibacillaceae*	*Treponema*
37	*Porphyromonas*	Mogibacterium	*Gemella*	*Gemella*
38	*Prevotellaceae*	*Porphyromonas*	*Acidobacteria_Gp16*	*Acidobacteria_Gp16*
39	Chloroflexi	*Fusobacterium*	*Acidobacteria_Gp2*	*Acidobacteria_Gp17*
40	*Clostridia*	*Dialister*	*Acidobacteria_Gp23*	*Acidobacteria_Gp2*
41	*Clostridiales;Incertae Sedis XI*	*Granulicatella*	*Acidobacteria_Gp3*	*Acidobacteria_Gp23*
42	*Clostridiales;Incertae Sedis XIII*	*Aggregatibacter*	*Acidobacteria_Gp4*	*Acidobacteria_Gp3*
43	*Clostridiales;Incertae Sedis XIV*	*Bifidobacteriaceae*	*Porphyromonadaceae*	*Acidobacteria_Gp7*
44	*Butyrivibrio*	*Paenibacillaceae*	*Bacteroidales_incertae_sedis*	*Bacillaceae*
45	*Catonella*	*Oribacterium*	*Capnocytophaga*	Bacilli
46	*Johnsonella*	*Enterococcaceae*	*Tannerella*	*Bacteroidales_incertae_sedis*
47	*Oribacterium*	*Treponema*	*Porphyromonas*	*Tannerella*
48	*Ruminococcaceae*	*Catonella*	Parvimonas	*Clostridiales;Incertae Sedis XIV*
49	*Syntrophomonas*	*Filifactor*	Mogibacterium	*Butyrivibrio*
50	*Deferribacteraceae*	*Gemella*	*Clostridiales;Incertae Sedis XIV*	*Catonella*

Table S4.2. Taxon numbers from the rank abundance curve (Figure S4.2b), comparing the ancient data to modern individuals with periodontal disease.

Taxon Rank	Modern-Plaque from individuals with periodontal disease	Medieval	Bronze Age	Neolithic
51	Firmicutes	*Acidobacteria_Gp16*	*Catonella*	*Johnsonella*
52	*Fusobacterium*	*Acidobacteria_Gp17*	*Oribacterium*	*Oribacterium*
53	*Leptotrichia*	*Bacillaceae*	*Filifactor*	*Filifactor*
54	*Sneathia*	*Bacteroidales_incertae_sedis*	*Dialister*	*Megasphaera*
55	*Enterococcaceae*	*Clostridiales;Incertae Sedis XIV*	*Megasphaera*	*Deferribacteraceae*
56	*Lactobacillus*	*Syntrophomonas*	*Deferribacteraceae*	*Erysipelotrichaceae*
57	*Brochothrix*	*Propionivibrio*	*Erysipelotrichaceae*	Firmicutes
58	*Kingella*	*Sorangiineae*	Firmicutes	*Leptotrichia*
59	*Neisseria*	*Oceanospirillales*	*Granulicatella*	*Sneathia*
60	*Paenibacillaceae*	*Acidobacteria_Gp4*	*Enterococcaceae*	*Enterococcaceae*
61	*Nitrosospira*	*Acidobacteria_Gp6*	*Brochothrix*	*Lactobacillus*
62	*Azospira*	*Bacillales_incertae_sedis*	*Kingella*	*Brochothrix*
63	*Propionivibrio*	*Megasphaera*	*Nitrosospira*	*Kingella*
64	*Sorangiineae*	*Deferribacteraceae*	*Azospira*	*Neisseria*
65	*Campylobacter*	*Nitrosospira*	*Propionivibrio*	*Nitrosospira*
66	*Aggregatibacter*	*SR1;SR1_genera_incertae_sedis*	*Sorangiineae*	*Azospira*
67	*Cardiobacteriaceae*	*Acidobacteria_Gp2*	*Campylobacter*	*Propionivibrio*
68	*Haemophilus*	*Acidobacteria_Gp23*	*Aggregatibacter*	*Sorangiineae*
69	*Oceanospirillales*	*Acidobacteria_Gp3*	*Cardiobacteriaceae*	*Aggregatibacter*
70	*Pasteurellaceae*	Bacilli	*Haemophilus*	*Cardiobacteriaceae*
71	*Vibrio*	*Erysipelotrichaceae*	*Oceanospirillales*	*Haemophilus*
72	Proteobacteria	*Sneathia*	*Pasteurellaceae*	*Oceanospirillales*
73	*Pseudoramibacter*	*Brochothrix*	*Vibrio*	*Pasteurellaceae*
74	*Euryarchaeota*	*Vibrio*	Proteobacteria	*Vibrio*
75	*SR1_genera_incertae_sedis*	*Euryarchaeota*	*Pseudoramibacter*	*Pseudoramibacter*
76	*Staphylococcus*	*Lactococcus*	*Euryarchaeota*	*Euryarchaeota*
77	*Lactococcus*	*Acidobacteria_Gp7*	*Treponema*	*SR1_genera_incertae_sedis*
78	*Synergistaceae*	*Azospira*	*SR1;SR1_genera_incertae_sedis*	*Staphylococcus*

Table S4.2. Taxon numbers from the rank abundance curve (Figure S4.2b), comparing the ancient data to modern individuals with periodontal disease.

Taxon Rank	Modern-Plaque from individuals with periodontal disease	Medieval	Bronze Age	Neolithic
79	*Mycoplasma*	*Staphylococcus*	*Lactococcus*	*Lactococcus*
80	*TM7_genera_incertae_sedis*	*Mycoplasma*	*Mycoplasma*	*Mycoplasma*

Table S4.1. The change in -diversity (Shannon Index) over time for the classified bacterial sequences, a restricted subset of only Gram-positive taxa and a restricted comparison of the ancient data to only modern individuals with periodontal disease. Variation in -diversity was calculated by comparing Shannon Index values at four time periods (Neolithic, Bronze Age, Medieval and Modern) by using a NP-MANOVA. The NP-MANOVA was calculated using a Bray-Curtis distance matrix with 100,000 permutations generated from the Shannon Index values. The grey-shaded values indicate the Bonferroni-corrected values.

Shannon Index- complete dataset	Neolithic	Bronze Age	Medieval	Modern
Neolithic	0	0.5656	0.0252	0.0001
Bronze Age	1	0	0.1664	0.0005
Medieval	0.1512	0.9984	0	0
Modern	0.0006	0.00024	0.003	0

Shannon Index- Gram-positive taxa only	Neolithic	Bronze Age	Medieval	Modern
Neolithic	0	0.8905	0.02656	0.00272
Bronze Age	1	0	0.4088	0.006
Medieval	0.1594	0.06813	0	0
Modern	0.01632	0.036	0	0

Shannon Index- modern periodontal disease only	Neolithic	Bronze Age	Medieval	Modern
Neolithic	0	0.5598	0.0245	0.0002
Bronze Age	1	0	0.0002	0.0008
Medieval	0.147	0.9552	0	0
Modern	0.0012	0.0048	0	0

SUPPLEMENTARY MATERIAL SECTION 5. Discriminant analysis of phylogenetic variation between the archaeological groups investigated.

Linear functional discriminant analysis (PASW 17 Statistics) was used to examine whether the microbial makeup of dental calculus was more similar between individuals from the same archaeological grouping as opposed to different groups. Discriminant analysis identifies variation between groups, whereas PCoA displays all variation within the data[31]. Discriminant analysis requires data to be mutivariate-Normal and uncorrelated[31]. Genetic data usually have neither of these traits. Hence, we performed discriminant analysis on the principal components generated from the PCoA of the unweighted UniFrac distances, which are Normally distributed and uncorrelated. Discriminant analysis was applied to each region of 16S rRNA gene, both with (Figure S5.1) and without Gram-negative lineages (Figure S5.2).

All discriminant analyses clustered individuals by their archaeological grouping. When comparing groups from the Medieval, the urban populations (Jewbury and St Helen on the Walls) consistently clustered together, distinct from the rural group (Raunds). Medieval urban populations in Britain more frequently display skeletal and dental signs of stress compared to rural groups[32]. Therefore, this finding highlights the potential use of dental calculus to investigate prehistoric human lifestyle.

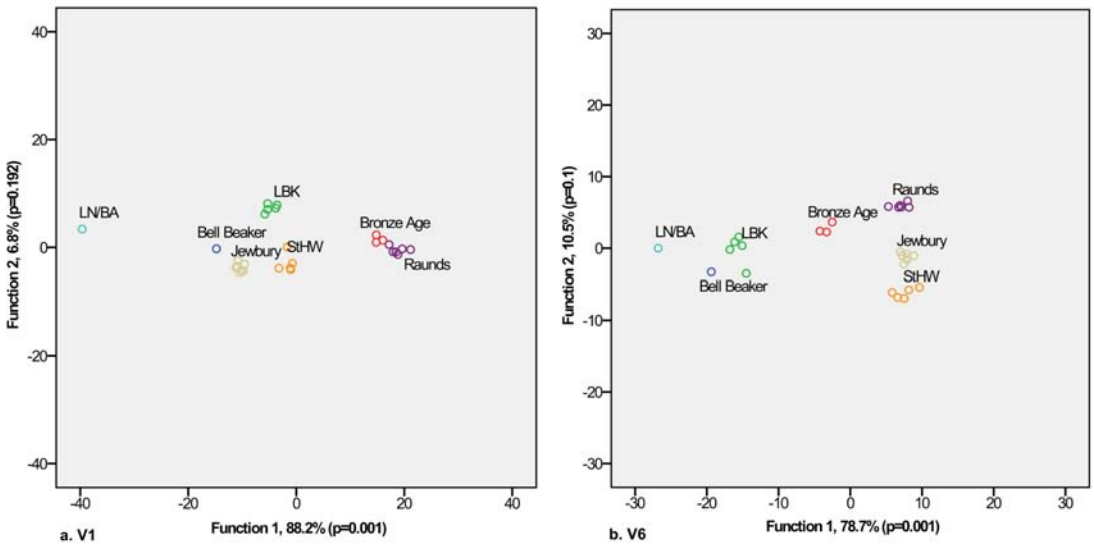

Figure S5.1. Discriminant analysis of the microbial makeup of ancient dental calculus deposits shows clustering of individuals by their archaeological grouping. We performed discriminant analyses on the principal components generated from the unweighted UniFrac distances produced from phylogenetic trees of the V1 (a) and V6 (b) regions of the 16S rRNA gene. Each individual is represented by a circle and coloured according to archaeological grouping (LBK, green; Bell Beaker, blue; Unetice, aqua; Bronze Age, red; Raunds, purple; Jewbury, brown and St Helen on the Walls (StHW), orange).

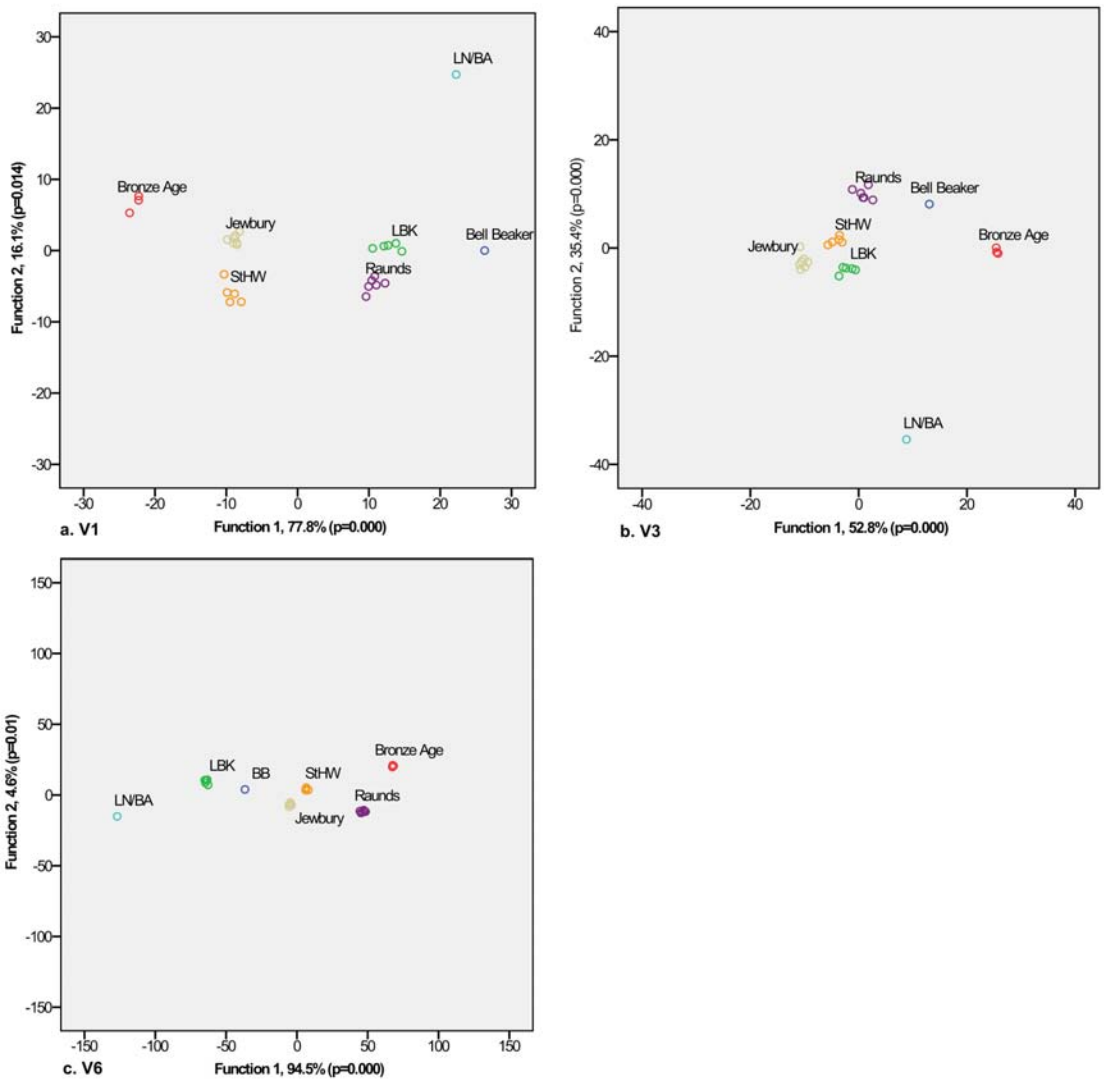

Figure S5.2. Discriminant analysis of the Gram-positive microbes in ancient dental calculus samples shows clustering of individuals by their archaeological grouping. We performed discriminant analyses on the principal components generated from the unweighted UniFrac distances produced from phylogenetic trees of the V1 (a), V3 (b) and V6 (c) regions of the 16S rRNA gene. Each individual is represented by a circle and coloured according to archaeological grouping (LBK, green; Bell Beaker, blue; Unetice, aqua; Bronze Age, red; Raunds, purple; Jewbury, brown and St Helen on the Walls (StHW), yellow).

REFERENCES

1 Bogucki, P. *Forest farmers and stockherders: early agriculture and its consequences in north-central Europe*. (University Press, 1988).

2 Ximenez-Fyvie, L. A., Haffajee, A. D. & Socransky, S. S. Comparison of the microbiota of supra- and subgingival plaque in health and periodontitis. *J Clin Periodontol* **27**, 648-657, (2000).

3 Quince, C. *et al.* Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* **6**, 639-641, (2009).

4 Lilley, J., Stroud, G. & Brothwell, D. in *The Arcaheology of York* Vol. 12 (eds P. V. Addyman & V. A. Kinsler) (Council for British Archaeology, 1994).

5 Boddington, A. *Raunds Furnells. The Anglo-Saxon church and churchyard*. (English Heritage, 1996).

6 Dawes, J. D. & Magilton, J. R. in *The Archaeology of York, The Medieval Cemeteries* Vol. 12 (Council for British Archaeology for York Archaeological Trust, 1980).

7 Yoshida, A. *et al.* Development of a 5' nuclease-based real-time PCR assay for quantitative detection of cariogenic dental pathogens Streptococcus mutans and Streptococcus sobrinus. *J Clin Microbiol* **41**, 4438-4441, (2003).

8 Yoshida, A. *et al.* Development of a 5' fluorogenic nuclease-based real-time PCR assay for quantitative detection of Actinobacillus actinomycetemcomitans and Porphyromonas gingivalis. *J Clin Microbiol* **41**, 863-866, (2003).

9 White, J. R., Nagarajan, N. & Pop, M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* **5**, e1000352, (2009).

10 Ochman, H., Elwyn, S. & Moran, N. A. Calibrating bacterial evolution. *Proc Natl Acad Sci U S A* **96**, 12638-12643, (1999).

11 Whittaker, R. H. Evolution and measurement of species diversity. *Taxon* **21**, 213-251, (1972).

12 Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I. & Dewhirst, F. E. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* **43**, 5721-5732, (2005).

13 Aas, J. A. *et al.* Bacteria of dental caries in primary and permanent teeth in children and young adults. *J Clin Microbiol* **46**, 1407-1417, (2008).

14 Faveri, M. *et al.* Microbiological diversity of generalized aggressive periodontitis by 16S rRNA clonal analysis. *Oral Microbiol Immunol* **23**, 112-118, (2008).

15 Schloss, P. D. & Handelsman, J. Toward a census of bacteria in soil. *PLoS Comput Biol* **2**, e92, (2006).

16 Ellis, R. J., Morgan, P., Weightman, A. J. & Fry, J. C. Cultivation-dependent and -independent approaches for determining bacterial diversity in heavy-metal-contaminated soil. *Appl Environ Microbiol* **69**, 3223-3230, (2003).

17 Nogales, B. *et al.* Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. *Appl Environ Microbiol* **67**, 1874-1884, (2001).

18 Elshahed, M. S. *et al.* Novelty and uniqueness patterns of rare members of the soil biosphere. *Appl Environ Microbiol* **74**, 5422-5428, (2008).

19 Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308**, 554-557, (2005).

20 Beier, S., Witzel, K. P. & Marxsen, J. Bacterial community composition in Central European running waters examined by temperature gradient gel electrophoresis and sequence analysis of 16S rRNA genes. *Appl Environ Microbiol* **74**, 188-199, (2008).

21 Lazarevic, V., Whiteson, K., Hernandez, D., Francois, P. & Schrenzel, J. Study of inter- and intra-individual variations in the salivary microbiota. *BMC Genomics* **11**, 523, (2010).

22 Will, C. *et al.* Horizon-Specific Bacterial Community Composition of German Grassland Soils, as Revealed by Pyrosequencing-Based Analysis of 16S rRNA Genes. *Applied and Environmental Microbiology* **76**, 6751-6759, (2010).

23 Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335-336, (2010).

24 Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science* **320**, 1647-1651, (2008).

25 Stamatakis, A., Hoover, P. & Rougemont, J. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* **57**, 758-771, (2008).

26 Lozupone, C., Hamady, M. & Knight, R. UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**, 371, (2006).

27 Sidaway, D. A. A microbiological study of dental calculus. I. The microbial flora of mature calculus. *J Periodontal Res* **13**, 349-359, (1978).

28 Faith, D. P. Conservation evaluation and phylogenetic diversity. *VBiological Conservation* **61**, 1-10, (1992).

29 Dahlen, G. *et al.* A microbiological study in relation to the presence of caries and calculus. *Acta Odontol Scand* **68**, 199-206, (2010).

30 Jin, Y. & Yip, H. K. Supragingival calculus: formation and control. *Crit Rev Oral Biol Med* **13**, 426-441, (2002).

31 Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* **11**, 94, (2010).

32 Lewis, M. E. Impact of industrialization: comparative study of child health in four sites from medieval and postmedieval England (A.D. 850-1859). *Am J Phys Anthropol* **119**, 211-223, (2002).

CHAPTER FIVE

Conclusion

This thesis contains a portfolio of chapters in the form of an existing publication, a paper in the process of submission and a traditionally formatted chapter. Each chapter contains detailed discussions and conclusions of the research presented therein. The work of the preceding chapters is drawn together here, to emphasise how each piece has contributed to the overall aims of this thesis. In this closing chapter, the significance, limitations and future directions of the research are considered.

5.1 Overall summary of the aims of the thesis

In this thesis, ancient DNA was used to investigate the impact of the development of agriculture in Europe during the Neolithic period, on both the population structure and health of prehistoric and modern Europeans. The Neolithic Revolution was investigated because it represents one of the most significant events in human history, heralding the transition from a hunter-gatherer to agriculturist lifestyle (Flannery, 1973, Jones and Liu, 2009). The transition to an agriculture lifestyle was significant because it largely ended the practice of collecting wild foods from the environment, which is how humans had obtained food for the majority of their history (Milton, 1999). The development of agriculture aside from altering human diet also changed many aspects of human lifestyle and in doing so established the social (Bar-Yosef and Belfer-Cohen, 1992), economic (Weisdorf, 2005) and biological (Taylor et al., 2001, Wolfe et al., 2007) characteristics of modern societies.

This thesis has used three disparate areas of ancient DNA research to investigate how the Neolithic Revolution has shaped European population structure and health. As all the research presented in this thesis relies on the application of ancient DNA techniques, the methods used to recover DNA from preserved hard remains were assessed. Quantification of DNA from human skeletal remains, in a variety of tissues and from samples that had been either drilled or pulverised, showed that common sampling techniques were detrimental to the recovery of DNA from preserved hard remains. To describe the population dynamics following the-Early Neolithic, ancient human DNA was recovered from Early and Late Neolithic/Early Bronze Age skeletal remains from Central Europe. Population genetic analysis of these samples revealed that the Early Neolithic farming cultures were not fully incorporated into succeeding populations. To assess the role of the development of agriculture on not only the population structure, but also on the health of prehistoric Europeans, ancient oral microbial DNA was

recovered from archaeological dental calculus deposits. The microbial content of dental calculus samples from prehistoric European agricultural communities in the Neolithic, Bronze Age and Medieval periods suggested that periodontal disease was an early consequence of an agricultural diet. For each of the aims presented in the introductory chapter, detailed summaries of the research, the significance of the work, and the future directions and limitations are outlined below.

