Towards a new history and geography of human genes informed by ancient DNA

Joseph K. Pickrell[1,2*], David Reich[3,4,5*]

[1]New York Genome Center, New York, NY
[2]Department of Biological Sciences, Columbia University, New York, NY
[3]Department of Genetics, Harvard Medical School, Boston, MA
[4]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA
[5]Broad Institute of MIT and Harvard, Cambridge, MA

*To whom correspondence should be addressed: jkpickrell@nygenome.org (JKP),
reich@genetics.med.harvard.edu (DR)

1

Abstract

Genetic information contains a record of the history of our species, and technological advances have transformed our ability to access this record. Many studies have used genome-wide data from populations today to learn about the peopling of the globe and subsequent adaptation to local conditions. Implicit in this research is the assumption that the geographic locations of people today are informative about the geographic locations of their ancestors in the distant past. However, it is now clear that long-range migration, admixture and population replacement have been the rule rather than the exception in human history. In light of this, we argue that it is time to critically re-evaluate current views of the peopling of the globe and the importance of natural selection in determining the geographic distribution of phenotypes. We specifically highlight the transformative potential of ancient DNA. By accessing the genetic make-up of populations living at archaeologically-known times and places, ancient DNA makes it possible to directly track migrations and responses to natural selection.

Introduction

Within the past 100,000 years, anatomically modern humans have expanded to occupy every habitable area of the globe. The history of this expansion has been explored with tools from a number of disciplines, including linguistics, archaeology, physical anthropology, and genetics. How did we get to where we are today?

Genetic information, since it is passed from generation to generation so faithfully, contains powerful information about population history. Technological advances have now made all of this information easily accessible. The great synthesis of genetic data with historical, archaeological and linguistic information, "The History and Geography of Human Genes" (Cavalli-Sforza et al., 1994), was written based on data from around one hundred protein polymorphisms. However, it is now possible to genotype millions of polymorphisms in thousands of individuals using high-throughput sequencing at an affordable cost (Abecasis et al., 2010).

With the advent of extensive genome-wide data, geneticists have been able to propose more refined models for the history of the human expansion across the globe. Probably the most popular model of this type, the "serial founder effect" model, is one in which humans expanded around the world in a series of successive population splits and range expansions into unoccupied territory, with little subsequent migration. Under this model, the present-day inhabitants of any region are largely descended from the first inhabitants of the region (Harpending and Eller, 2000; Harpending and Rogers, 2000; Hellenthal et al., 2008; Henn et al., 2012a; Prugnolle et al., 2005; Ramachandran et al., 2005).

It is now clear that such models are inaccurate for many populations. The present-day inhabitants of many regions of the world do not descend in simple ways from the populations that lived in the same locations in the distant past. Instead, human populations have periodically experienced major demographic upheavals, such that much of the geographic information about the first human migrations has been overwritten by subsequent population admixture or replacement. This idea has a long (albeit controversial) history in archaeology; it has been debated for decades whether sudden changes in material culture apparent in the archaeological record can be attributed to population movements (Renfrew and Bahn, 1996). In genetics, the series of discoveries documenting the importance of mixture and migration in human history have come in part from analysis of populations living today (e.g. Hellenthal et al., 2014; Loh et al., 2013; Patterson et al., 2012), but also increasingly from genetic analysis of ancient human remains (e.g. Raghavan et al., 2013; Skoglund et al., 2012).

In this paper, we argue that these results motivate a systematic re-evaluation of human history using modern genomic tools—a new "History and Geography of Human Genes" that exploits many orders of magnitude more data than the original synthesis—and we emphasize the potential for using ancient DNA to answer outstanding questions. Other recent reviews have had different emphases, or have

3

covered methodological approaches in more detail than we do here (Alves et al., 2012; Colonna et al., 2011; Novembre and Ramachandran, 2011; Pinhasi et al., 2012; Stoneking and Krause, 2011; Veeramah and Hammer, 2014; Wall and Slatkin, 2012)

Re-evaluation of the "serial founder effect" model

For about the last ten years, genetic studies of early human history have often been framed around the "serial founder effect" model. This model, initially proposed by Harpending, Eller, and Rogers (Harpending and Eller, 2000; Harpending and Rogers, 2000), gained popularity with the publication of two papers (Prugnolle et al., 2005; Ramachandran et al., 2005) that observed that heterozygosity (the average number of differences between two random copies of a genome in a population) declines approximately linearly with geographic distance from Africa. This pattern, initially identified in genome-wide microsatellite genotypes from 53 worldwide human populations, was subsequently confirmed in large single nucleotide polymorphism (SNP) datasets (Li et al., 2008).