5.2 Summary of Aim 1

To quantify the amount of DNA in skeletal remains and assess the impact of commonly used skeletal sampling techniques on the recovery of DNA.

Chapter 2 demonstrated that some of the current methods used in ancient DNA and forensic science to recover genetic information from skeletons are highly detrimental to DNA survival. The main finding is that drilling of bone and teeth under standard conditions can reduce DNA amounts by over five orders of magnitude. In addition, it was shown that current approaches do not sample the most DNA rich tissues (the cementum layers of tooth roots) and have failed to accurately establish where DNA survives or how it degrades over time. To analyse this issue further, the distribution of DNA quantity at a range of fragment lengths in human remains recovered from a variety of post-mortem environments (e.g. Europe, Asia and South America) were used to reveal an exponential pattern of DNA degradation, which can be used to generate site-specific decay rates. This pattern provides a new empirical platform for the analysis of DNA survival in the fields of archaeology, forensic science and ancient DNA.

5.2.1 Significance

Beyond investigating the impact of the Neolithic Revolution on prehistoric and modern European populations, the research presented in this thesis has general implications for the field of ancient DNA research. This has included revealing that commonly used skeletal sampling techniques for DNA analysis can be detrimental to the recovery of DNA from preserved remains.

Chapter 2 attempted to stimulate a new standardised approach to the genetic analysis of ancient human specimens, in particular by investigating the methods used to recover DNA from preserved remains. Ancient DNA and forensic science studies may have obtained false negative DNA results, especially from important specimens that are commonly sampled with drills to minimize signs of damage. The findings presented in this thesis provide a rationale for the re-sampling of specimens from which DNA has not been retrieved and were originally

sampled using non-ideal techniques, such as high speed drilling or using the dentine tooth tissue as opposed to the cementum tooth tissue for DNA analysis. For example, the original DNA analysis of the *Homo floresiensis* remains were unsuccessful, however the researchers targeted the dentine tooth tissue

The results presented in Chapter 2 also provide the means to make the process of recovering DNA from skeletal remains less subjective than is presently the case. There is now quantified information about the amount of mtDNA in different tooth tissues, the impact of powdering skeletal remains with methods such as drilling and pulverising on the recovery of DNA from these remains and the distribution of the quantity of ancient DNA at different fragment lengths. This information will assist ancient DNA researchers when they are approaching museums about sampling skeletal remains for genetic analysis, as they can now provide quantified reasons for why they would like to sample certain skeletal elements and with specific powdering techniques. The information will also assist archaeologists and physical anthropologists in selecting samples for genetic analysis. The results presented in Chapter 2 provide information about the behaviour of DNA recovered from preserved remains. This knowledge is fundamental to ancient DNA research, as the field is almost entirely built on the process of extracting DNA from skeletal remains.

5.2.2 Limitations of the research and future directions

In attempting to standardise approaches to sampling skeletal remains for genetic analysis it would be useful to investigate several research avenues that were not addressed in this thesis. These include investigating the impact of a wider range of sampling methods on the recovery of DNA from skeletal remains, the influence of the time since death of a sample on the amount of DNA in the cellular tooth tissues and the distribution of DNA across the whole skeleton.

The most common skeletal crushing techniques, drilling and pulverising, were investigated in this thesis, however only a few variables were assessed, such as drill speed. It would be useful to explore the effect of a larger range of drilling speeds (only two have been assessed at present) on the recovery of DNA, as although low speed drilling of skeletal remains has been found to be favourable for the recovery of DNA, there may be a point at which decreasing drill speed becomes detrimental to DNA survival. When hard remains are drilled at low speed, it takes longer to powder the remains and hence increases the time of exposure of the skeletal elements to heat generated during drilling. To investigate this, skeletal remains could be drilled at a range of speeds, from 100 to 1000 rpm at 200 rpm intervals, and the DNA quantified from these tissues. Aside from the speed, the type of drill bit used to powder

skeletal remains may also alter the force imposed while powdering hard tissues, and hence the heat generated during this process. The drilling of skeletal elements with a large amount of force has been found to be associated with increased temperature (Hillery and Shuaib, 1999). Hence, the size and the material (i.e. tungsten versus wood) of the drill bit may alter the force imposed on the skeletal element while drilling and hence the amount of heat and therefore, DNA damage, inflicted during the powdering of the tissue. To determine the impact of the drill bit on DNA recovery, a range of drills of different sizes and materials could be used to powder skeletal remains. Another method that has been used to prevent heat build-up while pulverising skeletal remains in a bone mill is the pre-cooling of canisters with liquid nitrogen (Hoss and Paabo, 1993). The heat generated when pulverising is probably going to quickly warm the chilled canisters, however the effectiveness of this technique has not been quantified. Therefore, it would be useful to determine if cooling pulverising equipment aids the recovery of DNA from skeletal remains.

The amount of mtDNA in the dentine and cementum tooth tissues was quantified using samples that had similar times since death. As such, these results did not demonstrate whether the DNA within dentine and cementum degraded at different rates. The rate of DNA degradation may vary between dentine and cementum as the tissues have different cellular properties (Berkovitz et al., 1992, De Leo et al., 2000, Carda and Peydro, 2006). For example, the mtDNA in dentine is trapped within mineralised odontoblastic tubules (Berkovitz et al., 1992), whereas in cementum, the cementoblasts and cementocytes are located in mineralised small channels called lucane (Bosshardt and Schroeder, 1991, De Leo et al., 2000). If the DNA in dentine and cementum degrade at different rates, then choosing which tooth tissue to sample for DNA analysis would depend on how long the tooth has been in the post-mortem environment. The rate of DNA degradation in dentine and cementum could be determined by quantifying the amount of DNA in both tooth tissues from a temporal range of samples, spanning forensic and historic (10 – 100 years) remains to archaeological (over 100 years) specimens. While the results presented in Chapter 2 have shown the relative quantity of mtDNA in dentine and cementum for very old remains (7,500 years), the same pattern may not be observed in historic samples. Hence, this would have important ramifications for the application of the findings to forensic science, where they are commonly dealing with samples which have spent under 100 years in the post-mortem environment.

Although determination of the distribution of DNA in the cellular tooth tissues is useful for the fields of ancient DNA research and forensic science, as teeth are frequently used for genetic analysis (Gilbert et al., 2003, Rudbeck et al., 2005), DNA is also routinely recovered from the rest of the skeleton. There have been attempts to compare the amount of DNA in

different skeletal elements (Pruvost et al., 2008). Pruvost and colleagues (2008) study investigated which skeletal element contained the most DNA by comparing the amount of DNA in a range of different skeletal elements from individuals sampled at different times since death and from a variety of post-mortem environments. Therefore, the recommendations given by this study concerning which skeletal element contained the most DNA was confounded by variables such as the varying post-mortem environments of the individuals that were compared. To determine the distribution of DNA across the skeleton, DNA would need to be sampled from complete (or near to complete) skeletons. To minimise the influence of individual variation it would be preferable to determine the distribution of DNA across a number of skeletons which all had a similar time since death and had been recovered from similar post-mortem environments. However, as human remains are valuable to multiple fields (Wood et al., 1992, Donlon, 1994, Bentley, 2006), it would be extremely difficult to gain access to complete skeletons for the purpose of performing repeated destructive analysis.

This thesis also found that DNA degradation follows an exponential process of fragmentation which was consistent across a variety of samples from different post-mortem environments. However, the pattern of DNA degradation was described using models based on only four data points (the four fragments of DNA quantified). It would be more realistic to fit the models of DNA degradation examined to data obtained from next generation sequencing, shotgun sequencing of degraded samples, as this type of dataset would contain a distribution of the fragment lengths present within the sample. Analysis of DNA degradation from a highly resolved distribution of DNA fragment lengths from ancient samples, compared to the dataset generated in Chapter 2, would provide a more accurate description of DNA fragmentation and, enable a more precise calculation of the rate of DNA degradation.

To determine if the protocols recommended in this thesis improve the ability to recover DNA from skeletal remains it would be useful to test the recommended methods on samples which have yet to yield DNA. In particular, the testing should be performed on skeletal remains which were originally sampled using techniques which are now known to degrade genetic material, such as high speed drilling or targeted areas of low DNA concentration, such as the dentine tooth tissue.

5.3 Summary of Aim 2

To determine if a single, continuous population has existed in Central Europe since the introduction of agriculture in the Early Neolithic until today.

Population continuity from the introduction of agriculture in the Early Neolithic until today was found to be an unlikely model of demographic history, as indicated by the results from population genetic and coalescent analyses of ancient mtDNA retrieved from Early Neolithic (n=59) and Late Neolithic/Early Bronze Age (n=49) cultures that were sampled from Central Europe, and compared to extant populations from the same region. Internal population changes in Europe between the Early Neolithic and Late Neolithic/Early Bronze Age appear to have contributed substantially to the population structure of extant Central Europeans, as all the Late Neolithic/Early Bronze Age cultures examined (Bell Beakers, Corded Ware or Unetice) were found to be more likely ancestors of modern Central Europeans than either of the Early Neolithic cultures (LBK and Rössen) investigated. Haplogroup distributions suggest that Palaeolithic mtDNA haplogroups which were infrequent in the Early Neolithic, such as haplogroups H and U, became more frequent during the Late Neolithic. These findings indicate that population change between the Early and Late Neolithic/Early Bronze Age has had a major influence on the population structure of modern Central Europeans.

5.3.1 Significance of the research

The general approach of using ancient human DNA from temporally spanned samples to investigate the Neolithic Revolution provides new and significant insights into both the lives of past humans. In particular, the results presented in Chapter 3 demonstrated that there was population change between the Early Neolithic and the Late Neolithic/Early Bronze Age, and identified that specific oral pathogens were associated with the introduction of agriculture.

Chapter 3 revealed that there was a degree of genetic discontinuity between the Early Neolithic and Late Neolithic/Early Bronze Age cultures in Central Europe, based on haplogroup distributions and genetic distances. Before this study, it was unknown if population changes occurred after the introduction of farming during the Early Neolithic in Central Europe or how these putative changes may have influenced the settlement history of the continent. There was no evidence that substantial population change had occurred after the Early Neolithic in Central Europe from the genetic analysis of modern European populations (Ammerman and Cavalli-Sforza, 1984, Richards et al., 1996, Richards et al., 2000, Torroni et al., 2001, Richards et al., 2002). Therefore, the settlement events suggested to have formed the genetic makeup of modern day Europe were restricted to pre-Late Neolithic events, including the introduction of agriculture during the Early Neolithic (Menozzi et al., 1978,

Cavalli-Sforza et al., 1993, Cavalli-Sforza et al., 1994), and Palaeolithic events, including the original settlement of Europe (Rootsi et al., 2004, Achilli et al., 2005, Soares et al., 2010) and the post-LGM re-colonisation of the continent (Torroni et al., 1998, Richards et al., 2000, Achilli et al., 2004, Rootsi et al., 2004, Achilli et al., 2005, Soares et al., 2010). However, analysis of ancient mtDNA from early farmers (Haak et al., 2010) and post-LGM hunter-gatherers (Bramanti et al., 2009) from Central Europe has revealed a degree of genetic discontinuity based on genetic distance comparisons ($p<0.05$) between both past groups and modern day Central Europeans. These findings indicated that population changes following the Early Neolithic may have contributed to shaping the mtDNA landscape in Europe, however both these studies were unable to directly identify when this occurred or how the population changes would have influenced modern Central Europeans. The analysis of ancient mtDNA from Late Neolithic/Bronze Age remains recovered from Central Europe has revealed that the degree of genetic discontinuity observed between the Early Neolithic and modern populations was at least partially attributable to population changes which occurred between the Early and Late Neolithic. These findings alter the traditional view of which past settlement events in Europe have contributed to the current genetic structure of the continent and demonstrate that past events which are not associated with major climatic or economic changes in human history, such as internal migrations in Europe during the Late Neolithic/Early Bronze Age, can also substantially alter population structure.

The research presented in Chapter 3 also assists in understanding the mode by which agriculture spread to Europe. Previous analysis of ancient mtDNA from Early Neolithic (Haak et al., 2010) and post-LGM hunter-gatherer (Bramanti et al., 2009) remains from Central Europe found population discontinuity based on genetic distance and haplogroup distribution comparisons. The genetic differences observed between early agriculturists and hunter-gatherers have been used to support the demic diffusion model that farming spread to Europe by the movement of people and not by the transmission of ideas. The findings in this thesis also supported that agriculture spread to Central Europe in accordance with the demic diffusion model as the haplogroup signature of the early Neolithic cultures investigated, LBK and Rössen, were similar to past investigations of the LBK (Haak et al., 2005, Haak et al., 2010), containing a high frequency of haplogroup N1a. The discontinuity in haplogroup makeup between the Early Neolithic and Late Neolithic cultures investigated in this thesis indicates that although the introduction of agriculture in Central Europe was associated with population change, as opposed to the transmission of ideas, that this population change did not greatly influence Europe's population structure because these early farming communities were incompletely incorporated into the succeeding populations of the Late Neolithic/Early Bronze Age.

The incomplete incorporation of the Early Neolithic cultures into the succeeding populations of the Late Neolithic/Early Bronze Age may explain why the majority of Europe's genetic diversity today is found to be indigenous to the continent (Torroni et al., 1998, Richards et al., 2000, Semino et al., 2000, Achilli et al., 2004, Rootsi et al., 2004, Achilli et al., 2005, Soares et al., 2010). The Late Neolithic and Early Bronze Age cultures were found to have a higher frequency of mtDNA lineages associated with Palaeolithic settlement events (Achilli et al., 2004, Loogvali et al., 2004, Achilli et al., 2005), including haplogroup H which was particularly frequent amongst individuals of the Bell Beaker culture, and haplogroup U that was frequent among individuals from the Corded Ware and Unetice cultures. Hence, the presence of Palaeolithic mtDNA lineages during the Late Neolithic and Early Bronze Age may reflect the emergence (or re-emergence) of indigenous European populations into Central Europe, following the influx of farming populations that were of Near Eastern descent into this region during the Early Neolithic (Haak et al., 2010). The direct genetic analysis of Late Neolithic and Early Bronze Age individuals from Central Europe has provided a way to trace the past movement of mtDNA lineages and determine when population changes occurred which have contributed to the current European population structure.

5.3.2 Limitations of the research and future directions

Chapter 3 revealed that in Central Europe, population changes occurred between the Early Neolithic and Late Neolithic/Early Bronze Age. However, these results do not demonstrate what population movements were occurring during the Middle Neolithic or if population changes were also occurring outside Central Europe during the Neolithic. Furthermore, the results presented in this thesis are based on the analysis of maternally inherited mtDNA and hence provide no direct information about the population history of males during the Neolithic and Bronze Age in Central Europe.

Analysis of a wider range of temporally spaced samples than were examined in Chapter 3 would provide a detailed picture of human population dynamics during the Neolithic and Bronze Age. Multiple cultures existed in Central Europe during the Neolithic and Early Bronze Age, as detailed in Table 5.1. Coalescent analysis of an ancient DNA dataset that includes genetic information from Middle Neolithic cultures in addition to Early and Late Neolithic/Early Bronze Age groups could be used to determine the timing and nature of the genetic discontinuity observed between the Early and Late Neolithic. In addition, the recovery of DNA from Middle Neolithic cultures would help to resolve the contribution of this period to the population structure of modern Central Europeans. The creation of a detailed temporal transect of ancient DNA that spans the Early Neolithic to Bronze Age, could be used to investigate changes in effective population size over time through the use of Bayesian Skyline

plot analysis (Drummond et al., 2005). By performing Skyline plot analyses on an ancient DNA dataset that covers the Early Neolithic to Bronze Age and combining this with already published genetic data from hunter-gatherers (Bramanti et al., 2009), the changes in effective population size of Central Europeans since the post-LGM could be studied. This type of analysis would enable inferences to be made about the impact of past events, such as the re-colonisation of Europe, the introduction of agriculture and potentially internal migrations in Europe during the Late Neolithic/Bronze Age, on effective population size.

Culture	Time (BC)	Period
Linienbandkeramik Culture (LBK)	5450–4775	Early Neolithic
Stichbandkeramik Culture (SBK)	4925–4550	
Rossen Culture (RSK)	4625–4475/4250	
Gaterslebener Culture (GLK)	4475–3950	
Lengyel	4250-3950	
Michelsberger Culture (MBK)	4250–3500	Middle Neolithic
Jordansmühler Culture (JMK)	4100–3700	
Baalberger Culture (BAK)	3950–3400	
Salzmünder Culture (SMK)	3400–3100/3025	
Tiefstichkeramik Culture (TSK)	3650–3325	
Walternienburger Culture (WBK)	3325–3100	
Bernburger Culture (BBK)	3100–2650	
Elb-Havel-Culture (EHK)	3100–2650	
Schnurkeramik Culture (SKK)	2800–2050	Late Neolithic
Einzelgrab Culture (EGK)	2800–2200	
Schönfelder Culture (SFK)	2800–2050	
Corded Ware	2700-2000	
Ammenslebener Gruppe	2650–2200	
Bell Beaker (GBK)	2500–2050	
Unetice	2050–1800	Early Bronze Age

Table 5.1. Cultures present during the Neolithic in Central Europe (Germany).

The temporal transect approach, where DNA is sampled from chronologically spaced human remains recovered from a restricted geographic area, could also be applied to investigate population dynamics during the Neolithic in areas outside Central Europe. The application of this approach to multiple areas would show whether population change was occurring across Europe during the Neolithic or if this was only a characteristic of Central Europe. Furthermore, it would help to resolve how migrations within Europe during the Neolithic and Bronze Age may have contributed to the current European population structure. For example, although ancient DNA from Early Neolithic and Late Neolithic/Early Bronze Age individuals revealed a degree of population discontinuity between these periods in Central Europe, I could only infer the migratory sources of the Late Neolithic and Bronze Age cultures. For example, the Bell Beakers were postulated to have a Western European origin due to

archaeological evidence (Gimbutas, 1963) and the high frequency of haplogroup H among this culture, which resulted in the Bell Beakers having a similar haplogroup composition to modern (Achilli et al., 2004, Loogvali et al., 2004, Soares et al., 2010) and Neolithic (Sampietro et al., 2007) Western European populations. To trace the movement of Neolithic cultures in Europe, ancient DNA would need to be recovered from Early to Late Neolithic remains sampled across the continent.

The degree of population discontinuity detected between the Early Neolithic and Late Neolithic/Early Bronze Age was only based on the analysis of ancient mtDNA. Therefore, the findings reflect the population history of females during the Neolithic. The movement of males and females during the Neolithic may have varied. For example, strontium isotope analysis of enamel from Corded Ware culture individuals (Haak et al., 2008), which were included in the genetic analyses presented in Chapter 3, revealed that adult females had spent their childhood in a different place. In comparison, the Corded Ware culture males were considered to have lived their whole life in the same area, having strontium isotope ratio values which were similar to the sediments retrieved from the grave site (Haak et al., 2008). The finding suggests the Corded Ware culture had a practice of exogamy and patrilocality, where females move to where their male partner lived. However, strontium isotope analysis of another Late Neolithic culture, the Bell Beakers, has shown no evidence of exogamy and patrilocality, with both males and females found to be highly mobile (Price et al., 2004).

To determine whether males and females had distinct population histories during the Neolithic period, YDNA would need to be recovered from skeletal remains. However, obtaining YDNA from skeletal remains is extremely difficult, as there is only a single copy of the Y chromosome per cell, whereas there are thousands of copies of mtDNA per cell (Shadel and Clayton, 1997). The study presented in Appendix B (Haak et al., 2010) highlighted the difficulty of obtaining YDNA from skeletal remains. Of the eight samples identified as males from morphological analysis, YDNA could only be amplified from three individuals. This low success rate was achieved even though very short fragments of DNA were targeted for amplification (less than 90 base pairs). Unsuccessful attempts were made to obtain YDNA from the samples examined in Chapter 3, using the same short fragments. The development of next generation sequencing technologies, which enable deep sequencing of samples, has provided a way to obtain nuclear DNA from ancient samples, such as from the woolly mammoth (Miller et al., 2008), Neanderthal (Green et al., 2010) and Palaeo-Eskimo remains (Rasmussen et al., 2010). Metagenomic sequencing of ancient DNA extracts increases the chance of recovering nuclear DNA by providing deep sequencing of all areas of the genome,

compared to PCR based approaches, which although could be used to amplify all areas of the genome, would be extremely costly and time consuming.

5.4 Summary of Aim 3

To demonstrate that dental calculus preserves ancient oral microbial DNA. Investigate whether temporal variation can be detected in the composition and diversity of oral microbiota recovered from ancient dental calculus deposits from early agriculturist populations and, if so, whether this could be linked to dietary changes brought about by the Neolithic Revolution.

Chapter 4 demonstrated that dental calculus recovered from human remains (dated between 5, 500 BC – 850 AD) contains preserved DNA from ancient oral microbes, with a similar phylum makeup to modern dental plaque and saliva samples. Importantly, the microbial composition was distinct from laboratory reagents (non-template controls) and environmental samples, such as soils, sediments and freshwater. The diversity of bacteria present in the mouths of Neolithic, Bronze Age and Medieval Europeans was significantly higher than that found today. Between the Neolithic and Medieval period, there was a dominance of bacteria involved in oral health and periodontal disease. In contrast, modern Europeans have a restricted suite of bacteria, mainly associated with tooth decay (caries). The results presented in this chapter suggest that periodontal disease was an early consequence of an agricultural diet, as periodontal disease is largely absent prior to the Neolithic as indicated by morphological studies (Kerr, 1991, Larsen, 1995, Aufderheide et al., 1998, Kerr, 1998). In comparison, dental decay-associated bacteria only became abundant recently, and this was potentially caused by the introduction of milled flour and processed sugar to the diet during the Industrial Revolution (150-200 years ago). These results suggest that changes in human ecology have resulted in the disruption of co-evolved mutualism between humans and their microbiota, as suggested by the change in oral microbial diversity and composition since the introduction of an agricultural diet in the Neolithic compared to today. The development of treatments for diseases caused by commensal microbiota requires an understanding of their shared evolutionary history with humans, so these findings have significant implications for human health, providing insight into when, how, and why microbiota became pathogenic. Furthermore, as the evolution of human microbiota becomes obscured by increasing admixture amongst present-day populations and globalisation of Western-type diets, archaeological dental calculus may be the only surviving evidence of the indigenous microbiota of distinct human groups in the pre-colonial era.

5.4.1 Significance of the research

As well as highlighting the complexity of population dynamics during the Neolithic in Central Europe, this thesis has revealed the impact of the introduction of an agriculture diet on oral health. Chapter 4 has shown that at the onset of farming there was a high frequency of periodontal disease-associated pathogens. This finding is confirmed by many morphological studies, which have consistently found that agricultural populations have a higher frequency of periodontal disease than hunter-gatherers (Kerr, 1991, Larsen, 1995, Aufderheide et al., 1998, Kerr, 1998). Although the morphological assessment of skeletal remains can be used to identify the presence of periodontal disease, this type of analysis cannot be used to identify the specific pathogens which are associated with the disease. Furthermore, the information obtained from morphological studies cannot be used to determine the overall composition and diversity of commensal microbial communities in the mouth. Prior to the genetic analysis of archaeological dental calculus, there was no direct information about the past composition of oral microbiota or knowledge about whether the composition and diversity of oral microbiota had changed since the development of an agricultural diet.

The identification of dental calculus as a previously untapped source of ancient oral pathogen and commensal microbial DNA provides a new way to study the evolution of disease and the response of human microbiota to changes in human ecology, such as diet. Presently, ancient pathogen DNA studies have had limited success, and therefore we have limited knowledge about the genetic evolution of disease. These studies have focused on attempting to extract ancient pathogen DNA from skeletal remains that morphologically present with a specific disease, such as tuberculosis (Salo et al., 1994, Rothschild et al., 2001, Zink et al., 2001, Brosch et al., 2002, Taylor et al., 2005), leprosy (Haas et al., 2000, Montiel et al., 2003), malaria (Sallares, 2001) and syphilis (Kolman et al., 1999). However, a limited number of these studies have been able to authenticate their findings, as the majority of ancient pathogen DNA investigations do not comply with established ancient DNA authentication protocols (Roberts and Ingham, 2008). Furthermore, these past studies may not have been able to retrieve ancient pathogen DNA from skeletal remains (Bouwman and Brown, 2005) because the pathogen may no longer be present in the bone or is there at extremely low concentrations (von Hunnius, 2007). In comparison, dental calculus is known to be a dense source of calcified oral microbiota from studies of modern dental calculus (Sidaway, 1978, Lieverse, 1999) and microscopic studies of archaeological dental calculus deposits (Dobney and Brothwell, 1986). The genetic analysis of ancient dental calculus deposits confirms past microscopic analysis, that dental calculus is a dense source of oral mcirobiota.

The genetic analysis of dental calculus from early agriculturists has also shown that the ecosystem of the mouth has drastically altered, becoming less diverse, since humans increased their consumption of carbohydrates (Tauber, 1981, Macko et al., 1999, Lillie and Richards, 2000, Richards, 2002). For example, the genetic analysis of dental calculus from Neolithic, Bronze Age and Medieval remains revealed that in the pre-industrial stages of agriculture (before the introduction of processed sugar and flour), there was a lower frequency of decay-associated bacteria and higher overall oral bacterial diversity. Therefore, reducing the intake of dietary sugar could assist in decreasing the prevalence of caries and potentially improving overall oral microbial diversity. In general, a less diverse ecosystem is less resilient to perturbations (Petchey and Gaston, 2009), which in the oral environment would typically be caused by dietary changes that may alter the pH of the mouth and by the invasion (Loreau et al., 2002) of pathogenic bacterial species. However, the reduced level of oral microbial diversity observed today compared to in pre-Industrial populations may also be attributed to the recent development of oral hygiene practices, such as tooth brushing, flossing and the use of anti-bacterial mouth washes. All these oral hygiene practices disrupt plaque formation and hence bacterial succession within the oral biofilm (Loe and Schiott, 1970, Addy et al., 1987, Stewart and Wolfe, 1989, McBain et al., 2003).

The results presented in this thesis provide information about the wider diversity of commensal human bacteria and hence assist in understanding the co-evolution of humans and their microbiota. Understanding the evolution of human microbiota is important as it has been shown that our commensal bacteria have the ability to modify our genomes (Lander et al., 2001) and cause disease (Aas et al., 2005). As human microbiota influence many aspects of human genetics and physiology, it has been postulated that to understand human evolution, we also need to understand how humans have co-evolved with their microbiota (Turnbaugh et al., 2007). The genetic analysis of dental calculus provides previously unattainable information about the impact of past selective pressures, such as diet, population change and behaviour, on commensal oral bacteria.

The genetic analysis of archaeological dental calculus samples may also be used to study population change during the Neolithic. Oral microbiota may be used to investigate population change because commensal oral bacteria are predominantly inherited (Suchett-Kaye et al., 1999). An individual's oral microbiota has been found to be primarily transferred vertically in early childhood from their primary care-giver(s), with a lesser proportion being transferred horizontally later in life, as indicated by the analysis of oral microbiota within families (Suchett-Kaye et al., 1999). This feature of oral microbiota may be used to investigate population change between the Early Neolithic and Late Neolithic/Early Bronze

Age cultures in Central Europe. If the Late Neolithic/Bronze Age cultures represent a distinct population to the Early Neolithic cultures, as indicated by the analysis of ancient human DNA, then the oral microbial composition may also be distinct. The bacterial composition of dental calculus was found to vary significantly between the Early Neolithic culture, the LBK and Late Neolithic/Bronze Age cultures, which included the Bell Beakers although the sample sizes were extremely small. However, this preliminary evidence suggests that the information obtained from the genetic analysis of ancient dental calculus has the potential to be used to trace population movement in a similar manner to mtDNA or YDNA, in addition to being used to investigate the response of microbiota to cultural and dietary changes brought about by the Neolithic Revolution.

5.4.2 Limitations of the research and future directions

In this thesis, I presented the proof-of-concept work to demonstrate that archaeological dental calculus preserves ancient oral microbial DNA, and as such, there are many areas of research that were not addressed in the study. The investigation was performed using a relatively small dataset, restricted to only European agriculturists and only one gene (16S rRNA) was assessed. Hence, from the research presented so far, the state of the oral environment before the introduction of an agricultural diet is unknown, as is the impact of the introduction of different cereals, other than wheat and barley, to the diet on the oral environment. Furthermore, although the data suggested that the genetic information from dental calculus could be used to study population change, it was not clearly demonstrated. Also, although the results were critically assessed, not all the ancient DNA authentication criteria were able to be applied, therefore new criteria need to be established for the authentication of genetic results from dental calculus. In addition, the changes in function of the oral pathogens over time cannot be determined from the analysis of the 16S gene, which only provides taxonomic information about the presence or absence of bacterial species.

The research presented in Chapter 4 was restricted to the analysis of dental calculus from agricultural populations. Hence, this provides no information about the composition of oral microbiota in pre-agricultural, hunter-gatherer populations. Since humans have spent the majority of their existence as hunter-gatherers (Milton, 1999), it is important to investigate the state of the oral environment before the development of an agricultural diet, as that will enable us to determine the degree to which the oral environment has been altered by this dietary transition. It is expected that the microbial makeup of dental calculus in hunter-gatherers would be considerably different from agriculturists, as morphological studies have shown that oral diseases, such as caries and periodontal disease, were rare amongst hunter-gatherer populations (Kerr, 1991, Larsen, 1995, Aufderheide et al., 1998, Kerr, 1998, Cordain

et al., 2005). However, the acquisition of dental calculus from hunter-gatherer remains will be more difficult than from agriculturists, as hunter-gatherers tend to have lower levels of calculus deposition (Aufderheide et al., 1998). In addition, there are much fewer skeletal remains recovered from prehistoric hunter-gatherer compared to agricultural populations, because they tended to not inhumane their dead.

Aside from only investigating agricultural populations, this study was restricted to investigating European farming populations. Hence, for all the individuals investigated, the primary carbohydrate source would have been obtained from wheat and barley (Willcox, 1996, Smith, 2001, Kuijt and Goring-Morris, 2002, Cordain et al., 2005). The source of carbohydrate is known to have varied effects on dental health (Tayles et al., 2000). For example, the introduction of maize in the Americas (Schmidt, 2001) and wheat and barley in the Near East and Europe (Smith, 1984) has been found to be associated with the increased incidence of caries. In contrast, the introduction of rice to the diet in Asia was not correlated with an increased frequency of dental caries (Tayles et al., 2000). The impact of different cereals on the oral environment could be determined by genetically analysing dental calculus from the Americas and Asia, in addition to Europe.

The results presented in Chapter 4 also indicated that dental calculus may be used to investigate population change as there was a significantly greater similarity in the composition of oral microbiota between individuals from the same archaeological group compared to individuals from different groups. Although the dataset was too small to draw substantial conclusions, the finding highlights the potential of using ancient oral microbial DNA to investigate past population dynamics. It would be interesting to compare the oral microbial composition of past groups that are thought to be mixing, such as the Corded Ware and Unetice cultures (Gimbutas, 1963, Bogucki and Grygiel, 1993, Gallagher et al., 2009), and those which are thought to be distinct, such as the Corded Ware and Bell Beakers (Vander Linden, 2007), as indicated by archaeological evidence and ancient human DNA analysis.

The oral micorbiota in dental calculus may also reveal phylogeographic information about human populations. It has been postulated that any phylogeographic separation which was previously present in the commensal microbiota of populations has probably been erased today due to the high levels of admixture amongst people (Ley et al., 2008). Hence, the genetic analysis of ancient oral microbial DNA may be the only direct way to determine the indigenous state of commensal bacteria for modern populations. Ancient oral microbial DNA retrieved from dental calculus from geographically distinct populations could be compared to mtDNA and YDNA from the same populations, to provide an alternative view of population

history, which could potentially highlight behavioural similarities between groups or trade routes that may not be visible from the analysis of human DNA.

The potential for dental calculus to be used as a population maker lends itself to the interesting application of studying admixture between humans and Neandertals. There have been several findings of dental calculus deposits on Neandertal teeth; from El Sidrón in Spain (Rosas et al., 2006), Shanidar Cave in Iraq, Spy Cave in Belgium (Henry et al., 2011) and from Croatia (Frayer and Russell, 1987). The oral microbial composition of dental calculus from archaic humans and Neandertals could be used to assess the level of admixture between them. Aside for assessing admixture, the type of oral microbiota present in Neandertal dental calculus would provide information about their diet and oral health. This information could be used in conjunction with previous findings of microfossils in Neandertal dental calculus (Henry et al., 2011) to build a picture of the past diet and lifestyle of Neandertals.

The use of dental calculus for genetic analysis provides challenges which have yet to be faced in ancient DNA research. One of the key authentication criteria in ancient DNA research is the independent replication of results (Cooper and Poinar, 2000). However, independent replication of results may not be feasible for dental calculus samples which are often extremely small meaning that often only a single sample may be available for genetic analysis. Furthermore, performing an independent extraction in another laboratory may also be difficult. The successful recovery of endogenous ancient oral microbial DNA from dental calculus requires a work-room dedicated to the study of environmental samples. This study has relied on the use of non-template controls to assess if the dental calculus samples contained endogenous oral microbiota. The non-template controls were deeply sequenced, with on average 3000 sequences generated for each non template control, for each region of 16S investigated. The non-template control sequences were compared to the dental calculus samples through taxonomic and phylogenetic analysis. If the dental calculus samples contained significantly different sequences to the non-template controls, then the sequences were considered to be endogenous to the sample and not contaminants. In the future, it may be useful to set a benchmark for the level of background sequences within a dental calculus sample, for example at 5%, or that the sequences which are thought to be environmental background are removed prior to analysis. Additionally, to authenticate the findings, the sequences from the ancient dental calculus samples were compared to published sequences from environmental samples (Nogales et al., 2001, Ellis et al., 2003, Tringe et al., 2005, Schloss and Handelsman, 2006, Elshahed et al., 2008), such as soil and water. For future projects, when archaeological samples are being excavated, it would be desirable to collect a

soil sample from the site for microbial analysis and this will assist in identifying the presence of environmental contaminants in the dental calculus samples.

While this study provided fundamental information about the presence or absence of bacteria in dental calculus from amplification of the 16S rRNA gene, these results provide no information about the changes in function of oral pathogens during ecological changes, such as dietary shifts. Functional genes would need to be analysed to investigate the evolution of virulence in oral pathogens. An example of an oral pathogen which appears to have altered in frequency since the introduction of carbohydrate to the diet is *Streptococcus mutans.* The known virulence gene of *Streptococcus mutans,* is *gtfB* (Yoshida et al., 2003) and this encodes glycosyltransfareses that catalyse the production of extracellular glucan and facilitates the adherence of the bacteria to smooth surfaces (Ajdic et al., 2002, Napimoga et al., 2005). It is thought that the *gtfB* gene evolved via gene duplication and subsequent sequence change in response to relatively recent introduction of processed sugar to the human diet (Yamashita et al., 1993). This hypothesis could be tested using the ancient sequences obtained from dental calculus. Furthermore, by investigating the changes in *gtfB* over time, the selective pressures driving the changes could be identified. Using the genetic information obtained from ancient dental calculus is one of the only ways currently available to determine the timing, mechanism of acquisition and assembly of known virulence genes from human pathogens in real time.

5.5 Concluding remarks

Ancient DNA methods were used in this thesis to provide novel information about the impact of the Neolithic Revolution on both past and present, European populations. Ancient human DNA from Early Neolithic and Late Neolithic/Early Bronze Age remains showed that the Neolithic period was a time of population change, and that while the findings supported the demic diffusion model of the spread of farming, they also demonstrated that the early farming populations were not fully incorporated into the succeeding populations of the Late Neolithic and Early Bronze Age. The research presented in this thesis also suggest that periodontal disease was an early consequence of dietary changes brought about by the introduction of agriculture, as determined by the genetic analysis of ancient dental calculus. Furthermore, the genetic analysis of ancient dental calculus deposits revealed that the oral environment today is substantially different compared to during the early stages of agriculture, between the Neolithic and Medieval periods, as the mouth is now less diverse and dominated by decay-associated bacteria, which is possibly due to the introduction of sugar during the Industrial

Revolution. The application of ancient DNA techniques to examine dental calculus, enabled for the first time, the direct analysis of the past state of human associated microbiota, which will have multiple implications for the study of prehistoric human lifestyle. This thesis also contributed more generally to the field of ancient DNA by highlighting that commonly used sampling techniques were detrimental to the recovery of DNA from skeletal remains. By combining three different approaches to ancient DNA research, this thesis has provided original insights into the lives of prehistoric humans, which contributes information to the fields of archaeology, anthropology, modern and ancient population genetics, forensic science and medical research.

REFERENCES

Aas, J A, Paster, B J, Stokes, L N, Olsen, I & Dewhirst, F E 2005. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol,* 43**,** 5721-32.

Achilli, A, Rengo, C, Battaglia, V, Pala, M, Olivieri, A, Fornarino, S, Magri, C, Scozzari, R, Babudri, N, Santachiara-Benerecetti, A S, Bandelt, H J, Semino, O & Torroni, A 2005. Saami and Berbers--an unexpected mitochondrial DNA link. *Am J Hum Genet,* 76**,** 883-6.

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, Loogvali, 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**,** 910-8.

Addy, M, Griffiths, G, Dummer, P, Kingdom, A & Shaw, W C 1987. The distribution of plaque and gingivitis and the influence of toothbrushing hand in a group of South Wales 11-12 year-old children. *J Clin Periodontol,* 14**,** 564-72.

Ajdic, D, Mcshan, W M, Mclaughlin, R E, Savi, G, Chang, J, Carson, M B, Primeaux, C, Tian, R, Kenton, S & Jia, H 2002. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 14434.

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

Aufderheide, A C, Rodriguez-Martin, C & Langsjoen, O 1998. *The Cambridge encyclopedia of human paleopathology* Cambridge Cambridge University Press.

Bar-Yosef, O & Belfer-Cohen, A 1992. From foraging to farming in the Mediterranean Levant. *In:* GEBAUER, A. B. & PRICE, T. D. (eds.) *Transitions to Agriculture in Prehistory.* Madison: Prehistory Press.

Bentley, R A 2006. Strontium Isotopes from the Earth to the Archaeological Skeleton: A Review. *J Archaeol Method Theory,* 13.

Berkovitz, B, Holland, G & Moxham, B 1992. *A colour atlas and textbook of oral anatomy, histology and embryology,* London, Wolfe.

Bogucki, P & Grygiel, R 1993. The first farmers of Central Europe: A survey article *Journal of Field Archaeology,* 20**,** 399-426.

Bosshardt, D D & Schroeder, H E 1991. Initiation of acellular extrinsic fiber cementum on human teeth. A light- and electron-microscopic study. *Cell Tissue Res,* 263**,** 311-24.

Bouwman, A S & Brown, T A 2005. The limits of biomolecular palaeopathology: ancient DNA cannot be used to study venereal syphilis *Journal of Archaeological Science,* 32**,** 703-713.

Bramanti, B, Thomas, M G, Haak, W, Unterlaender, M, Jores, P, Tambets, K, Antanaitis-Jacobs, I, Haidle, M N, Jankauskas, R, Kind, C J, Lueth, F, Terberger, T, Hiller, J, Matsumura, S, Forster, P & Burger, J 2009. Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. *Science*.

Brosch, R, Gordon, S V, Marmiesse, M, Brodin, P, Buchrieser, C, Eiglmeier, K, Garnier, T, Gutierrez, C, Hewinson, G, Kremer, K, Parsons, L M, Pym, A S, Samper, S, Van Soolingen, D & Cole, S T 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proc Natl Acad Sci U S A,* 99**,** 3684-9.

Carda, C & Peydro, A 2006. Ultrastructural patterns of human dentinal tubules, odontoblasts processes and nerve fibres. *Tissue Cell,* 38**,** 141-50.

Cavalli-Sforza, L L, Menozzi, P & Piazza, A 1993. Demic expansions and human evolution. *Science,* 259**,** 639-46.

Cavalli-Sforza, L L, Menozzi, P & Piazza, A 1994. *The history and geography of human genes,* Princeton, Princeton University Press.

Cooper, A & Poinar, H N 2000. Ancient DNA: do it right or not at all. *Science,* 289**,** 1139.

Cordain, L, Eaton, S B, Sebastian, A, Mann, N, Lindeberg, S, Watkins, B A, O'keefe, J H & Brand-Miller, J 2005. Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr,* 81**,** 341-54.

De Leo, D, Turrina, S & Marigo, M 2000. Effects of individual dental factors on genomic DNA analysis. *Am J Forensic Med Pathol,* 21**,** 411-5.

Dobney, K M & Brothwell, D R 1986. Dental calculus: its relevance to ancient diet and oral ecology. *In:* CRUWYS, E. & FOLEY, R. A. (eds.) *Teeth and Anthropology.* B.A.R. International Series.

Donlon, D 1994. Aboriginal skeletal collections and research in physical anthropology: An historical perspective. *Australian Archaeology,* 39**,** 73-82

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**,** 1185-92.

Ellis, R J, Morgan, P, Weightman, A J & Fry, J C 2003. Cultivation-dependent and -independent approaches for determining bacterial diversity in heavy-metal-contaminated soil. *Appl Environ Microbiol,* 69**,** 3223-30.

Elshahed, M S, Youssef, N H, Spain, A M, Sheik, C, Najar, F Z, Sukharnikov, L O, Roe, B A, Davis, J P, Schloss, P D, Bailey, V L & Krumholz, L R 2008. Novelty and uniqueness patterns of rare members of the soil biosphere. *Appl Environ Microbiol,* 74**,** 5422-8.

Flannery, K V 1973. The origins of agriculture. *An. Rev. Anthropol.,* 2**,** 271 – 310.

Frayer, D W & Russell, M D 1987. Artificial grooves on the Krapina Neanderthal teeth. *Am J Phys Anthropol,* 74**,** 393-405.

Gallagher, A, Gunther, M M & Bruchhaus, H 2009. Population continuity, demic diffusion and Neolithic origins in central-southern Germany: the evidence from body proportions. *Homo,* 60**,** 95-126.

Gilbert, M T, Willerslev, E, Hansen, A J, Barnes, I, Rudbeck, L, Lynnerup, N & Cooper, A 2003. Distribution patterns of postmortem damage in human mitochondrial DNA. *Am J Hum Genet,* 72**,** 32-47.

Gimbutas, M 1963. The Indo-Europeans: Archeological Problems *American Anthropologist,* 65**,** 815-836.

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, Prufer, K, Meyer, M, Burbano, H A, Good, J M, Schultz, R, Aximu-Petri, A, Butthof, A, Hober, B, Hoffner, 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, Lalueza-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 & Paabo, S 2010. A draft sequence of the Neandertal genome. *Science,* 328**,** 710-22.

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 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biol,* 8**,** e1000536.

Haak, W, Brandt, G, De Jong, H N, Meyer, C, Ganslmeier, R, Heyd, V, Hawkesworth, C, Pike, A W, Meller, H & Alt, K W 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**,** 18226-31.