The observation of declining human genetic variation with increasing geographic distance from Africa was interpreted as being consistent with a "serial founder effect" model for human dispersal. In this type of model (Figure 1A), the peopling of the globe proceeded by an iterative process in which small bands of individuals pushed into unoccupied territory, experienced population expansions, and subsequently gave rise to new small bands of individuals who then pushed further into unoccupied territory. This model has two important features: a large number of expansions into new territory by small groups of individuals (and concurrent bottlenecks), and little subsequent migration. As Prugnolle et al. (2005) wrote: "what is clear…is that [this] pattern of constant loss of genetic diversity along colonisation routes could only have arisen through successive bottlenecks of small amplitude as the range of our species increased… The pattern we observe also suggests that subsequent migration was limited or at least very localized".

This model has been influential in many fields. If the serial founder event model is correct, then the difficult problem of identifying the geographic origin of all modern humans is reduced to the simpler problem of finding the geographic region where people have the most genetic diversity (Ramachandran et al., 2005). This idea has been used extensively in discussions of human origins (e.g. Henn et al., 2011; Schlebusch et al., 2012). Outside of genetics, the serial founder effect model has been used as a framework to interpret data from linguistics (Atkinson, 2011), physical anthropology (Betti et al., 2012; von Cramon-Taubadel and Lycett, 2008; Hanihara, 2008; Manica et al., 2007), material culture (Rogers et al., 2009), and economics (Ashraf and Galor, 2013).

A caveat of the serial founder effect model is that many other historical models can produce qualitatively similar patterns of genetic diversity (e.g. Amos and Hoffman, 2010; DeGiorgio et al., 2009). Producing this empirical pattern simply requires that the average time to the most recent common ancestor between two chromosomes

4

in a population depend on the distance of that population from Africa (DeGiorgio et al., 2009). To illustrate this point, we simulated genetic data using *ms* (Hudson, 2002) under three demographic models: the serial founder effect model (Figure 1A), a model with two large founder effects and extensive subsequent population mixture (Figure 1B), and a model without bottlenecks but with archaic admixture (from an anciently diverged population like Neanderthals) as well as more extensive recent population mixture (Figure 1C). Details of the simulation parameters are in the Supplementary Information. We simulated 1,000 regions of 100kb under each model, and plotted the average heterozygosity in each population. In all three cases (Figure 1D,E,F) we recapitulate an approximately linear decline in heterozygosity with distance from a source. These simulations show that the main observation that has been marshaled in support of the serial founder effect model is also consistent with very different histories. Other types of evidence are necessary to distinguish between these very distinct and in our view equally plausible models.
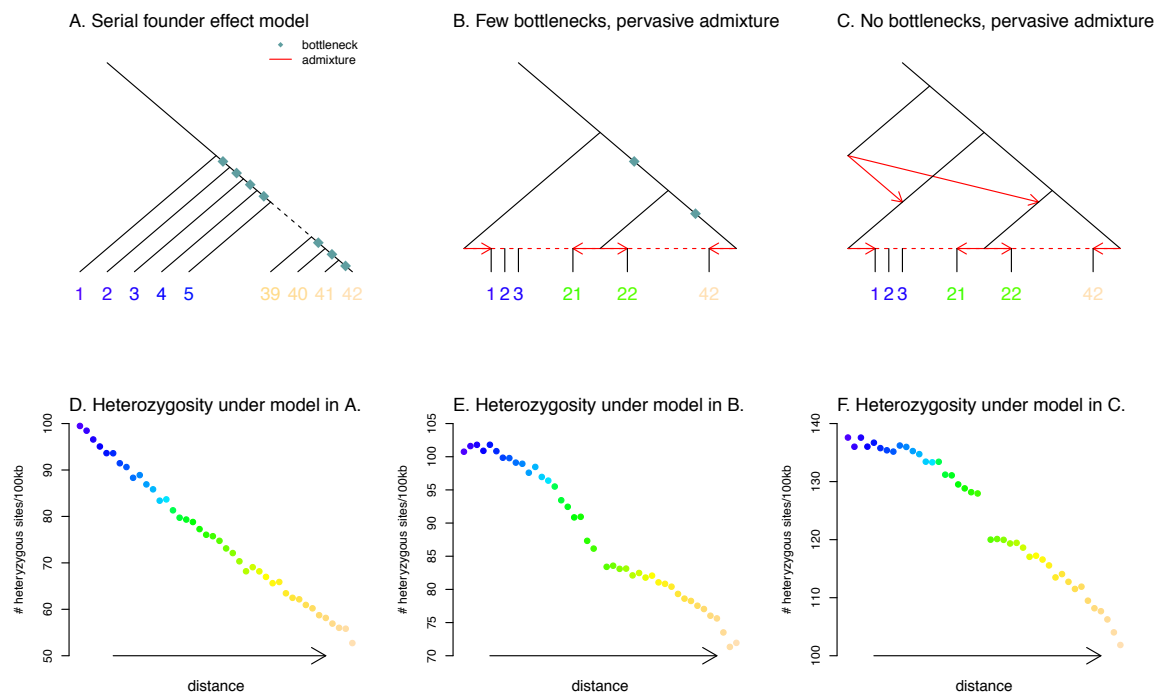