Haak, W, Forster, P, Bramanti, B, Matsumura, S, Brandt, G, Tanzer, 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**,** 1016-8.

Haas, C J, Zink, A, Palfi, G, Szeimies, U & Nerlich, A G 2000. Detection of leprosy in ancient human skeletal remains by molecular identification of Mycobacterium leprae. *Am J Clin Pathol,* 114**,** 428-36.

Henry, A G, Brooks, A S & Piperno, D R 2011. Microfossils in calculus demonstrate consumption of plants and cooked foods in Neanderthal diets (Shanidar III, Iraq; Spy I and II, Belgium). *Proc Natl Acad Sci U S A,* 108**,** 486-91.

Hillery, M T & Shuaib, I 1999. Temperature effects in the drilling of human and bovine bone. *J Mater Process Technol,* 92-93**,** 302-308.

Hoss, M & Paabo, S 1993. DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Res,* 21**,** 3913-4.

Jones, M K & Liu, X 2009. Archaeology. Origins of agriculture in East Asia. *Science,* 324**,** 730-1.

Kerr, N W 1991. Prevalence and natural history of periodontal disease in Scotland--the medieval period (900-1600 A.D.). *J Periodontal Res,* 26**,** 346-54.

Kerr, N W 1998. Prevalence and natural history of periodontal disease in prehistoric Scots (pre-900 AD). *J Periodontal Res,* 33**,** 131-7.

Kolman, C J, Centurion-Lara, A, Lukehart, S A, Owsley, D W & Tuross, N 1999. Identification of Treponema pallidum subspecies pallidum in a 200-year-old skeletal specimen. *J Infect Dis,* 180**,** 2060-3.

Kuijt, I & Goring-Morris, N 2002. Foraging, farming, and social complexity in the pre-pottery Neolithic of the Southern Levant: A review and synthesis *Journal of World Prehistory,* 16**,** 361 - 440

Lander, E S, Linton, L M, Birren, B, Nusbaum, C, Zody, M C, Baldwin, J, Devon, K, Dewar, K, Doyle, M, Fitzhugh, W, Funke, R, Gage, D, Harris, K, Heaford, A, Howland, J, Kann, L, Lehoczky, J, Levine, R, Mcewan, P, Mckernan, K, Meldrim, J, Mesirov, J P, Miranda, C, Morris, W, Naylor, J, Raymond, C, Rosetti, M, Santos, R, Sheridan, A, Sougnez, C, Stange-Thomann, N, Stojanovic, N, Subramanian, A, Wyman, D, Rogers, J, Sulston, J, Ainscough, R, Beck, S, Bentley, D, Burton, J, Clee, C, Carter, N, Coulson, A, Deadman, R, Deloukas, P, Dunham, A, Dunham, I, Durbin, R, French, L, Grafham, D, Gregory, S, Hubbard, T, Humphray, S, Hunt, A, Jones, M, Lloyd, C, Mcmurray, A, Matthews, L, Mercer, S, Milne, S, Mullikin, J C, Mungall, A, Plumb, R, Ross, M, Shownkeen, R, Sims, S, Waterston, R H, Wilson, R K, Hillier, L W, Mcpherson, J D, Marra, M A, Mardis, E R, Fulton, L A, Chinwalla, A T, Pepin, K H, Gish, W R, Chissoe, S L, Wendl, M C, Delehaunty, K D, Miner, T L, Delehaunty, A, Kramer, J B, Cook, L L, Fulton, R S, Johnson, D L, Minx, P J, Clifton, S W, Hawkins, T, Branscomb, E, Predki, P, Richardson, P, Wenning, S, Slezak, T, Doggett, N, Cheng, J F, Olsen, A, Lucas, S, Elkin, C, Uberbacher, E, Frazier, M, et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

Larsen, C S 1995. Biological changes in human populations with agriculture. *Ann Rev of Anthr,* 24**,** 185-213.

Ley, R E, Lozupone, C A, Hamady, M, Knight, R & Gordon, J I 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol,* 6**,** 776-88.

Lieverse, A R 1999. Diet and the aetiology of dental calculus *International Journal of Osteoarchaeology,* 9**,** 219 - 232.

Lillie, M C & Richards, M 2000. Stable isotope analysis and dental evidence of diet at the Mesolithic–Neolithic transition in Ukraine. *Journal of Archaeological Science,* 27**,** 965-972

Loe, H & Schiott, C R 1970. The effect of mouthrinses and topical application of chlorhexidine on the development of dental plaque and gingivitis in man. *J Periodontal Res,* 5**,** 79-83.

Loogvali, E L, Roostalu, U, Malyarchuk, B A, Derenko, M V, Kivisild, T, Metspalu, E, Tambets, K, Reidla, M, Tolk, H V, Parik, J, Pennarun, E, Laos, S, Lunkina, A, Golubenko, M, Barac, L, Pericic, M, Balanovsky, O P, Gusar, V, Khusnutdinova, E K, Stepanov, V, Puzyrev, V, Rudan, P, Balanovska, E V, Grechanina, E, Richard, C, Moisan, J P, Chaventre, A, Anagnou, N P, Pappa, K I, Michalodimitrakis, E N, Claustres, M, Golge, M, Mikerezi, I, Usanga, E & Villems, R 2004. Disuniting

uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol,* 21**,** 2012-21.

Loreau, M, Downing, A, Emmerson, M, Gonzalez, A, Hughes, J, Inchausti, P, Joshi, J, Norberg, J & Sala, O 2002. A new look at the relationship between diversity and stability. *In:* LOREAU, M., NAEEM, S. & INCHAUSTI, P. (eds.) *Biodiversity and Ecosystem Functioning. Synthesis and Perspectives.* Oxford: Oxford University Press.

Macko, S A, Lubec, G, Teschler-Nicola, M, Andrusevich, V & Engel, M H 1999. The Ice Man's diet as reflected by the stable nitrogen and carbon isotopic composition of his hair. *FASEB J,* 13**,** 559-62.

Mcbain, A J, Bartolo, R G, Catrenich, C E, Charbonneau, D, Ledder, R G & Gilbert, P 2003. Effects of a chlorhexidine gluconate-containing mouthwash on the vitality and antimicrobial susceptibility of in vitro oral bacterial ecosystems. *Appl Environ Microbiol,* 69**,** 4770-6.

Menozzi, P, Piazza, A & Cavalli-Sforza, L 1978. Synthetic maps of human gene frequencies in Europeans. *Science,* 201**,** 786-92.

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**,** 387-90.

Milton, K 1999. Nutritional characteristics of wild primate foods: do the diets of our closest living relatives have lessons for us? *Nutrition,* 15**,** 488-98.

Montiel, R, Garcia, C, Canadas, M P, Isidro, A, Guijo, J M & Malgosa, A 2003. DNA sequences of Mycobacterium leprae recovered from ancient bones. *FEMS Microbiol Lett,* 226**,** 413-4.

Napimoga, M H, Höfling, J F, Klein, M I, Kamiya, R U & Gonçalves, R B 2005. Transmission, diversity and virulence factors of *Streptococcus mutans* genotypes. *Journal of Oral Science,* 47**,** 59-64.

Nogales, B, Moore, E R, Llobet-Brossa, E, Rossello-Mora, R, Amann, R & Timmis, K N 2001. Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. *Appl Environ Microbiol,* 67**,** 1874-84.

Petchey, O & Gaston, K 2009. Effects on ecosystem resilience of biodiversity, extinctions, and the structure of regional species pools. *Theoretical Ecology,* 2**,** 177-187.

Price, T D, Knipper, C, Grupe, G & Smrcka, V 2004. Strontium isotopes and prehistoric human migration: The Bell Beaker period in central Europe. *European Journal of Archaeology,* 7**,** 9-40.

Pruvost, M, Schwarz, R, Correia, V B, Champlot, S, Grange, T & Geigl, E 2008. DNA diagenesis and palaeogenetic analysis: Critical assessment and methodological progress *Palaeogeogr Palaeoclimatol Palaeoecol,* 266**,** 211-219.

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, Gronnow, 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-Ponten, T, Villems, R, Nielsen, R, Krogh, A, Wang, J & Willerslev, E 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature,* 463**,** 757-62.

Richards, M, Corte-Real, H, Forster, P, Macaulay, V, Wilkinson-Herbots, H, Demaine, A, Papiha, S, Hedges, R, Bandelt, H J & Sykes, B 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet,* 59**,** 185-203.

Richards, M, Macaulay, V, Hickey, E, Vega, E, Sykes, B, Guida, V, Rengo, C, Sellitto, D, Cruciani, F, Kivisild, T, Villems, R, Thomas, M, Rychkov, S, Rychkov, O, Rychkov,

Y, 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 Rienzo, A, Novelletto, A, Oppenheim, A, Norby, S, Al-Zaheri, N, Santachiara-Benerecetti, S, Scozari, R, Torroni, A & Bandelt, H J 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet,* 67**,** 1251-76.

Richards, M, Macaulay, V, Torroni, A & Bandelt, H J 2002. In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet,* 71**,** 1168-74.

Richards, M P 2002. A brief review of the archaeological evidence for Palaeolithic and Neolithic subsistence. *Eur J Clin Nutr,* 56**,** 16 p following 1262.

Roberts, C & Ingham, S 2008. Using ancient DNA analysis in palaeopathology: A critical analysis of published papers, with recommendations for future work. *International Journal of Osteoarchaeology,* 18**,** 600–613.

Rootsi, S, Magri, C, Kivisild, T, Benuzzi, G, Help, H, Bermisheva, M, Kutuev, I, Barac, L, Pericic, M, Balanovsky, O, Pshenichnov, A, Dion, D, Grobei, M, Zhivotovsky, L A, Battaglia, V, Achilli, A, Al-Zahery, N, Parik, J, King, R, Cinnioglu, C, Khusnutdinova, E, Rudan, P, Balanovska, E, Scheffrahn, W, Simonescu, M, Brehm, A, Goncalves, R, Rosa, A, Moisan, J P, Chaventre, A, Ferak, V, Furedi, S, Oefner, P J, Shen, P, Beckman, L, Mikerezi, I, Terzic, 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**,** 128-37.

Rosas, A, Martinez-Maza, C, Bastir, M, Garcia-Tabernero, A, Lalueza-Fox, C, Huguet, R, Ortiz, J E, Julia, R, Soler, V, De Torres, T, Martinez, E, Canaveras, J C, Sanchez-Moral, S, Cuezva, S, Lario, J, Santamaria, D, De La Rasilla, M & Fortea, J 2006. Paleobiology and comparative morphology of a late Neandertal sample from El Sidron, Asturias, Spain. *Proc Natl Acad Sci U S A,* 103**,** 19266-71.

Rothschild, B M, Martin, L D, Lev, G, Bercovier, H, Bar-Gal, G K, Greenblatt, C, Donoghue, H, Spigelman, M & Brittain, D 2001. Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. *Clin Infect Dis,* 33**,** 305-11.

Rudbeck, L, Gilbert, M T, Willerslev, E, Hansen, A J, Lynnerup, N, Christensen, T & Dissing, J 2005. mtDNA analysis of human remains from an early Danish Christian cemetery. *Am J Phys Anthropol,* 128**,** 424-9.

Sallares, R a G, S. 2001. Biomolecular archaeology of Malaria. *Ancient Biomolecules,* 3**,** 195-213.

Salo, W L, Aufderheide, A C, Buikstra, J & Holcomb, T A 1994. Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. *Proc Natl Acad Sci U S A,* 91**,** 2091-4.

Sampietro, M L, Lao, O, Caramelli, D, Lari, M, Pou, R, Marti, M, Bertranpetit, J & Lalueza-Fox, C 2007. Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc Biol Sci,* 274**,** 2161-7.

Schloss, P D & Handelsman, J 2006. Toward a census of bacteria in soil. *PLoS Comput Biol,* 2**,** e92.

Schmidt, C 2001. Dental microwear evidence for a dietary shift between two nonmaize-reliant prehistoric human populations from Indiana. *Am J Phys Anthropol,* 114**,** 139-45.

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**,** 1155-9.

Shadel, G S & Clayton, D A 1997. Mitochondrial DNA maintenance in vertebrates. *Annu Rev Biochem,* 66**,** 409-35.

Sidaway, D A 1978. A microbiological study of dental calculus. I. The microbial flora of mature calculus. *J Periodontal Res,* 13**,** 349-59.

Smith, B 2001. The transition to food production. *In:* IN FEINMAN, G. M., AND PRICE, T. D (ed.) *Archaeology at the Millennium: A Sourcebook.* New York: Kluwer Academic/Plenum.

Smith, B H 1984. Patterns of molar wear in hunger-gatherers and agriculturalists. *Am J Phys Anthropol,* 63**,** 39-56.

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**,** R174-83.

Stewart, J E & Wolfe, G R 1989. The retention of newly-acquired brushing and flossing skills. *J Clin Periodontol,* 16**,** 331-2.

Suchett-Kaye, G, Decoret, D & Barsotti, O 1999. Intra-familial distribution of Fusobacterium nucleatum strains in healthy families with optimal plaque control. *J Clin Periodontol,* 26**,** 401-4.

Tauber, H 1981. 13C evidence for dietary habits of prehistoric man in Denmark. *Nature,* 292**,** 332-3.

Tayles, N, Domett, K & Nelsen, K 2000. Agriculture and dental caries? The case of rice in prehistoric Southeast Asia. *World Archaeol,* 32**,** 68-83.

Taylor, G M, Young, D B & Mays, S A 2005. Genotypic analysis of the earliest known prehistoric case of tuberculosis in Britain. *J Clin Microbiol,* 43**,** 2236-40.

Taylor, L H, Latham, S M & Woolhouse, M E 2001. Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci,* 356**,** 983-9.

Torroni, A, Bandelt, H J, D'urbano, L, Lahermo, P, Moral, P, Sellitto, D, Rengo, C, Forster, P, Savontaus, M L, Bonne-Tamir, B & Scozzari, R 1998. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet,* 62**,** 1137-52.

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, Koivumaki, 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**,** 844-52.

Tringe, S G, Von Mering, C, Kobayashi, A, Salamov, A A, Chen, K, Chang, H W, Podar, M, Short, J M, Mathur, E J, Detter, J C, Bork, P, Hugenholtz, P & Rubin, E M 2005. Comparative metagenomics of microbial communities. *Science,* 308**,** 554-7.

Turnbaugh, P J, Ley, R E, Hamady, M, Fraser-Liggett, C M, Knight, R & Gordon, J I 2007. The human microbiome project. *Nature,* 449**,** 804-10.

Vander Linden, M 2007. What linked the Bell Beakers in third millennium BC Europe? . *Antiquity,* 81**,** 343–352.

Von Hunnius, T E, Yang, D.,, Eng, B., Waye, J. S., and Saunders, S. 2007. Digging deeper into the limits of ancient DNA research on syphilis *Journal of Archaeological Science,* 34**,** 2091-2100.

Weisdorf, J L 2005. From foraging to farming: Explaining the neolithic revolution. *Journal of Economic Surveys,* 19**,** 561-586.

Willcox, G 1996. Evidence for plant exploitation and vegetation history from three Early Neolithic pre-pottery sites on the Euphrates (Syria) *Vegetation History and Archaeobotany,* 5**,** 143-152.

Wolfe, N D, Dunavan, C P & Diamond, J 2007. Origins of major human infectious diseases. *Nature,* 447**,** 279-83.

Wood, J W, Milner, G R, Harpending, H C & Weiss, K M 1992. The osteological paradox: Problems of inferring prehistoric health from skeletal samples. *Current Anthropology,* 33**,** 343–370.

Yamashita, Y, Bowen, W, Burne, R & Kuramitsu, H 1993. Role of the *Streptococcus mutans* gtf genes in caries induction in the specific-pathogen-free rat model. *Infection and Immunity,* 61**,** 3811-3817.

Yoshida, A, Suzuki, N, Nakano, Y, Kawada, M, Oho, T & Koga, T 2003. Development of a 5' nuclease-based real-time PCR assay for quantitative detection of cariogenic dental pathogens Streptococcus mutans and Streptococcus sobrinus. *J Clin Microbiol,* 41**,** 4438-41.

Zink, A, Haas, C J, Reischl, U, Szeimies, U & Nerlich, A G 2001. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *J Med Microbiol,* 50**,** 355-66.

APPENDIX A

Published online 5 January 2011 | Nature | doi:10.1038/news.2011.702

News

Researchers to drill for hobbit history

Prospects of recovering ancient DNA from Homo floresiensis boosted by study on teeth.

Cheryl Jones

Homing in on tooth cementum may help researchers extract DNA from Homo floresiensis remains. Peter Schouten/National Geographic Society/Reuters/Corbis

Scientists are planning an attempt to extract DNA from the 'hobbit' Homo floresiensis, the 1-metre-tall extinct distant relative of modern humans that was unearthed in Indonesia, following a study that suggests problems in standard sampling methods in ancient-DNA research could have thwarted previous efforts.

This year, geneticists at the Australian Centre for Ancient DNA (ACAD) at the University of Adelaide hope to recover DNA from a roughly 18,000-year-old H. floresiensis tooth, which was excavated in 2009 from the Liang Bua site on the Indonesian island of Flores.

The premolar has been kept cold, and has been handled as little as possible to prevent contamination with modern DNA. But little, if any, of the ancient DNA is likely to have survived the heat and moisture of the tropics, and any that has may be highly fragmented.

Tony Djubiantono, director of the Indonesian National Centre for Archaeology in Jakarta, where the tooth is held, says that developments in DNA extraction techniques could overcome previous sampling problems, and have exciting potential for understanding the evolutionary history of H. floresiensis.

If the DNA can be extracted, comparing its sequence to that of other species could settle disputes over classification. For instance, Peter Brown, a palaeoanthropologist at the University of New England in Armidale, Australia, who described and named the species in 2004[1], is rethinking his initial classification. At first he put the species in the human genus Homo, but he now suspects that the hobbit's ancestors left Africa before Homo evolved so the species could belong to a different or new genus.

Teething troubles

Five years ago, two teams, one from ACAD and one from the Max Planck Institute of Evolutionary Anthropology in Leipzig, Germany, attempted to recover DNA from another H. floresiensis tooth excavated in 2003. Both attempts failed.

Now, a team led by Christina Adler, a geneticist at ACAD, has found that standard sampling procedures could be responsible for the failure to get DNA from the hobbit and some other ancient specimens.

Adler's team — which included some researchers involved in the original H. floresiensis DNA recovery attempt — compared the impacts of various sampling techniques on DNA from the mitochondria of 40 human specimens from around the world, which had been dated up to 7,500 years old. The results have been accepted for publication in the Journal of Archaeological Science[2].

Most genetics research on ancient teeth has focused on the inner tooth tissue, dentine, but Adler's team found that cementum, the coating of the root, was a richer source of DNA.

Drilling is a technique commonly used to sample teeth and bone, because it minimizes damage to the precious specimen. But Adler's team found that the heat generated at standard drill speeds of more than 1,000 revolutions per minute (RPM) destroys DNA rapidly, causing yields to be up to 30 times lower than for samples pulverized in a mill. Reducing the drill speed to 100 RPM alleviated the problem.

The Max Planck team sampled dentine from the hobbit tooth in its early attempt to recover DNA, but it is unclear what drill speed was used. And although the ACAD scientists used the lower drill speed, they also concentrated on dentine. They will target cementum in their next attempt.

Small chance

Adler says that several ancient specimens that previously failed to yield DNA might now warrant re-sampling. She is surprised that ancient-DNA researchers commonly choose to drill at high speeds for samples, because dentists have long known that this harms their patients' teeth. "This is a case of a lack of communication between two specialist fields that are both working on similar things independently," she says.

ADVERTISEMENT

Matthew Collins, a specialist in ancient-protein analysis who is based at the University of York, UK, says that Adler's team's results will "help to ensure that we minimize the destruction of molecules during sampling of precious fossils, and potentially enable us to reach even further back in time to recover sequence information".

However, he is pessimistic about the chances of winkling DNA out of H. floresiensis, saying that the molecules are probably too fragmented owing to high temperatures at the excavation site.

But the ACAD scientists think that it is worth making an attempt on their H. floresiensis tooth. They have been encouraged by their successful extraction of DNA from a 6,000-year-old pig tooth from the site in 2007[3].

- ## References

1. Brown, P. *et al*. Nature 431, 1055-1061 (2004). | Article
2. Adler, C. J. , Haak, W. , Donlon, D. & Cooper, A. J. Archaeol. Sci. doi:10.1016/j.jas.2010.11.010 (in the press).
3. Larson, G. *et al*. Proc. Natl Acad. Sci. USA 104 4834-4839 (2007). | Article | PubMed

APPENDIX B

Ancient DNA from European Early Neolithic Farmers Reveals Their Near Eastern Affinities

Wolfgang Haak[1]*, Oleg Balanovsky[2], Juan J. Sanchez[3], Sergey Koshel[4], Valery Zaporozhchenko[2,5], Christina J. Adler[1], Clio S. I. Der Sarkissian[1], Guido Brandt[6], Carolin Schwarz[6], Nicole Nicklisch[6], Veit Dresely[7], Barbara Fritsch[7], Elena Balanovska[2], Richard Villems[8], Harald Meller[7], Kurt W. Alt[6], Alan Cooper[1], the Genographic Consortium[¶]

1 Australian Centre for Ancient DNA, School of Earth and Environmental Sciences, University of Adelaide, Adelaide, Australia, **2** Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia, **3** National Institute of Toxicology and Forensic Sciences, Canary Islands Delegation, Campus de Ciencias de la Salud, La Laguna, Tenerife, Spain, **4** Faculty of Geography, Moscow State University, Moscow, Russia, **5** Research Centre for Drug Evaluation, Ministry of Public Health of the Russian Federation, Moscow, Russia, **6** Institute for Anthropology, Johannes Gutenberg University of Mainz, Mainz, Germany, **7** Landesamt für Denkmalpflege und Archaeologie und Landesmuseum für Vorgeschichte, Halle (Saale), Germany, **8** Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu and Estonian Biocentre, Tartu, Estonia

Abstract

In Europe, the Neolithic transition (8,000–4,000 B.C.) from hunting and gathering to agricultural communities was one of the most important demographic events since the initial peopling of Europe by anatomically modern humans in the Upper Paleolithic (40,000 B.C.). However, the nature and speed of this transition is a matter of continuing scientific debate in archaeology, anthropology, and human population genetics. To date, inferences about the genetic make up of past populations have mostly been drawn from studies of modern-day Eurasian populations, but increasingly ancient DNA studies offer a direct view of the genetic past. We genetically characterized a population of the earliest farming culture in Central Europe, the Linear Pottery Culture (LBK; 5,500–4,900 calibrated B.C.) and used comprehensive phylogeographic and population genetic analyses to locate its origins within the broader Eurasian region, and to trace potential dispersal routes into Europe. We cloned and sequenced the mitochondrial hypervariable segment I and designed two powerful SNP multiplex PCR systems to generate new mitochondrial and Y-chromosomal data from 21 individuals from a complete LBK graveyard at Derenburg Meerenstieg II in Germany. These results considerably extend the available genetic dataset for the LBK ($n = 42$) and permit the first detailed genetic analysis of the earliest Neolithic culture in Central Europe (5,500–4,900 calibrated B.C.). We characterized the Neolithic mitochondrial DNA sequence diversity and geographical affinities of the early farmers using a large database of extant Western Eurasian populations ($n = 23{,}394$) and a wide range of population genetic analyses including shared haplotype analyses, principal component analyses, multidimensional scaling, geographic mapping of genetic distances, and Bayesian Serial Simcoal analyses. The results reveal that the LBK population shared an affinity with the modern-day Near East and Anatolia, supporting a major genetic input from this area during the advent of farming in Europe. However, the LBK population also showed unique genetic features including a clearly distinct distribution of mitochondrial haplogroup frequencies, confirming that major demographic events continued to take place in Europe after the early Neolithic.

Citation: Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, et al. (2010) Ancient DNA from European Early Neolithic Farmers Reveals Their Near Eastern Affinities. PLoS Biol 8(11): e1000536. doi:10.1371/journal.pbio.1000536

Academic Editor: David Penny, Massey University, New Zealand

Received March 18, 2010; **Accepted** September 27, 2010; **Published** November 9, 2010

Copyright: © 2010 Haak et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by the German Research Foundation, the Geocycles Research Centre at the University of Mainz, and The Genographic Project. The Genographic Project is supported by funding from the National Geographic Society, IBM, and the Waitt Family Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: ABC, approximate Bayesian computation; ACAD, Australian Centre for Ancient DNA; aDNA, ancient DNA; AIC, Akaike information criterion; BayeSSC, Bayesian Serial Simcoal; cal B.C., calibrated B.C.; hg, haplogroup; HVS-I, hypervariable segment I; LBK, Linear Pottery Culture; MDS, multidimensional scaling; mtDNA, mitochondrial DNA; np, nucleotide position(s); PC, principal component; PCA, principal component analysis; qPCR, quantitative real-time PCR; SBE, single base extension; SNP, single nucleotide polymorphism.

* E-mail: wolfgang.haak@adelaide.edu.au

¶ Membership for the Genographic Consortium is listed in the Acknowledgments section.

Introduction

The transition from a hunter–gatherer existence to a "Neolithic lifestyle," which was characterized by increasing sedentarism and the domestication of animals and plants, has profoundly altered human societies around the world [1,2]. In Europe, archaeological and population genetic views of the spread of this event from the Near East have traditionally been divided into two contrasting positions. Most researchers have interpreted the Neolithic transition as a period of substantial demographic flux (demic diffusion) potentially involving large-scale expansions of farming populations from the Near East, which are expected to have left a detectable genetic footprint [3,4]. The alternative view (cultural diffusion model; e.g., [5]) suggests that indigenous Mesolithic

Author Summary

The transition from a hunter–gatherer existence to a sedentary farming-based lifestyle has had key consequences for human groups around the world and has profoundly shaped human societies. Originating in the Near East around 11,000 y ago, an agricultural lifestyle subsequently spread across Europe during the New Stone Age (Neolithic). Whether it was mediated by incoming farmers or driven by the transmission of innovative ideas and techniques remains a subject of continuing debate in archaeology, anthropology, and human population genetics. Ancient DNA from the earliest farmers can provide a direct view of the genetic diversity of these populations in the earliest Neolithic. Here, we compare Neolithic haplogroups and their diversity to a large database of extant European and Eurasian populations. We identified Neolithic haplotypes that left clear traces in modern populations, and the data suggest a route for the migrating farmers that extends from the Near East and Anatolia into Central Europe. When compared to indigenous hunter–gatherer populations, the unique and characteristic genetic signature of the early farmers suggests a significant demographic input from the Near East during the onset of farming in Europe.

hunter–gatherer groups instead adopted new subsistence strategies with relatively little, or no, genetic influence from groups originating in the Near East.

Genetic studies using mitochondrial DNA (mtDNA) and Y-chromosomal data from modern populations have generated contradictory results, and as a consequence, the extent of the Neolithic contribution to the gene pool of modern-day Europeans is still actively debated [6–8]. Studies that suggest that the genetic variation in modern-day Europe largely reflects farming communities of the Early Neolithic period [9–11] contrast strongly with others that consider the input from the Near East an event of minor importance and ascribe the European genetic variation and its distribution patterns to the initial peopling of Europe by anatomically modern humans in the Upper Paleolithic [12–15]. These patterns are also likely to have been significantly impacted by the early Holocene re-expansions of populations out of southerly refugia formed during the Last Glacial Maximum (~25,000 y ago) and by the numerous demographic events that have taken place in post-Neolithic Europe.

The genetics of prehistoric populations in Europe remain poorly understood, restricting real-time insights into the process of the Neolithic transition [16–21]. As a result, most attempts to reconstruct history have been limited to extrapolation from allele frequencies and/or coalescent ages of mitochondrial and Y chromosome haplogroups (hgs) in modern populations. Ancient DNA (aDNA) analyses now provide a powerful new means to directly investigate the genetic patterns of the early Neolithic period, although contamination of specimens with modern DNA remains a major methodical challenge [22].

A previous genetic study of 24 individuals from the early Neolithic Linear Pottery Culture (LBK; 5,500–4,900 calibrated B.C. [cal B.C.]) in Central Europe detected a high frequency of the currently rare mtDNA hg N1a, and proposed this as a characteristic genetic signature of the Early Neolithic farming population [19]. This idea was recently supported by the absence of this particular lineage (and other now more common European hgs) among sequences retrieved from neighboring Mesolithic populations [20,21]. However, a study of 11 individuals from a Middle/Late Neolithic site on the Iberian Peninsula (3,500–3,000 cal B.C.) did not find significant differences from modern populations, supporting a quite different population genetic model for the Neolithic transition in Iberia [18].

To gain direct insight into the genetic structure of a population at the advent of farming in Central Europe we analyzed a complete graveyard from the Early Neolithic LBK site at Derenburg Meerenstieg II (Harzkreis, Saxony-Anhalt) in Germany. The archaeological culture of the LBK had its roots in the Transdanubian part of the Carpathian Basin in modern-day Hungary approximately 7,500–8,000 y ago and spread during the subsequent five centuries across a vast area ranging from the Paris Basin to the Ukraine [23,24]. The graveyard samples provide a unique view of a local, closed population and permit comparisons with other specimens of the LBK archaeological culture (the contemporaneous meta-population) and with modern populations from the same geographical area (covering the former range of the LBK), as well as groups across the wider context of Western Eurasia. Our primary aim was to genetically characterize the LBK early farming population: by applying comprehensive phylogeographic and population genetic analyses we were able to locate its origins within the broader Eurasian region, and to trace its potential dispersal routes into Europe.

Results/Discussion

We used standard approaches to clone and sequence the mitochondrial hypervariable segment I (HVS-I) and applied quantitative real-time PCR (qPCR) as an additional quality control. In addition, we developed two new multiplex typing assays to simultaneously analyze important single nucleotide polymorphisms (SNPs) within the mtDNA coding region (22 SNPs: GenoCoRe22) and also the Y chromosome (25 SNPs: GenoY25). In addition to minimizing the risk of contamination, the very short DNA fragments (average 60–80 bp) required by this approach maximize the number of specimens that can be genetically typed.

We successfully typed 17 individuals for mtDNA, which together with a previous study [19] provided data for 22 individuals from the Derenburg graveyard (71% of all samples collected for genetic analysis; Tables 1 and S1), and significantly extended the genetic dataset of the LBK ($n = 42$), to our knowledge the largest Neolithic database available. Sequences have been deposited in GenBank (http://www.ncbi.nlm.nih.gov/genbank/; accession numbers HM009339–HM009341, HM009343–HM009355, and HM009358), and detailed alignments of all HVS-I clone sequences from Derenburg are shown in Dataset S1.

Multiplex SNP Typing Assays

All of the mtDNA SNP typing results were concordant with the hg assignments based on HVS-I sequence information (Tables 1 and S1) and the known phylogenetic framework for the SNPs determined from modern populations [25]. The tight hierarchical structure of the latter provides a powerful internal control for contamination or erroneous results. Overall, both multiplex systems proved to be extremely time- and cost-efficient compared to the standard approach of numerous individual PCRs, and required 22–25 times less aDNA template while simultaneously reducing the chances of contamination dramatically. Also, both multiplex assays proved to be a powerful tool for analyzing highly degraded aDNA, and the GenoCoRe22 assay was able to unambiguously type four additional specimens that had failed to amplify more than 100 bp (Table 1) from two independent extractions. However, for reasons of overall data comparability, we could not include these specimens in downstream population

Table 1. Summary of archaeological, genetic, and radiocarbon data.

Sample	Feature	Grave	Age, Sex[a]	Radiocarbon Date (Laboratory Code) (Uncalibrated BP, Cal B.C.) [73]	HVS-I Sequence (np 15997–16409), Minus np 16000	Hg HVS-I	Hg GenoCoRe22	Hg GenoY25
deb09	420	9	Adult, f		rCRS	H	H	
deb06	421	10	Adult/mature, n.d.		Ambiguous	n.d.	H	—
deb11	569	16	Adult, f?		n.d.	n.d.	T	
deb10	566	17	Adult, m		093C, 224C, 311C	K	K	—
deb23	565	18	Infans I, m?		093C, 223T, 292T	W	W	—
deb12I	568	20	Infans I, m?	6,015±35 BP (KIA30400), 4,910±50 cal B.C.	298C	V	V	—
deb03	591	21	Adult, f	6,147±32 BP (KIA30401), 5,117±69 cal B.C.	147A, 172C, 223T, 248T, 320T, 355T	N1a	n.d.	
deb15	593	23	Infans I, f?		126C, 294T, 296T, 304C	T2	T	—
deb05	604/2	29	Infans II, f??		311C	HV	HV[b]	
deb22	604/3	30	Adult/mature, f		092C, 129A, 147A, 154C, 172C, 223T, 248T, 320T, 355T	N1a	N1	—
deb20	599	31	Adult, m	6,257±40 BP (KIA30403), 5,247±45 cal B.C.	311C	HV	HV	F*(xG,H,I,J,K)
deb21	600	32	Mature, f	6,151±27 BP (KIA30404), 5,122±65 cal B.C.	rCRS	H	H	
deb01	598	33	Infans II/Juvenile, f??		147A, 172C, 223T, 248T, 355T	N1a	N1	
deb04	596	34	Adult, m	6,141±33 BP (KIA30402), 5,112±73 cal B.C.	311C	HV	HV[b]	
deb26	606	37	Juvenile, m??		069T, 126C	J	J	—
deb32	640	38	Adult/mature f	6,142±34 BP (KIA30405), 5,112±73 cal B.C.	n.d.	n.d.	T	
deb30	592	40	Adult, f?		069T, 126C	J	J	—
deb29II	649	41	Adult, f?	6,068±31 BP (KIA30406), 4,982±38 cal B.C.	n.d.	n.d.	K	
deb34II	484	42	Adult/mature, m		093C, 223T, 292T	W	W	G2a3
deb33	483	43	Juvenile II, f??		126C, 147T, 293G, 294T, 296T, 297C, 304C	T2	T	—
deb02	644	44	Mature, f		224C, 311C	K	K	—
deb36	645	45	Mature, f		093C, 256T, 270T, 399G	U5a1a	U	
deb38	665	46	Adult/mature, m		093C, 224C, 311C	K	K	F*(xG,H,I,J,K)
deb35II	662	47	Adult, f?		126C, 189C, 294T, 296T	T	T	
deb37I	643	48	Adult/mature f		069T, 126C	J	J	
deb39	708	49	Adult/mature, f	6,148±33 BP (KIA30407), 5,117±69 cal B.C.	126C, 294T, 296T, 304C	T2	T	—

Italicized samples had been described previously [19].
[a]One versus two question marks after sex indicate two levels of insecurity in sexing.
[b]Previously analyzed diagnostic SNP sites at np 7028 AluI (hg H) and np 12308 HinfI (hg U) per restriction fragment length polymorphism.
BP, before present; f, female; m, male; n.d., not determined.
doi:10.1371/journal.pbio.1000536.t001

genetic analyses, which required HVS-I sequence data. The only artifacts detected were occasional peaks in the electropherograms of the SNaPshot reactions outside the bin range of expected signals. These were probably due to primers and were mainly present in reactions from extracts with very little or no DNA template molecules; they were not observed with better preserved samples or modern controls.

In contrast, Y chromosome SNPs could be typed for only three out of the eight male individuals (37.5%; Table S2) identified through physical anthropological examination, reflecting the much

lower copy number of nuclear loci [22]. After typing with the GenoY25 assay, individual deb34 was found to belong to hg G (M201), whereas individuals deb20 and deb38 both fall basally on the F branch (derived for M89 but ancestral for markers M201, M170, M304, and M9), i.e., they could be either F or H (Table 1). To further investigate the hg status beyond the standard GenoY25 assay, we amplified short fragments around SNP sites M285, P287, and S126 to further resolve deb34 into G1, G2*, and G2a3, and around SNP site M69 to distinguish between F and H [26]. deb34 proved to be ancestral for G1-M285 but derived for G2*-P287 and additional downstream SNP S126 (L30), placing it into G2a3. deb20 and deb38 were shown to be ancestral at M69 and hence basal F (M89), and remained in this position because we did not carry out further internal subtyping within the F clade.

The multiplexed single base extension (SBE) approach with its shortened flanking regions around targeted SNPs significantly increases the chance of successful Y-chromosomal amplifications, which have remained problematic for aDNA studies, as have nuclear loci in general, because of the much lower cellular copy number compared to mitochondrial loci. The multiplexed SBE approach promises to open the way to studying the paternal history of past populations, which is of paramount importance in determining how the social organization of prehistoric societies impacted the population dynamics of the past.

Quantitative Real-Time PCR

Results of the qPCR revealed significantly ($p = 0.012$, Wilcoxon signed-ranks test) more mtDNA copies per microliter of each extract for the shorter fragment (141 bp) than for the longer (179 bp), with an average 3.7×10^4–fold increase (detailed results are shown in Table S3). This finding is consistent with previous observations demonstrating a biased size distribution for authentic aDNA molecules [22,27,28] and suggests that any contaminating molecules, which would also result in higher copy numbers in the larger size class, did not significantly contribute to our amplifications.

Population Genetic Analyses

To analyze the Neolithic mtDNA sequence diversity and characterize modern geographical affinities, we applied a range of population genetic analyses including shared haplotype analyses, principal component analyses (PCAs), multidimensional scaling (MDS), geographic mapping of genetic distances, and demographic modeling via Bayesian Serial Simcoal (BayeSSC) analyses (Table 2).

Shared Haplotype Analyses

We prepared standardized modern population datasets of equal size ($n = \sim 500$) from 36 geographical regions in Eurasia ($n = 18{,}039$; Table S4) to search for identical matches with each LBK haplotype. Out of 25 different haplotypes present in 42 LBK samples, 11 are found at high frequency in nearly all present-day populations under study, a further ten have limited geographic distribution, and the remaining four haplotypes are unique to Neolithic LBK populations (Table S4). The 11 widespread haplotypes are mainly basal (i.e., constituting a basal node within the corresponding hg) for Western Eurasian mitochondrial hgs H, HV, V, K, T, and W. While these haplotypes are relatively uninformative for identifying genetic affiliations to extant populations, this finding is consistent within an ancient population (5,500–4,900 cal B.C., i.e., prior to recent population expansions), in which basal haplotypes might be expected to be more frequent than derived haplotypes (e.g., end tips of branches within hgs). The next ten LBK haplotypes were unequally spread among present-day populations and for this reason potentially contain information about geographical affinities. We found nine modern-day population pools in which the percentage of these haplotypes is significantly higher than in other population pools ($p > 0.01$, two-tailed z test; Figure 1; Table S4): (a) North and Central English, (b) Croatians and Slovenians, (c) Czechs and Slovaks, (d) Hungarians and Romanians, (e) Turkish, Kurds, and Armenians, (f) Iraqis, Syrians, Palestinians, and Cypriotes, (g) Caucasus (Ossetians and Georgians), (h) Southern Russians, and (i) Iranians. Three of these pools (b–d) originate near the proposed geographic center of the earliest LBK in Central Europe and presumably represent a genetic legacy from the Neolithic. However, the other matching population pools are from Near East regions (except [a] and [h]), which is consistent with this area representing the origin of the European Neolithic, an idea that is further supported by Iranians sharing the highest number of informative haplotypes with the LBK (7.2%; Table S4). The remaining pool (a) from North and Central England shares an elevated frequency of mtDNA T2 haplotypes with the LBK, but otherwise appears inconsistent with the proposed origin of the Neolithic in the Near East. It has been shown that certain alleles (here hgs) can accumulate in frequency while surfing on the wave of expansion, eventually resulting in higher frequencies relative to the proposed origin [29,30]. Several of the other population pools also show a low but nonsignificant level of matches, which may relate to pre-Neolithic distributions or subsequent demographic movements (Figure 1).

Of the four unique mtDNA haplotypes, two were from an earlier study of the LBK (16286-16304 and 16319-16343; Table S5 and [19]). The haplotype 16286-16304 has many one- or two-step derivates in all parts of Europe and is therefore rather uninformative for inferring further geographical affinities. The only relatively close neighbor of haplotype 16319-16343 is found in Iraq (16129-16189-16319-16343), in agreement with the Near Eastern affinities of the informative LBK haplotypes. The other two unique LBK haplotypes belong to N1a, the characteristic LBK hg. The frequency of N1a was 13.6% for Derenburg samples (3/22) and 14.3% for all LBK samples published to date (6/42). Notably, N1a has not yet been observed in the neighboring hunter–gatherer populations of Central Europe before, during, or after the Early Neolithic [20] nor in the early Neolithic Cardial Ware Culture from Spain [18].

The Y chromosome hgs obtained from the three Derenburg early Neolithic individuals are generally concordant with the mtDNA data (Table 1). Interestingly, we do not find the most common Y chromosome hgs in modern Europe (e.g., R1b, R1a, I, and E1b1), which parallels the low frequency of the very common modern European mtDNA hg H (now at 20%–50% across Western Eurasia) in the Neolithic samples. Also, while both Neolithic Y chromosome hgs G2a3 and F* are rather rare in modern-day Europe, they have slightly higher frequencies in populations of the Near East, and the highest frequency of hg G2a is seen in the Caucasus today [15]. The few published ancient Y chromosome results from Central Europe come from late Neolithic sites and were exclusively hg R1a [31]. While speculative, we suggest this supports the idea that R1a may have spread with late Neolithic cultures from the east [31].

Principal Component Analysis and Multidimensional Scaling

Four Neolithic datasets were constructed (Table 2) and compared with 55 present-day European and Near Eastern populations and one Mesolithic hunter–gatherer population [20] in a PCA (Figure 2). The PCA accounted for 39% of the total genetic variation, with the first principal component (PC)

Table 2. Summary statistics, overview of population genetic analyses, and summary of haplogroup frequencies used for comparison with PCA vector loadings.

Category	Variable, Simulation, or Hg	Modern Datasets				Ancient Datasets[a]				
		Total Dataset	Pooled Geographic Sets of Equal Size ($n = \sim 500$)	Pooled European Dataset	Pooled Near East Populations	DEB22	LBK20	LBK42	LBK34	Hunter–Gatherers
Summary statistics	Populations	55	37	41	14	1	1	1	1	1
	Samples	23,394	18,039			22	20	42	34	20
Population genetic analysis & simulations	Shared haplotypes		X					X		
	PCA	X				X	X	X	X	X
	Relative hg frequencies			X	X	X	X	X	X	X
	MDS	X				X	X	X	X	
	Genetic distance maps	X				X		X		
	BayeSSC			X[b]	X[b]			X		X
	Haplotype diversity *h*					0.957	0.989	0.969	0.982	0.932
	Tajima's *D*					−0.91645	−0.90573	−0.91374	−0.88555	−1.05761
Relative hg frequencies	Asian hgs			1.62	2.09	0.00	0.00	0.00	0.00	0.00
	African hgs			0.65	6.43	0.00	0.00	0.00	0.00	0.00
	R0/preHV			0.37	3.26	0.00	0.00	0.00	0.00	0.00
	H			43.35	23.74	13.64	25.00	19.05	17.65	0.00
	HV			1.40	5.80	13.64	0.00	7.14	2.94	0.00
	J			8.49	10.59	13.64	5.00	9.52	5.88	4.76
	T			9.26	8.91	13.64	25.00	19.05	23.53	9.52
	I			2.23	1.97	0.00	0.00	0.00	0.00	0.00
	N1a			0.30	0.32	13.64	15.00	14.29	17.65	0.00
	K			5.39	6.67	13.64	15.00	14.29	14.71	4.76
	V			4.35	0.77	4.55	5.00	4.76	5.88	0.00
	W			2.03	2.25	9.09	5.00	7.14	5.88	0.00
	X			1.22	2.52	0.00	0.00	0.00	0.00	0.00
	U2			1.04	1.52	0.00	0.00	0.00	0.00	0.00
	U3			1.26	4.43	0.00	5.00	2.38	2.94	0.00
	U4			4.04	2.10	0.00	0.00	0.00	0.00	9.52
	U5a			5.46	2.53	4.55	0.00	2.38	2.94	23.80

Table 2. Cont.