Figure 1: A negative correlation between heterozygosity and geographic distance from a source population can be generated by qualitatively different, historically plausible demographic models. We simulated genetic data under different demographic models and calculated the average heterozygosity in each simulated population. **A.** Schematic of a serial founder effect model. **B.** Schematic of a demographic model with two bottlenecks and extensive admixture. **C.** Schematic of a demographic model with no bottlenecks and extensive admixture. **D,E,F**. Average heterozygosity in each population simulated under the demographic models in A,B, and C respectively. Each point represents a population, ordered along the x-axis according as in A.

The current inhabitants of a region are often poor representatives of the historical populations who lived in the same locations

A parsimonious assumption underlying much work on human population history is that populations living in a geographic region today are largely descended from the

5

first anatomically modern humans to arrive in the region. Recent work, however, has shown that the truth is far from parsimonious. Instead, long-range migration and concomitant population replacement or admixture has occurred often enough in human history that the present-day inhabitants of many places in the world often bear little genetic relationship to the more ancient peoples of the same region.

The present-day populations of the Americas represent a recent example of this phenomenon. It is well known that the populations of the Americas experienced demographic upheaval after the arrival of Europeans and Africans around 500 years ago. Accounting for mixture between Europeans, Africans and Native Americans is crucial to understanding population structure in the Americas. In fact, most of the ancestry of present-day populations in the Americas is not derived from the Native Americans who were the sole inhabitants of the region less than half a millennium ago (Bryc et al., 2010; Gravel et al., 2013; Johnson et al., 2011; Moreno-Estrada et al., 2013; Price et al., 2007; Wang et al., 2007). Another recent example is Australia, where accounting for massive European migration over the last couple of hundred years is crucial for understanding population structure (McEvoy et al., 2010).

Further back in time, the present-day hunter-gatherer and pastoralist populations of Siberia are often treated as surrogates for the populations that first crossed the Bering land bridge to people the Americas beginning more than 15,000 years ago (e.g. Lell et al., 2002; Santos et al., 1999; Starikovskaya et al., 2005; Wang et al., 2007). However, just in the last year, genetic data showed that this assumption is wrong. The sequencing of DNA from an individual who lived ~24,000 years before the present in the Lake Baikal region of Siberia, and a ~17,000 year old individual from the same region, showed that these ancient individuals (from time points prior to the time the Americas were thought to be peopled) are phylogenetically more closely related to present-day Native Americans than are present-day Siberians (Raghavan et al., 2013). These ancient individuals appear to have been members of an "Ancient North Eurasian" population that no longer exists in unmixed form, but that admixed substantially with the ancestors of both present-day Europeans and Native Americans (Patterson et al., 2012; Raghavan et al., 2013; Lazaridis et al., 2013). In contrast, most of the present-day indigenous populations of Siberia are more closely related to East Asians, indicating that the present-day indigenous population of Siberia is descended in large part from populations that arrived in the region after the end of the last ice age (Raghavan et al., 2013).