Category	Variable, Simulation, or Hg	Modern Datasets				Ancient Datasets[a]				
		Total Dataset	Pooled Geographic Sets of Equal Size ($n = \sim 500$)	Pooled European Dataset	Pooled Near East Populations	DEB22	LBK20	LBK42	LBK34	Hunter–Gatherers
	U5b			3.89	0.64	0.00	0.00	0.00	0.00	28.57
	Other rare hgs			3.67	13.45	0.00	0.00	0.00	0.00	19.05

X's indicate which datasets were used in the genetic analyses.

[a]For explanation of datasets, see Materials and Methods.

[b]For BayeSSC analyses, representative samples of the key areas were randomly drawn from the larger meta-population pool (Table S6).

doi:10.1371/journal.pbio.1000536.t002

separating Near Eastern populations from Europeans (24.9%), and with LBK populations falling closer to Near Eastern ones. However, the second PC (17.4%) clearly distinguished the four Neolithic datasets from both Near East and European populations. An MDS plot (Figure S1) showed similar results, with the Near Eastern affinities of the LBK populations even more apparent.

To better understand which particular hgs made the Neolithic populations appear either Near Eastern or (West) European, we compared average hg frequencies of the total LBK (LBK42) and Derenburg (DEB22) datasets to two geographically pooled meta-population sets from Europe and the Near East (Tables 2 and S6; 41 and 14 populations, respectively). PC correlates and component loadings (Figure 2) showed a pattern similar to average hg frequencies (Table 2) in both large meta-population sets, with the LBK dataset grouping with Europeans because of a lack of mitochondrial African hgs (L and M1) and preHV, and elevated frequencies of hg V. In contrast, low frequencies of hg H and higher frequencies for HV, J, and U3 promoted Near Eastern resemblances. Removal of individuals with shared haplotypes within the Derenburg dataset (yielding dataset LBK34) did not noticeably decrease the elevated frequencies of J and especially HV in the Neolithic data.

Most importantly, PC correlates of the second component showed that elevated or high frequencies of hgs T, N1a, K, and W were unique to LBK populations, making them appear different from both Europe and Near East. The considerable within-hg diversity of all four of these hgs (especially T and N1a; Table 1) suggests that this observation is unlikely to be an artifact of random genetic drift leading to elevated frequencies in small, isolated populations.

The pooled European and Near Eastern meta-populations are necessarily overgeneralizations, and there are likely to be subsets of Near Eastern populations that are more similar to the Neolithic population. Interestingly, both the PCA and the MDS plots identified Georgians, Ossetians, and Armenians as candidate populations (Figures 2 and S1).

Mapping Genetic Distances

We generated genetic distance maps to visualize the similarity/distance of the LBK and Derenburg populations (datasets LBK42 and DEB22) to all modern populations in the large Western Eurasian dataset (Figure 3). In agreement with the PCA and MDS analyses, populations from the area bounding modern-day Turkey, Armenia, Iraq, and Iran demonstrated a clear genetic similarity with the LBK population (Figure 3A). This relationship was even stronger in a second map generated with just the Neolithic Derenburg individuals (Figure 3B). Interestingly, the map of the combined LBK data also suggested a possible geographic route for the dispersal of Neolithic lineages into Central Europe: genetic distances gradually increase from eastern Anatolia westward across the Balkans, and then northwards into Central Europe. The area with lower genetic distances follows the course of the rivers Danube and Dniester, and this natural corridor has been widely accepted as the most likely inland route towards the Carpathian basin as well as the fertile Loess plains further northwest [23,32,33].

Bayesian Serial Simcoal Analysis

While an apparent affinity of Neolithic farmers to modern-day Near East populations is revealed by the shared haplotype analyses, PCA, MDS, and genetic distance maps, the population-specific pairwise F_{ST} values among ancient populations (hunter–gatherers and LBK) and the modern population pools (Central Europe and Near East) tested were all significant ($p > 0.05$; Table 3), suggesting a degree of genetic discontinuity between ancient and modern-day populations. The early farmers were closer to the modern Near

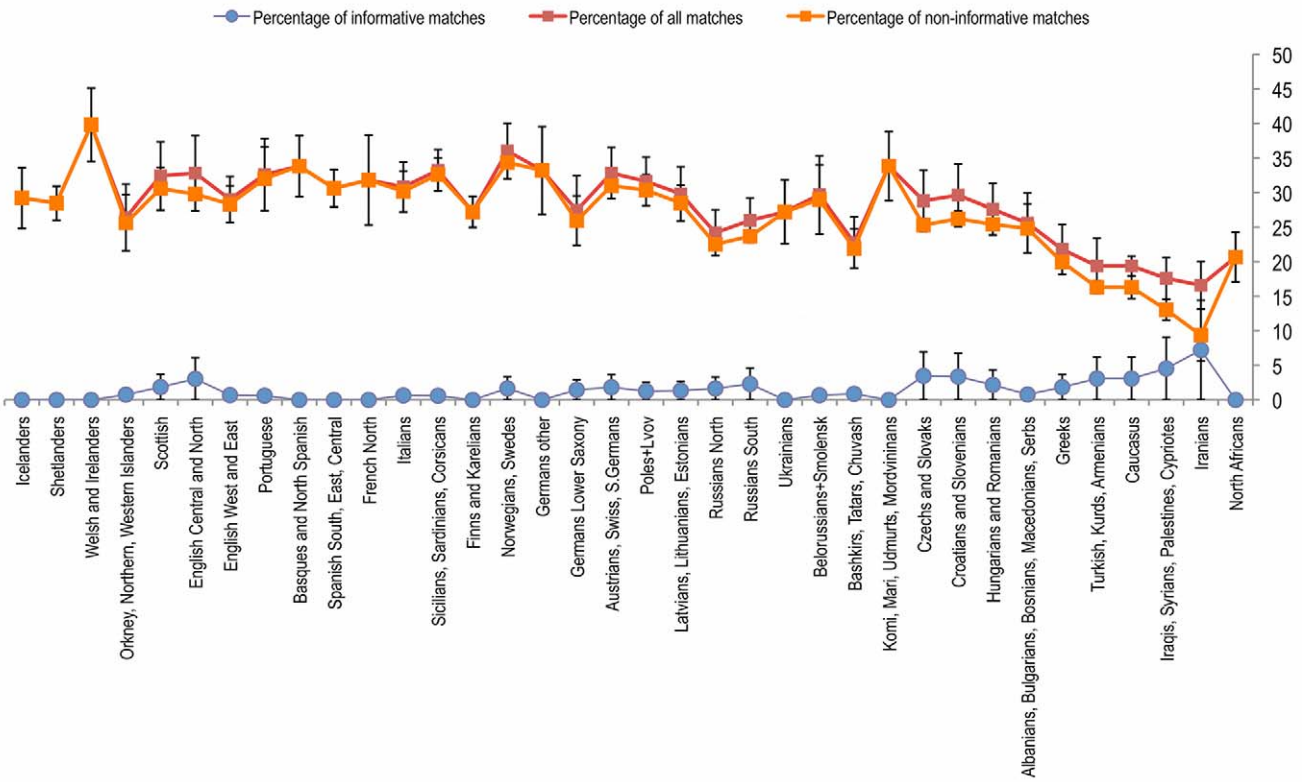

Figure 1. Percentages of shared haplotype matches per population. Populations are plotted on a northwest–southeast axis. Note that the percentage of non-informative matches (orange) is nearly identical to the percentage of all shared haplotypes (red) in most populations, whereas we observe elevated frequencies of informative matches (blue) in Southeast European and Near Eastern population pools, culminating in Iranians.
doi:10.1371/journal.pbio.1000536.g001

Eastern pool ($F_{ST} = 0.03019$) than hunter–gatherers were ($F_{ST} = 0.04192$), while both ancient populations showed similar differences to modern Central Europe, with the hunter–gatherers slightly closer ($F_{ST} = 0.03445$) than the early farmers ($F_{ST} = 0.03958$). The most striking difference was seen between Mesolithic hunter–gatherers and the LBK population itself ($F_{ST} = 0.09298$), as previously shown [20]. We used BayeSSC analyses to test whether the observed F_{ST} values can be explained by the effects of drift or migration under different demographic scenarios (Figure S2). This encompassed comparing F_{ST} values derived from coalescent simulations under a series of demographic models with the observed F_{ST} values in order to test which model was the most likely, given the data. By using an approximate Bayesian computation (ABC) framework we were able to explore priors for initial starting deme sizes and dependent growth rates to maximize the credibility of the final results. The Akaike information criterion (AIC) was used to evaluate a goodness-of-fit value of the range of models in the light of the observed F_{ST} values. In addition, a relative likelihood estimate for each of the six models given the data was calculated via Akaike weights (ω). The highest AIC values, and therefore the poorest fit, were obtained for models representing population continuity in one large Eurasian meta-population through time (Models H_0a and H_0b; Table 4). Of note, the goodness of fit was better with a more recent population expansion (modeled at the onset of the Neolithic in Central Europe) and hence higher exponential growth rate (H_0a). The model of cultural transmission (H_1), in which a Central European deme including Neolithic farmers and hunter–gatherers coalesced with a Near Eastern deme in the Early Upper Paleolithic (1,500 generations, or ~37,500 y ago), resulted in intermediate goodness-of-fit values (H_1a and H_1b; Table 4; Figure S2). The best goodness-of-fit values were retrieved for models of demic diffusion (model H_2; Table 4) with differing proportions of migrants (25%, 50%, and 75% were tested) from the Near Eastern deme into the Central European deme around the time of the LBK (290 generations, ~7,250 y ago; Table 4). Notably, the models testing 50% and 75% migrants returned the highest relative likelihood values (42% and 52%, respectively), and therefore warrant further investigation. However, while the demic diffusion model H_2 produced values that approximated the observed F_{ST} between Neolithic farmers and the Near Eastern population pool, none of the models could account for the high F_{ST} between hunter–gatherers and early farmers or early farmers and modern-day Central Europeans.

The models we tested represent major oversimplifications and it should be noted that modeling human demographic history is notoriously difficult, especially given the complex history of Europe and the Near East over this time scale. The fact that no model explained the observed F_{ST} between ancient and modern-day populations particularly well suggests that the correct scenario has not yet been identified, and that there is also an obvious need for sampling of material from younger epochs. Additionally, sampling bias remains an issue in aDNA studies, and this is particularly true for the chronologically and geographically diverse hunter–gatherer dataset. In the light of the models tested (see also [19,20]), we would suggest that the basis of modern European mtDNA diversity was formed from the postglacial re-peopling of Europe (represented here by the Mesolithic hunter–gatherers) and the genetic input from the Near East during the Neolithic, but that demographic processes after the early Neolithic have contributed substantially to shaping Europe's contemporary genetic make up.

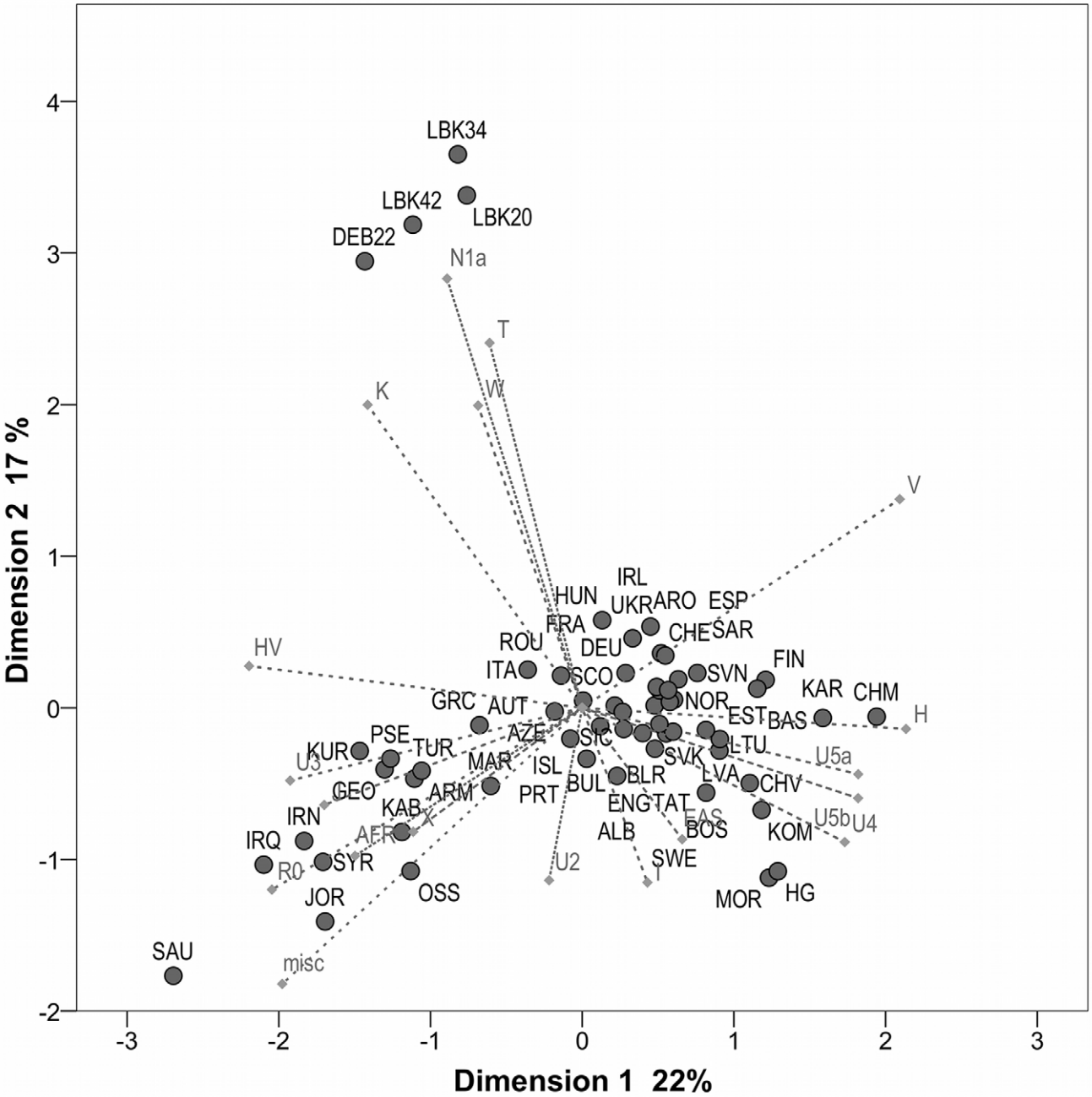

Figure 2. PCA plot based on mtDNA haplogroup frequencies. The two dimensions display 39% of the total variance. The contribution of each hg is superimposed as grey component loading vectors. Notably, the Derenburg dataset (DEB22) groups well with its meta-population (LBK20), supporting the unique status and characteristic composition of the LBK sample. Populations are abbreviated as follows (Table S6): ALB, Albanians; ARM, Armenians; ARO, Aromuns; AUT, Austrians; AZE, Azeris; BAS, Basques; BLR, Byelorussians; BOS, Bosnians; BUL, Bulgarians; CHE, Swiss; CHM, Mari; CHV, Chuvash; CRO, Croats; CZE, Czechs; DEB22, Derenburg; DEU, Germans; ENG, English; ESP, Spanish; EST, Estonians; FIN, Finns; FRA, French; GEO, Georgians; GRC, Greeks; HG, European Mesolithic hunter–gatherers.; HUN, Hungarians; IRL, Irish; IRN, Iranians; IRQ, Iraqis; ISL, Icelanders; ITA, Italians; JOR, Jordanians; KAB, Kabardinians; KAR, Karelians; KOM, Komis (Permyaks and Zyrian); KUR, Kurds; LBK20, LBK without Derenburg; LBK34, all LBK samples excluding potential relatives; LBK42, all LBK; LTU, Lithuanians; LVA, Latvians; MAR, Moroccans; MOR, Mordvinians; NOG, Nogais; NOR, Norwegians; OSS, Ossetians; POL, Poles; PRT, Portuguese; PSE, Palestinians; ROU, Romanians; RUS, Russians; SAR, Sardinians; SAU, Saudi Arabians; SCO, Scots; SIC, Sicilians; SVK, Slovaks; SVN, Slovenians; SWE, Swedes; SYR, Syrians; TAT, Tatars; TUR, Turkish; UKR, Ukrainians.
doi:10.1371/journal.pbio.1000536.g002

Synthesis of Population Genetic Analyses

The aDNA data from a range of Mesolithic hunter–gatherer samples from regions neighboring the LBK area have been shown to be surprisingly homogenous across space and time, with an mtDNA composition almost exclusively of hg U (~80%), particularly hg U4 and U5, which is clearly different from the LBK dataset as well as the modern European diversity (Table 2) [20]. The observation that hgs U4 and U5 are virtually absent in the LBK population (1/42 samples) is striking (Table 2). Given this clear difference in the mtDNA hg composition, it is not surprising

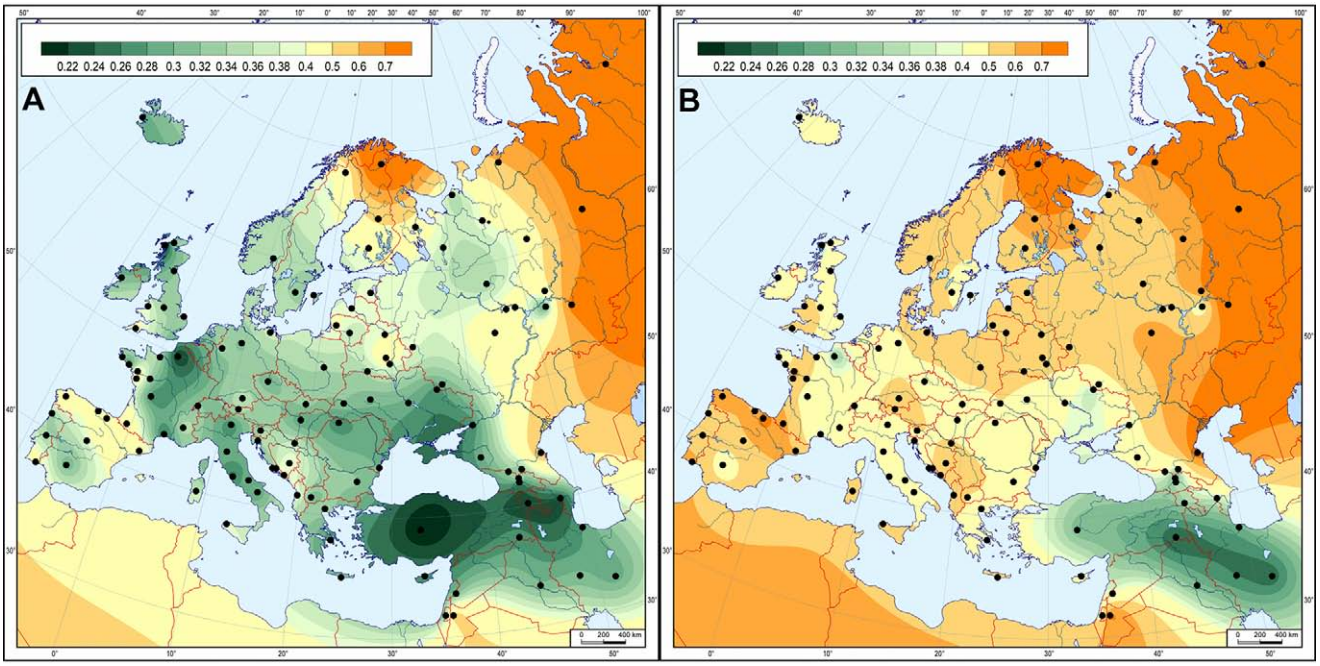

Figure 3. Genetic matrilineal distances between 55 modern Western Eurasian populations (Table S6) and Neolithic LBK samples. Mapped genetic distances are illustrated between 55 modern Western Eurasian populations and the total of 42 Neolithic LBK samples (A) or the single graveyard of Derenburg (B). Black dots denote the location of modern-day populations used in the analysis. The coloring indicates the degree of similarity of the modern local population(s) with the Neolithic sample set: short distances (greatest similarity) are marked by dark green and long distances (greatest dissimilarity) by orange, with fainter colors in between the extremes. Note that green intervals are scaled by genetic distance values of 0.02, with increasingly larger intervals towards the "orange" end of the scale.
doi:10.1371/journal.pbio.1000536.g003

that the pairwise F_{ST} between hunter–gatherers and the LBK population is the highest observed (0.09298) when we compared ancient populations with representative population pools from Central Europe and the Near East (Table 3; see also [20]). If the Mesolithic data are a genuine proxy for populations in Central Europe at the onset of the LBK, it implies that the Mesolithic and LBK groups had clearly different origins, with the former potentially representing the pre-Neolithic indigenous groups who survived the Last Glacial Maximum in southern European refugia. In contrast, our population genetic analyses confirm that the LBK shares an affinity with modern-day Near East and Anatolia populations. Furthermore, the large number of basal lineages within the LBK, a reasonably high hg and haplotype diversity generated through one- or two-step derivative lineages, and the negative Tajima's *D* values (Tables 1 and 2) indicate a recent expansion. These combined data are compatible with a model of Central Europe in the early Neolithic of indigenous populations plus significant inputs from expanding populations in the Near East [4,12,34]. Overall, the mtDNA hg composition of the LBK would suggest that the input of Neolithic farming cultures (LBK) to modern European genetic variation was much higher than that of Mesolithic populations, although it is important to note that the unique characteristics of the LBK sample imply that further significant genetic changes took place in Europe after the early Neolithic.

Table 3. Pairwise F_{ST} values between ancient and modern-day population pools as used for goodness-of-fit estimates in BayeSSC analyses.

	Hunter–Gatherers	Near East	LBK	Central Europe
Hunter–Gatherers	0	—	—	—
Near East	0.04192	0	—	—
LBK	0.09298	0.03019	0	—
Central Europe	0.03445	0.00939	0.03958	0

doi:10.1371/journal.pbio.1000536.t003

aDNA data offers a powerful new means to test evolutionary models and assumptions. The European lineage with the oldest coalescent age, U5, has indeed been found to prevail in the indigenous hunter–gatherers [12,35]. However, mtDNA hgs J2a1a and T1, which because of their younger coalescence ages have been suggested to be Neolithic immigrant lineages [8,12], are so far absent from the samples of early farmers in Central Europe. Similarly, older coalescence ages were used to support hgs K, T2, H, and V as "postglacial/Mesolithic lineages," and yet these have been revealed to be common only in Neolithic samples. The recent use of whole mitochondrial genomes and the refinement of mutation rate estimates have resulted in a general reduction in coalescence ages [8], which would lead to an improved fit with the aDNA data. However we advise caution in directly relating coalescence ages of specific hgs to evolutionary or prehistoric demographic events [36]. Significant temporal offsets can be caused by either observational bias (the delay between the actual split of a lineage and the eventual fixation and dissemination of this lineage) or calculation bias (incorrect coalescent age estimation). aDNA has considerable value not only for directly analyzing the presence or absence of lineages at points in the past but also for refining mutation rate estimates by providing internal calibration points [37].

Table 4. Details of the demographic models analyzed with BayeSSC and AIC goodness-of-fit estimates, and resulting model probabilities via Akaike weights.

Model	H_0a	H_0b	H_1	H_2	H_2	H_2
Prior N_e, time 0 ,deme 0	U[a]:100000,30000000	U:100000,30000000	U:100000,12000000	U:100000,12000000	U:100000,12000000	U:100000,12000000
Prior N_e, time 0, deme 1			U:100000,12000000	U:100000,12000000	U:100000,12000000	U:100000,12000000
Percent migrants from deme 0 to deme 1				25%	50%	75%
AIC	97.78	120.37	89.19	82.56	78.52	78.07
Akaike weight ω	$2.76164e^{-5}$	$3.42478e^{-10}$	0.002018032	0.055596369	0.418527622	0.52383036

Of note, the smaller the AIC value, the better the fit of the model. While no threshold value can be assigned to AIC values at which any model can be rejected, the Akaike weights estimate a model probability given the six models tested.
[a]U, uniform distribution of given range.
N_e, effective population size.
doi:10.1371/journal.pbio.1000536.t004

Archaeological and anthropological research has produced a variety of models for the dispersal of the Neolithic agricultural system ("process of Neolithization") into and throughout Europe (e.g., [1,2,38]). Our findings are consistent with models that argue that the cultural connection of the LBK to its proposed origin in modern-day Hungary, and reaching beyond the Carpathian Basin [23,32,38,39], should also be reflected in a genetic relationship (e.g., shared haplotype analyses; Table S4). Therefore at a large scale, a *demic diffusion* model of genetic input from the Near East into Central Europe is the best match for our observations. It is notable that recent anthropological research has come to similar conclusions [40,41]. On a regional scale, "leap-frog" or "individual pioneer" colonization models, where early farmers initially target the economically favorable Loess plains in Central Europe [33,42], would explain both the relative speed of the LBK expansion and the clear genetic Near Eastern connections still seen in these pioneer settlements, although the resolving power of the genetic data is currently unable to test the subtleties of these models.

In conclusion, the new LBK dataset provides the most detailed and direct genetic portrait of the Neolithic transition in Central Europe; analysis of this dataset reveals a clear demonstration of Near Eastern and Anatolian affinities and argues for a much higher genetic input from these regions, while also identifying characteristic differences from all extant (meta-)populations studied. Ancient genetic data from adjacent geographic regions and time periods, and especially from the Near East and Anatolia, will be needed to more accurately describe the changing genetic landscape during and after the Neolithic, and the new multiplexed SBE assays offer a powerful means to access this information.

Materials and Methods

Archaeological Background

The archaeological site Derenburg Meerenstieg II (Harzkreis, Saxony-Anhalt, Germany) was excavated during three campaigns in 1997–1999 comprising an area of 3 ha. The archaeological context at this site shows a record of settlement activity ranging from the Early Neolithic (LBK) and Middle Neolithic (Rössen and Ammensleben cultures) to Bronze and Iron Age [43]. However, the main features of Derenburg are the LBK graveyard and its associated partial settlement approximately 70 m southwest. The archaeological data revealed that the larger part of the settlement has not yet been excavated and lies outside the area covered during these campaigns. In contrast, the graveyard was recorded in its entire dimension (25×30 m) and encompassed a total of 41 graves. Two separate graves were found outside the graveyard (50 m WSW and 95 m SSE). Erosion and modern agricultural ploughing might have led to a loss of some graves at the plateau area. Here, the graves were shallow and in average state of preservation, whereas the graves embedded in deeper Loess layers showed an excellent state of preservation. In total, 32 single grave burials were found; there were also one double burial, one triple burial, two burials in settlement pits, two or three times additional singular bones in a grave, three burials with a secondary inhumation, and one empty grave. The majority of individuals (75%) at Derenburg were buried in East–West orientation in a varying flexed position. The duration of usage of the graveyard spans over the entire time frame of the LBK and is reflected by the typology of the ceramics and associated grave goods ranging from older LBK pottery (Flomborn style) to youngest LBK pottery. Absolute radiocarbon dates confirm the usage over three centuries (5,200–4,900 cal B.C.; see also Table 1 and [44]).

Ancient DNA Work

From an initial 43 graves in the Derenburg graveyard, 31 indicated morphological preservation suitable for sampling and aDNA analyses. Five individuals had already been sampled in 2003 for our previous study and showed excellent preservation of aDNA, a negligible level of contamination, and an unusual mtDNA hg distribution, thereby justifying further investigation [19]. Hence, 26 additional individuals were processed in this study (Table 1). We amplified, cloned, and sequenced mitochondrial HVS-I (nucleotide positions [np] 15997–16409; nucleotide position according to [45]) as described previously [19]. mtDNA hg assignments were further supported by typing with a newly developed multiplex of 22 mtDNA coding region SNPs (GenoCoRe22). In addition, we typed 25 Y chromosome SNPs using a second novel multiplex assay (GenoY25). Final refinement of Y chromosome hg assignments was performed via singleplex PCRs. Lastly, the amount of starting DNA template molecules was monitored using qPCR on seven random samples (Table S3). aDNA work was performed in specialized aDNA facilities at the Johannes Gutenberg University of Mainz and the Australian Centre for Ancient DNA (ACAD) at the University of Adelaide according to appropriate criteria. All DNA extractions as well as amplification, cloning, and sequencing of the mitochondrial control region HVS-I were carried out in the Johannes Gutenberg University of Mainz facilities. Additional singleplex, all multiplex, and quantitative real-time amplifications, SNP typing, and direct sequencing of Y chromosome SNPs were carried at the ACAD as described below.

SNP Selection and Multiplex Design

The technique of SNP typing via SBE reactions (also known as minisequencing) has proven a reliable and robust method for high throughput analyses of polymorphisms, e.g., human mitochondrial variation [46], human X- and Y-chromosomal SNPs [47,48], and human autosomal SNPs [49]. However, few SBE studies have addressed the special need for very short amplicon sizes to allow amplification from highly degraded DNA, as even forensic protocols have generally targeted relatively long amplicon sizes [50–54]. Our first multiplex (GenoCoRe22) was designed to type a panel of 22 mitochondrial coding region SNPs that are routinely typed within the Genographic Project [25], to allow for future maximum comparability with modern population data. A second multiplex (GenoY25) targeted a basal, but global, coverage of 25 commonly typed Y chromosome SNPs, for maximum comparability of paternal lineages. The aim of the SNP assay design was to produce highly efficient and sensitive protocols, capable of working on highly degraded DNA, that also allow modern human DNA contamination to be detected at very low levels and monitored [51]. The GenoCoRe22 SNP panel was chosen to cover the basal branches of mitochondrial hgs across modern human mtDNA diversity [25]. The chosen SNP sites were identical to the initial set (Figure 4 in [25]) except for hg W (SNP at np 8994 instead of np 1243) and hg R9 (SNP at np 13928 instead of np 3970), as a compromise arising from primer design within a multiplex assay. Selection of GenoY25 SNP panel for incorporation into the multiplex assay was performed using the highly resolved Y Chromosome Consortium tree and an extensive literature search for corresponding SNP allele frequencies in European populations [13,26,55].

Multiplex PCR Assays GenoCoRe22 and GenoY25

Multiplex assays were set up, established, and performed at the ACAD facilities. Multiplex PCR using Amplitaq Gold (Applied Biosystems) was conducted in 25-µl volumes using 1× Buffer Gold, 6 mM (GenoCoRe22) or 8 mM (GenoY25) $MgCl_2$, 0.5 mM dNTPs (Invitrogen), ≤0.2 µM of each primer, 1 mg/ml RSA (Sigma), 2 U of Amplitaq Gold Polymerase, and 2 µl of DNA extract. Thermocycling conditions consisted of an initial enzyme activation at 95°C for 6 min, followed by 40–45 cycles of denaturation at 95°C for 30 s, annealing at 60°C (GenoCoRe22) or 59°C (GenoY25) for 30 s, and elongation at 65°C for 30 s, with a single final extension time at 65°C for 10 min. Each PCR included extraction blanks as well as a minimum of two PCR negatives at a ratio of 5:1. PCRs were visually checked by electrophoresis on 3.5% agarose TBE gels. PCR products were purified by mixing 5 µl of PCR product with 1 U of SAP and 0.8 U of ExoI and incubating at 37°C for 40 min, followed by heat inactivation at 80°C for 10 min. Because of the sensitivity of the multiplex PCR (using fragment lengths of only 60–85 bp), and to be able to monitor potential human background contamination, usually all controls were included in downstream fragment analysis. Multiplex primer sequences and concentration are given in Table S7.

SNaPshot Typing

SBE reactions were carried out on the GenoCoRe22 and GenoY25 SNP multiplex assay using the ABI Prism SNaPshot multiplex reaction kit (Applied Biosystems) following the manufacturer's instructions, except that 10% 3 M ammonium sulfate was added to the extension primer mix to minimize artifacts [56]. SBE primers and concentrations are given in Table S7. Cycling conditions consisted of 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 reactions were purified using 1 U of SAP, incubating at 37°C for 40 min, followed by heat inactivation at 80°C for 10 min. Prior to capillary electrophoresis, 2 µl of purified SNaPshot product was added to a mix of 11.5 µl of Hi-Di Formamide (Applied Biosystems) and 0.5 µl of Gene-Scan-120 LIZ size standard (Applied Biosystems). Samples were run on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems) after a denaturation carried out according to the manufacturer's instructions using POP-6 (Applied Biosystems). Evaluation and analyses of SNaPshot typing profiles were performed using custom settings within the GeneMapper version 3.2 Software (Applied Biosystems).

Y Chromosome SNP Singleplex PCRs and Sequencing

Additional Y chromosome SNPs (M285, P287 S126, and M69) were tested to determine specific downstream subclades based on the initial multiplex results in order to gain further resolution. We chose appropriate SNP loci by following general criteria, trying to keep the PCR amplicon size smaller than 90 bp in size and flanking DNA sequences free from interfering polymorphisms, such as nucleotide substitutions in potential primer binding sites. We selected PCR amplification primers that have a theoretical melting temperature of around 60°C in neutral buffered solutions (pH 7–8), with monovalent cation (Na^+) concentrations at 50 mM and divalent cation (Mg^{++}) concentrations at 8 mM. All primer candidates were analyzed for primer–dimer formation, hairpin structures, and complementarities to other primers in the multiplex using Primer 3 (http://primer3.sourceforge.net/). Primer characteristics were chosen to ensure equal PCR amplification efficiency for all DNA fragments, as previously described [50]. The primers were HPLC-purified and checked for homogeneity by MALDI-TOF (Thermo). Table S7 shows the sequences and the concentrations of the amplification primers in the final multiplex PCR.

Additional Y chromosome SNP singleplex PCRs were carried out in the ACAD facilities. Standard PCRs using Amplitaq Gold (Applied Biosystems) were conducted in 25-µl volumes using 1× Buffer Gold, 2.5 mM $MgCl_2$, 0.25 mM of each dNTP (Fermentas), 400 µM of each primer (Table S7), 1 mg/ml RSA (Sigma-Aldrich), 2 U of Amplitaq Gold Polymerase, and 2 µl of DNA extract. Thermocycling conditions consisted of an initial enzyme activation at 95°C for 6 min, followed by 50 cycles of denaturation at 94°C for 30 s, annealing at 59°C for 30 s, and elongation at 72°C for 30 s, with a single final extension time at 60°C for 10 min. Each PCR reaction included extraction blanks as well as a minimum of two PCR negatives. PCR products were visualized and purified as described above and were directly sequenced in both directions using the Big Dye Terminator 3.1 Kit (Applied Biosystems) as per manufacturer's instructions. Sequencing products were purified using Cleanseq magnetic beads (Agencourt, Beckman Coulter) according to the manufacturer's protocol. Sequencing products were separated on a 3130xl Genetic Analyzer (Applied Biosystems), and the resulting sequences were edited and aligned relative to the SNP reference sequence (GenBank SNP accession numbers: M285, rs13447378; P287, rs4116820; S126 [also known as L30], rs34134567; and M69, rs2032673) using the software Sequencher 4.7 (Genecodes).

Quantitative Real-Time PCR

qPCR was used to determine the amount of DNA in the samples prior to amplification and to assess the authenticity based on the assumption that there is an inverse relationship between DNA quantity and fragment length for degraded aDNA [57,58]. Two different length fragments were amplified from the HVS-I: 141 bp (L16117/H16218) and 179 bp (L16209/H16348) [19,59].

All qPCR reactions were carried out in a 10-µl reaction volume containing 1× Express SYBR Green ER Supermix Universal (Invitrogen), rabbit serum albumin (10 mg/ml), forward and reverse primers (10 µM), and 1 µl of DNA extract. Thermocycling conditions consisted of an initial enzyme activation at 95°C for 5 min, followed by 50 cycles of 94°C for 10 s, 58°C for 20 s, and 72°C for 15 s. The primer specificity was assessed using a post-PCR melt curve to visualize the dissociation kinetics. The primers were validated using modern DNA, and a single peak was observed for both fragments, indicating specific binding. The dissociation temperature (T_M) was 80–80.3°C for the 141-bp fragment and 81.7–82.3°C for the 179-bp fragment. Both primer pairs showed an absence of primer dimers, indicated by the lack of a smaller peak on the melt curve (≈60°C) and a single band on a 2% agarose gel. The starting quantity of DNA in the ancient samples was determined by comparison to a standard curve of a known amount of DNA. The standard curves for the two fragments were created from modern human DNA. The DNA was extracted from a buccal cheek swab of a single individual using DNeasy Blood and Tissue Kit (Qiagen). mtDNA was amplified for the two fragments (141 bp and 179 bp) using 1× Hotmaster Buffer (Eppendorf), 0.5 U of Hotmaster Taq (5Prime), forward and reverse primers (10 µM), distilled water, and 2 µl of DNA extract. Thermocycling conditions consisted of an initial enzyme activation at 94°C for 2 min, followed by 30 cycles of 94°C for 20 s, 60°C for 10 s, and 65°C for 1 min. The PCR products were purified using Agencourt Ampure (Beckman Coulter) according to manufacturer's instructions. The DNA concentration for the 141-bp and 179-bp amplicons was measured twice at 1:1 and 1:10 dilutions with a Nanovue (GE Healthcare). Ten-fold serial dilutions, from 1×10^6 to 10 copies/µl, of the purified fragments were used to make the standards. These were run with the qPCR conditions described above. For each standard, each 10-fold dilution was run in triplicate and the qPCR was repeated on a separate day. All the standards met the following criteria: (1) there was a linear regression relationship between DNA quantity and cycle threshold (fluorescence above background), $R^2>0.95$, and (2) the reaction was efficient (i.e., a doubling of product per cycle in the exponential phase), between 90% and 110%. Ancient qPCRs were run in triplicate with extraction and PCR blanks, and PCR standards (positive control) run in duplicate. Amplifications were performed on Rotor-Gene 6000 and analyses on Rotor-Gene 6000 Series Software 1.7 (Corbett). The difference in mtDNA quantity between fragment lengths (141 and 179 bp) was assessed using a nonparametric version of a Student's *t* test, a Wilcoxon signed-ranks test. This test was selected because the data were not appropriate for a parametric test, displaying a mixture of normal (179 bp, $p=0.425$) and non-normal (141 bp, $p=0.012$) distributions, as determined from a Shapiro-Wilk *W* test, which is appropriate for testing the normality of groups with small sample sizes.

Authentication Criteria

In line with previous publications on aDNA and especially with criteria for working with human aDNA, it can be stated that a 100% authentication of ancient samples is virtually impossible [22,57,60]. However, we took all possible precautions to prevent modern contaminations, and we regard the results as authentically derived from endogenous DNA based on the following chain of evidence. (1) All samples were collected under DNA-free conditions after excavation. Samples were not washed, treated, or examined before taking DNA samples. (2) All preparation and analytical steps prior to DNA amplification were conducted in a clean room area solely dedicated to aDNA work located in a physically separated building without any modern DNA work (pre-PCR area). Amplification, cloning, and sequencing were carried out in the post-PCR lab. (3) All steps were monitored by non-template controls and by using bovid samples in parallel. (4) All individuals were sampled twice from anatomically independent regions and treated independently. At least eight independent PCR reactions were carried out (four overlapping fragments × two extractions) per individual. In case of successful amplification of all eight fragments, these were cloned and an average of eight clones per amplicons was sequenced to detect heterogeneous sequences due to DNA degradation or contamination. All replicable polymorphic sites were consistent with existing mtDNA haplotypes, ruling out post mortem DNA damage as a potential source for erroneous sequences. (5) The new multiplexes not only clearly confirm hg assignment but also provide an ideal monitoring system for ancient human DNA samples, as they directly target SNPs defining all potential contaminating lineages. (6) qPCR was carried out on a selection of samples to ensure appropriate levels of DNA quantity and to assess DNA quality. (7) Samples were collected and processed by W. H. exclusively (mtDNA hg H1, np 15997–16409: 16189C 16311C, and Y chromosome hg E1b1b1a-M78) after excavation; no other staff were involved in any of the pre-PCR steps. Eventually, all listed criteria indicating authenticity or at least the plausibility of having retrieved endogenous DNA were evaluated, together with the sample's post-excavation history [60].

Populations under Study

Four partly overlapping Neolithic datasets were analyzed: the 22 Derenburg individuals (DEB22); 20 individuals from other LBK populations previously published (LBK20; Table S5 and [19]); the combined LBK dataset (LBK42); and the combined LBK dataset excluding eight individuals of possible kinship (LBK34, see below) to avoid overestimation of haplotype frequencies. These four Neolithic sets were analyzed against extant population data from the MURKA mitochondrial DNA database and integrated software, currently containing 97,523 HVS-I records from published sources, and maintained by coauthors V. Z., E. B., and O. B. of the Russian Academy of Medical Sciences. Analyses were restricted to 390 populations from Europe and the Near East (35,757 mtDNAs). For detailed analysis of shared haplotypes, we included only sequences spanning from np 16069 to np 16365 (34,258 samples, haplotype dataset). aDNA sequences were trimmed to the same length. For frequency-based analyses (PCA, MDS, and genetic distance maps), we omitted mtDNAs whose hg affiliations were ambiguous (absence of information on coding region SNPs), resulting in our final hg frequency dataset of 23,394 individuals from 228 population studies, which subsequently were pooled into 55 populations based on ethnicity, language, and/or geographical criteria as described in the original publications (see Table S6).

Addressing Potential Kinship within the Derenburg Graveyard

The mtDNA and Y chromosome hg results were overlaid onto the map of the graveyard to elucidate the spatial relationships within the graveyard (Figure S3). Four haplotypes were shared by two individuals each, and two haplotypes by three individuals each, while the remaining eight individuals (36.4%) showed unique haplotypes within the Derenburg graveyard. A number of shared haplotypes is not surprising in a medium sized, closed LBK graveyard where the influence of genetic drift and a certain level of biological kinship are likely. However, little positional structuring according to maternal lineages was observed. A clustering of

mtDNA haplotypes H-rCRS (deb9 and deb21) and HV (deb4, deb20, and deb5) in the northwest corner of the cemetery is notable, whereas other shared haplotype "twins" or "trios" with a potential maternal relationship are spread across larger distances. However, it must be stated that there are many other factors influencing the layout of interments in a graveyard that cannot be unraveled by aDNA analyses. LBK burials commonly show a great variety of mortuary patterns or rites at the same site (e.g., burials within the settlement and burials in pits/middens), and it is therefore not clear whether individuals in the cemetery represent the norm or the exception, and how much of the initial genetic variation of the population is missing [44]. In any case, to avoid overestimation of haplotype frequencies in the LBK dataset, the eight duplicate haplotypes were excluded, and a reduced dataset (LBK34) was used in population genetic analyses alongside the complete set to account for a potential kinship effect.

Haplotype Diversity and Tajima's *D*

Haplotype diversity (*h*) and Tajima's *D* were calculated using DnaSP version 5 [61].

Shared Haplotype Analysis

In order to calculate the percentage of shared haplotypes between the LBK sample and modern-day populations, we chose modern populations of equal or larger sample sizes, resulting in 36 out of 55 pooled populations with sample size $n = 500$ or above. Pooling was based on geographic proximity and linguistic similarity. For population studies with $n > 500$, 500 samples were selected randomly. After pooling and random selection the dataset comprised 18,039 samples. A pivot table was created (4,140 haplotypes in rows and 36 populations in columns), and Neolithic LBK data were included. Similarity between LBK and other populations was described quantitatively in two ways: (1) indicating presence or absence (1/0), i.e., whether or not the particular Neolithic haplotype was found in a given modern population, and (2) indicating the number of hits, i.e., how many times the particular haplotype was found in a given population. The 25 different LBK sequence haplotypes were sorted into clusters of noninformative (11), informative (10), and unique (4) haplotypes (Table S4). We then calculated the relative frequency of each of the shared informative vs. noninformative LBK sequence haplotypes in each of the 36 modern-day populations (Table S4). A two-tailed *z* test (Excel version 12.1, Microsoft Office) was applied to determine which population pool showed a significantly higher or lower percentage of shared informative haplotypes (Table S4). Nonparametric bootstrapping of 100 replicates for each hg per population was used to generate the confidence intervals for the percentage of hgs that are shared between all matches, informative matches, and noninformative matches. Bootstrapping was performed in Excel.version 12.1.

Principal Component and Multidimensional Scaling Analyses

Classical and categorical PCAs and MDS were performed using the hg frequencies dataset. To avoid overpopulating graphs with 228 populations, populations were pooled into 55 groups defined by ethnicity, language, and/or geography as described in the original publications (see Table S6). To minimize statistical noise caused by very rare hgs, we considered only the following 19 hgs with average frequency above 1% in Europe and Near East: preHV, H, HV, J, T, I, N1a, K, V, W, X, U2, U3, U4, U5a, U5b, the group of African hgs (L and M1), the group of East Eurasian hgs (A, B, C, D, F, G, and Z), and the group of all other (rare) hgs. PCAs and categorical PCAs (used for the biplot graph in Figure 1, with default settings to correspond to a classical PCA) were performed and visualized using the software package SPSS Statistics 17.0. Nei's genetic distances [62] were calculated using the software program DJ, written by Yuri Seryogin (freely available at http://www.genofond.ru). The resulting distance matrix was visualized via MDS in SPSS Statistics 17.0.

Mapping Genetic Distances

The genetic distances from two Neolithic datasets (DEB22 and LBK42) to populations in the hg frequencies dataset (pooled into 120 populations with the average sample size $n = 196$ to gain a balanced geographical coverage) were calculated using the software DJ. Distances were plotted on a geographic map of Europe using the software GeneGeo written by S. K. This software is the renewed GGMAG package previously used for gene geographical studies ([63] and references therein).

Bayesian Serial Simcoal Analysis

We calculated population-specific pairwise genetic distances (F_{ST}) in Arlequin version 3.5 [64], using 377-bp HVS-I sequences (np 16069–16365) assigned to one of four populations (Table S6): modern Central Europeans from the LBK core area ($n = 1{,}030$), modern Near Easterners ($n = 737$), LBK samples ($n = 42$), and hunter–gatherers ($n = 20$). F_{ST} values were estimated using the Kimura two-parameter model [65] using a gamma distribution with shape parameter of 0.205 [66].

To test whether drift can account for the high F_{ST} values between ancient and contemporary populations from Central Europe and the Near East we modeled three alternative population histories (Figure S2) using simulated coalescent analyses in the program BayeSSC [67,68].

Under the null hypothesis (H_0) we considered one large continuous Eurasian population with an effective population size ranging from 100,000 to 30 million and an exponential growth starting from a small Palaeolithic deme of 5,000 females, 300 (H_0a) or 1,500 (H_0b) generations ago. Hypothesis 1 (H_1) assumes two exponentially growing populations, a Central European deme (100,000 to 12 million) and a Near Eastern deme (100,000 to 12 million), which coalesce 1,500 generations ago (37,500 y ago, assuming 25 y per generation) in an Early Upper Palaeolithic deme of 5,000 females and constant size. Here, ancient samples from hunter–gatherers and Neolithic farmers were included in the Central European deme; therefore, this model can be considered a test for genetic continuity of Central European lineages under a scenario of cultural diffusion/transmission. Alternatively, we modeled a contrasting ("demic diffusion") scenario (H_2), similar to H_1 in structure but allowing for migration from the Near Eastern deme 290 generations ago. We tested a contribution of 25%, 50%, and 75% migrants from the Near Eastern to the Central European deme.

Each model was simulated initially using BayeSSC for 100,000 genealogies and a fixed mutation rate of 7.5×10^{-6} per site per generation [66]. A uniform distribution was used for priors to estimate effective population sizes at time 0 for the Central European and Near Eastern demes (Table 4). To compare the simulated and observed data, five pairwise F_{ST} values were chosen that reflect population differentiation between each of the two ancient samples and modern populations (Table 3). The simulated and observed F_{ST} values were compared within an ABC framework [69], in which the top 1% of simulations were retained. Posterior distributions for each of the parameters with a prior were assessed. ABC was performed in R version 2.11.0 using scripts freely available at http://www.stanford.edu/group/hadlylab/ssc/index.html.

To compare the goodness of fit of each model using AIC [70] given the observed data, priors were removed from the model and replaced with absolute parameter values that gave the maximum likelihood. The model was rerun in BayeSSC for 1,000 genealogies. The AIC for each model was calculated in R, and Akaike weights ω to compare the relative likelihood of each model where calculated in Excel version 12.1 [71,72].

Supporting Information

Dataset S1 Sequence alignments of the Derenburg individuals.
Found at: doi:10.1371/journal.pbio.1000536.s001 (17.75 MB PDF)

Figure S1 Multidimensional scaling plot of genetic distances based on haplogroup frequencies (alienation = 0, 1117760; stress = 0, 1053030). Population abbreviations are consistent with Figure 1, and further population details and references are listed in Table S6.
Found at: doi:10.1371/journal.pbio.1000536.s002 (1.05 MB TIF)

Figure S2 Demographic models and population pairwise F_{ST} values used in BayeSSC analyses. CE_1, Central European deme 1; exp, exponential; HG, hunter–gatherers; M, migrants; Ne, effective population size; NE_0, Near Eastern deme 0; r, growth rate; UP, Upper Paleolithic.
Found at: doi:10.1371/journal.pbio.1000536.s003 (3.00 MB TIF)

Figure S3 Map of the Neolithic graveyard Derenburg Meerenstieg II.
Found at: doi:10.1371/journal.pbio.1000536.s004 (1.29 MB TIF)

Table S1 Results of mtDNA coding region SNP typing using the GenoCoRe22 assay. SNPs are detected in forward orientation (L-strand) unless stated otherwise (underlined), and SNP results are reported as typed in the SBE assay. Italicized samples were discarded from further analyses. Samples were typed twice from two independent extracts except for individuals deb1 and deb2. Empty cells indicate either allelic dropout or a relative fluorescence unit value below the threshold of 50. SNP 3594_L3'4 consistently yielded relative fluorescence unit values below 50, and was not reported. Subsequent primer mixes were adjusted for the suboptimal performance of SNP3594 (Table S7).
Found at: doi:10.1371/journal.pbio.1000536.s005 (0.26 MB DOC)

Table S2 Results of Y chromosome SNP typing using the GenoY25 assay. SNPs are detected in forward orientation unless stated otherwise (underlined), and SNP results are reported as typed in the SBE assay.
Found at: doi:10.1371/journal.pbio.1000536.s006 (0.21 MB DOC)

Table S3 Quantitative real-time PCR of Neolithic Samples from Derenburg.
Found at: doi:10.1371/journal.pbio.1000536.s007 (0.02 MB XLS)

Table S4 Shared haplotype analyses.
Found at: doi:10.1371/journal.pbio.1000536.s008 (0.08 MB XLS)

Table S5 Ancient samples from other LBK sites used for population genetics analyses [19].
Found at: doi:10.1371/journal.pbio.1000536.s009 (0.07 MB PDF)

Table S6 Details of Neolithic and modern-day populations used for comparison.
Found at: doi:10.1371/journal.pbio.1000536.s010 (0.14 MB XLS)

Table S7 GenoCoRe22 and GenoY25 multiplex assay and additional Y chromosome PCR primer information.
Found at: doi:10.1371/journal.pbio.1000536.s011 (0.24 MB XLS)

Acknowledgments

We thank Sabine Moeller-Rieker for help with cloning; Susanne Friederich, Erich Classen, and Marta Kasper for help with project logistics; Connie Bormans for providing modern control samples; Yuri Seryogin for software support and genetic distance calculations; and Pierre Zalloua and Christian Anderson for helpful comments.

Members of the Genographic Consortium. Syama Adhikarla, Madurai Kamaraj University, Madurai, Tamil Nadu, India; Doron M. Behar, Rambam Medical Center, Haifa, Israel; Jaume Bertranpetit, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain; Andrew C. Clarke, University of Otago, Dunedin, New Zealand; David Comas, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain; Matthew C. Dulik, University of Pennsylvania, Philadelphia, Pennsylvania, United States; Christoff J. Erasmus, National Health Laboratory Service, Johannesburg, South Africa; Jill B. Gaieski, University of Pennsylvania, Philadelphia, Pennsylvania, United States; ArunKumar GaneshPrasad, Madurai Kamaraj University, Madurai, Tamil Nadu, India; Angela Hobbs, National Health Laboratory Service, Johannesburg, South Africa; Asif Javed, IBM, Yorktown Heights, New York, United States; Li Jin, Fudan University, Shanghai, China; Matthew E. Kaplan, University of Arizona, Tucson, Arizona, United States; Shilin Li, Fudan University, Shanghai, China; Begoña Martínez-Cruz, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain; Elizabeth A. Matisoo-Smith, University of Otago, Dunedin, New Zealand; Marta Melé, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain; Nirav C. Merchant, University of Arizona, Tucson, Arizona, United States; R. John Mitchell, La Trobe University, Melbourne, Victoria, Australia; Amanda C. Owings, University of Pennsylvania, Philadelphia, Pennsylvania, United States; Laxmi Parida, IBM, Yorktown Heights, New York, United States; Ramasamy Pitchappan, Madurai Kamaraj University, Madurai, Tamil Nadu, India; Daniel E. Platt, IBM, Yorktown Heights, New York, United States; Lluis Quintana-Murci, Institut Pasteur, Paris, France; Colin Renfrew, University of Cambridge, Cambridge, United Kingdom; Daniela Rodrigues Lacerda, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; Ajay K. Royyuru, IBM, Yorktown Heights, New York, United States; Fabrício R. Santos, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; Theodore G. Schurr, University of Pennsylvania, Philadelphia, Pennsylvania, United States; Himla Soodyall, National Health Laboratory Service, Johannesburg, South Africa; David F. Soria Hernanz, National Geographic Society, Washington, District of Columbia, United States; Pandikumar Swamikrishnan, IBM, Somers, New York, United States; Chris Tyler-Smith, The Wellcome Trust Sanger Institute, Hinxton, United Kingdom; Kavitha Valampuri John, Madurai Kamaraj University, Madurai, Tamil Nadu, India; Arun Varatharajan Santhakumari, Madurai Kamaraj University, Madurai, Tamil Nadu, India; Pedro Paulo Vieira, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; R. Spencer Wells, National Geographic Society, Washington, District of Columbia, United States; Janet S. Ziegle, Applied Biosystems, Foster City, California, United States.

Author Contributions

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: WH OB JJS KWA AC. Performed the experiments: WH OB CJA. Analyzed the data: WH OB SK CJA CSIDS CS NN. Contributed reagents/materials/analysis tools: WH JJS SK VZ GB VD BF EB RV HM KWA. Wrote the paper: WH OB JJS CJA BF AC.

References

1. Price TD, ed (2000) Europe's first farmers. Cambridge: Cambridge University Press. 395 p.
2. Whittle AWR, Cummings V, eds (2007) Going over: the mesolithic-neolithic transition in North-West Europe. Oxford: Oxford University Press. 632 p.

3. Ammerman AJ, Cavalli-Sforza LL (1984) The neolithic transition and the genetics of populations in Europe. Princeton (New Jersey): Princeton University Press. 193 p.
4. Boyle K, Renfrew C, eds (2000) Archaeogenetics: DNA and the population prehistory of Europe. Cambridge: McDonald Institute for Archaeological Research. 342 p.
5. Whittle AWR (1996) Europe in the Neolithic: the creation of new worlds. Cambridge: Cambridge University Press. 443 p.
6. Dupanloup I, Bertorelle G, Chikhi L, Barbujani G (2004) Estimating the impact of prehistoric admixture on the genome of Europeans. Mol Biol Evol 21: 1361–1372.
7. Richards M (2003) The neolithic invasion of Europe. Annu Rev Anthropol 32: 135–162.
8. Soares P, Achilli A, Semino O, Davies W, Macaulay V, et al. (2010) The archaeogenetics of Europe. Curr Biol 20: R174–R183.
9. Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, Barbujani G (1998) Clines of nuclear DNA markers suggest a largely neolithic ancestry of the European gene pool. Proc Natl Acad Sci U S A 95: 9053–9058.
10. Chikhi L, Nichols RA, Barbujani G, Beaumont MA (2002) Y genetic data support the Neolithic demic diffusion model. Proc Natl Acad Sci U S A 99: 11008–11013.
11. Balaresque P, Bowden GR, Adams SM, Leung HY, King TE, et al. (2010) A predominantly neolithic origin for European paternal lineages. PLoS Biol 8: e1000285. doi:10.1371/journal.pbio.1000285.
12. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67: 1251–1276.
13. Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, et al. (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science 290: 1155–1159.
14. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, et al. (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. Am J Hum Genet 75: 910–918.
15. Battaglia V, Fornarino S, Al-Zahery N, Olivieri A, Pala M, et al. (2008) Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe. Eur J Hum Genet 8: 820–830.
16. Di Benedetto G, Nasidze IS, Stenico M, Nigro L, Krings M, et al. (2000) Mitochondrial DNA sequences in prehistoric human remains from the Alps. Eur J Hum Genet 8: 669–677.
17. Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, et al. (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.
18. Sampietro ML, Lao O, Caramelli D, Lari M, Pou R, et al. (2007) Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. Proc Biol Sci 274: 2161–2167.
19. Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, et al. (2005) Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. Science 310: 1016–1018.
20. Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, et al. (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science 326: 137–140.
21. Malmstrom H, Gilbert MT, Thomas MG, Brandstrom M, Stora J, et al. (2009) Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. Curr Biol 19: 1758–1762.
22. Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, et al. (2004) Genetic analyses from ancient DNA. Annu Rev Genet 38: 645–679.
23. Gronenborn D (1999) A variation on a basic theme: the transition to farming in southern central Europe. J World Prehist 13: 123–210.
24. Banffy E (2004) The 6th millenium BC boundary in western Transdanubia and its role in the Central European neolithic transition (the Szentgyörgyölgy-Pityerdomb settlement). Budapest: Archaeological Institute of the Hungarian Academy of Science.
25. Behar DM, Rosset S, Blue-Smith J, Balanovsky O, Tzur S, et al. (2007) The Genographic Project public participation mitochondrial DNA database. PLoS Genet 3: e104. doi/10.1371/3journal.pgen.0030104.
26. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, et al. (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18: 830–838.
27. Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, et al. (2005) Genomic sequencing of Pleistocene cave bears. Science 309: 597–599.
28. Malmstrom H, Svensson EM, Gilbert MT, Willerslev E, Gotherstrom A, et al. (2007) More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA. Mol Biol Evol 24: 998–1004.
29. Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. Proc Natl Acad Sci U S A 101: 975–979.
30. Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. Mol Biol Evol 23: 482–490.
31. Keyser C, Bouakaze C, Crubézy E, Nikolaev V, Montagnon D, et al. (2009) Ancient DNA provides new insights into the history of south Siberian Kurgan people. Hum Genet 126: 395–410.
32. Gronenborn D (2007) Beyond the models: 'Neolithisation' in Central Europe. In: Whittle A, Cummings V, eds. Going over: the Mesolithic-Neolithic transition in North-West Europe. Oxford: Oxford University Press. pp 73–98.
33. Bogucki P (2000) How agriculture came to North-Central Europe. In: Price TD, ed. Europe's first farmers. Cambridge: Cambridge University Press. pp 197–218.
34. Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, et al. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. Am J Hum Genet 69: 844–852.
35. Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, et al. (2010) A complete mtDNA genome of an early modern human from Kostenki, Russia. Curr Biol 20: 231–236.
36. Barbujani G, Bertorelle G, Chikhi L (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. Am J Hum Genet 62: 488–492.
37. Ho SY, Shapiro B, Phillips MJ, Cooper A, Drummond AJ (2007) Evidence for time dependency of molecular rate estimates. Syst Biol 56: 515–522.
38. Bocquet-Appel JP, Naji S, Vander Linden M, Kozlowski JK (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.
39. Bocquet-Appel JP (2002) Paleoanthropological traces of a Neolithic demographic transition. Curr Anthropol 43: 637–650.
40. Pinhasi R, Fort J, Ammerman AJ (2005) Tracing the origin and spread of agriculture in Europe. PLoS Biol 3: e410. doi/10.1371/journal.pbio.0030410.
41. Pinhasi R, von Cramon-Taubadel N (2009) Craniometric data supports demic diffusion model for the spread of agriculture into Europe. PLoS ONE 4: e6747. doi/10.1371/journal.pone.0006747.
42. Zvelebil M, Lillie M, Price TD (2000) Transition to agriculture in Eastern Europe. In: Price TD, ed. Europe's first farmers. Cambridge: Cambridge University Press. pp 57–92.
43. Müller U (2002) Vorläufige Ergebnisse der Untersuchungen in Derenburg, Meerenstieg II, Ldkr. Wernigerode. Jahresschr Mitteldt Vorgesch 85: 77–90.
44. Fritsch B, Claßen E, Müller U, Dresely V (2010) Die linienbandkeramischen Gräberfelder von Derenburg "Meerenstieg II" und Halberstadt "Sonntagsfeld", Lkr. Harz. Jahresschr Mitteldt Vorgesch;In press.
45. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23: 147.
46. Brandstatter A, Parsons TJ, 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.
47. Sanchez JJ, Borsting C, Hallenberg C, Buchard A, Hernandez A, et al. (2003) Multiplex PCR and minisequencing of SNPs—a model with 35 Y chromosome SNPs. Forensic Sci Int 137: 74–84.
48. Brion M, Sanchez JJ, Balogh K, Thacker C, Blanco-Verea A, et al. (2005) Introduction of an single nucleodite polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. Electrophoresis 26: 4411–4420.
49. Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, et al. (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27: 1713–1724.
50. Sanchez JJ, Endicott P (2006) Developing multiplexed SNP assays with special reference to degraded DNA templates. Nat Protoc 1: 1370–1378.
51. Endicott P, Metspalu M, Stringer C, Macaulay V, Cooper A, et al. (2006) Multiplexed SNP typing of ancient DNA clarifies the origin of andaman mtDNA haplogroups amongst south Asian tribal populations. PLoS ONE 1: e81. doi/10.1371/journal.pone.0000081.
52. Bouakaze C, Keyser C, Amory S, Crubezy E, Ludes B (2007) First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. Int J Legal Med 121: 493–499.
53. Bouakaze C, Keyser C, Crubezy 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: 315–325.
54. Gilbert MT, Sanchez JJ, Haselkorn T, Jewell LD, Lucas SB, et al. (2007) Multiplex PCR with minisequencing as an effective high-throughput SNP typing method for formalin-fixed tissue. Electrophoresis 28: 2361–2367.
55. Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, et al. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67: 1526–1543.
56. Doi Y, Yamamoto Y, Inagaki S, Shigeta Y, Miyaishi S, et al. (2004) A new method for ABO genotyping using a multiplex single-base primer extension reaction and its application to forensic casework samples. Leg Med (Tokyo) 6: 213–223.
57. Cooper A, Poinar HN (2001) Ancient DNA: do it right or not at all. Science 289: 1139.
58. Pruvost M, Geigl E-M (2004) Real-time quantitative PCR to assess the authenticity of ancient DNA amplification. J Archaeol Sci 31: 1191–1197.
59. Handt O, Krings M, Ward RH, Paabo S (1996) The retrieval of ancient human DNA sequences. Am J Hum Genet 59: 368–376.
60. Gilbert MTP, Bandelt H-J, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. Trends Ecol Evol 20: 541–544.
61. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.
62. Nei M (1972) Genetic distance between populations. Am Nat 106: 283–292.
63. Balanovsky O, Rootsi S, Pshenichnov A, Kivisild T, Churnosov M, et al. (2008) Two sources of the Russian patrilineal heritage in their Eurasian context. Am J Hum Genet 82: 236–250.

64. 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.
65. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111–120.
66. Ho SY, 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.
67. Chan YL, Anderson CN, Hadly EA (2006) Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. PLoS Genet 2: e59. doi/10.1371/journal.pgen.0020059.
68. Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA (2005) Serial SimCoal: A population genetic model for data from multiple populations and points in time. Bioinformatics 21: 1733–1734.
69. Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. Genetics 162: 2025–2035.
70. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19: 716–723.
71. Burnham KP, Anderson DR (2002) Model selection and multimodel inference: A practical information-theoretic approach, 2nd edition. New York: Springer.
72. Posada D, Buckley TR (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.
73. Danzeglocke U, Jöris O, Weninger B (2007) CalPal online: Cologne radiocarbon & paleoclimate research package. Available: http://www.calpal-online.de. Accessed 2 November 2009.