Ancient DNA studies have also revealed population turnover over shorter timescales (Bramanti et al., 2009; Brandt et al., 2013; Haak et al., 2005, 2010; Keller et al., 2012; Lazaridis et al., 2013; Malmström et al., 2009; Skoglund et al., 2012). A major debate in the last few decades among archaeologists and geneticists has been about whether the arrival of agriculture in Europe involved the spread of people or culture. Skoglund et al. (2012) addressed this controversy by analyzing genome-wide genetic data from Swedish hunter-gatherer and agriculturalist populations that lived around 5,000 years ago. The farmer population appears most genetically similar to southern Europeans today, while the hunter-gatherers are more similar to

6

northern Europeans. Thus, at least in Scandinavia, the spread of agriculture was accompanied by the spread of people. Mitochondrial DNA studies have shown that such population turnovers accompanied the arrival of agriculture throughout Europe (Bramanti et al., 2009; Brandt et al., 2013; Haak et al., 2005, 2010).

The arrival of farmers was not the end of pre-historic population turnover in Europe (and we do not even discuss here the turnovers that occurred in the few thousand years in Europe since the invention of writing; see Davies, 1998; Hellenthal et al., 2014; Moorjani et al., 2011; Ralph and Coop, 2013). The most notable study in this regard is the one of Brandt et al. (2013), which assembled mtDNA data from 364 human samples from archaeological cultures ranging from the early Neolithic to the Bronze Age from the same geographic area in Germany. This study documented discontinuity between people of early and late Neolithic cultures, with people of late Neolithic cultures bearing more relatedness to the present-day populations of Eastern Europe and Russia than do people of early Neolithic cultures. Thus, major demographic turnover has happened at least twice over the course of the last eight thousand years of European prehistory. This makes inferences about the inhabitants of Europe *tens* of thousands of years ago based on the locations of people today unreliable.

Evidence of major demographic changes in the last several thousand years has also accumulated in parts of the world where no ancient DNA is (yet) available:

- In India, nearly all people today are admixed between two distinct genetic groups, one most closely related to present-day Europeans, Central Asians, and Near Easterners, and one most closely related to an isolated population living in the Andaman islands (Reich et al., 2009). Moorjani et al., 2013 showed that much of this admixture occurred within the last 4,000 years.

- In North Africa, nearly all people today descend from admixture between populations related to those present today in western Africa and in the Near East (Henn et al., 2012b; Rosenberg et al., 2002; Tishkoff et al., 2009). Some of this admixture can be dated to within the last few thousand years (Hellenthal et al., 2014), indicating that much of the ancestry from these populations does not descend continuously from the Stone Age peoples of North Africa.

- In sub-Saharan Africa, genetic studies have documented multiple examples of populations with ancestry from disparate sources. Populations that speak languages with heavy use of click consonants in eastern Africa (Hadza and Sandawe) result from admixture involving a population related to hunter-gatherer populations southern Africa (Khoisan) (Pickrell et al., 2012), suggesting that Khoisan-related populations were once more widespread than they are today. Further, populations in eastern and southern Africa have been influenced by gene flow from west Eurasian populations in the last

7

3,000 years (Pagani et al., 2012; Pickrell et al., 2014). And in Madagascar, all populations derive approximately half of their ancestry from Austronesian speaking migrants from southeast Asia (Pierron et al., 2014).

Beyond case studies, Patterson et al. (2012), Loh et al. (2013), and Hellenthal et al. (2014) applied tests for admixture to diverse sets of population from around the globe and found that nearly all populations show evidence of admixture. To illustrate the degree to which this admixture involved populations that today are geographically distant, in Figure 2 we present an analysis based on combining data for 103 worldwide populations from a number of sources (Altshuler et al., 2010; Behar et al., 2010; Henn et al., 2011; Li et al., 2008; Pagani et al., 2012; Schlebusch et al., 2012) and running a simple three-population test for admixture (Reich et al., 2009) on all populations (see Supplementary information for details). For all populations with significant evidence of admixture, we identified the present-day populations that are most closely related to the admixing populations.
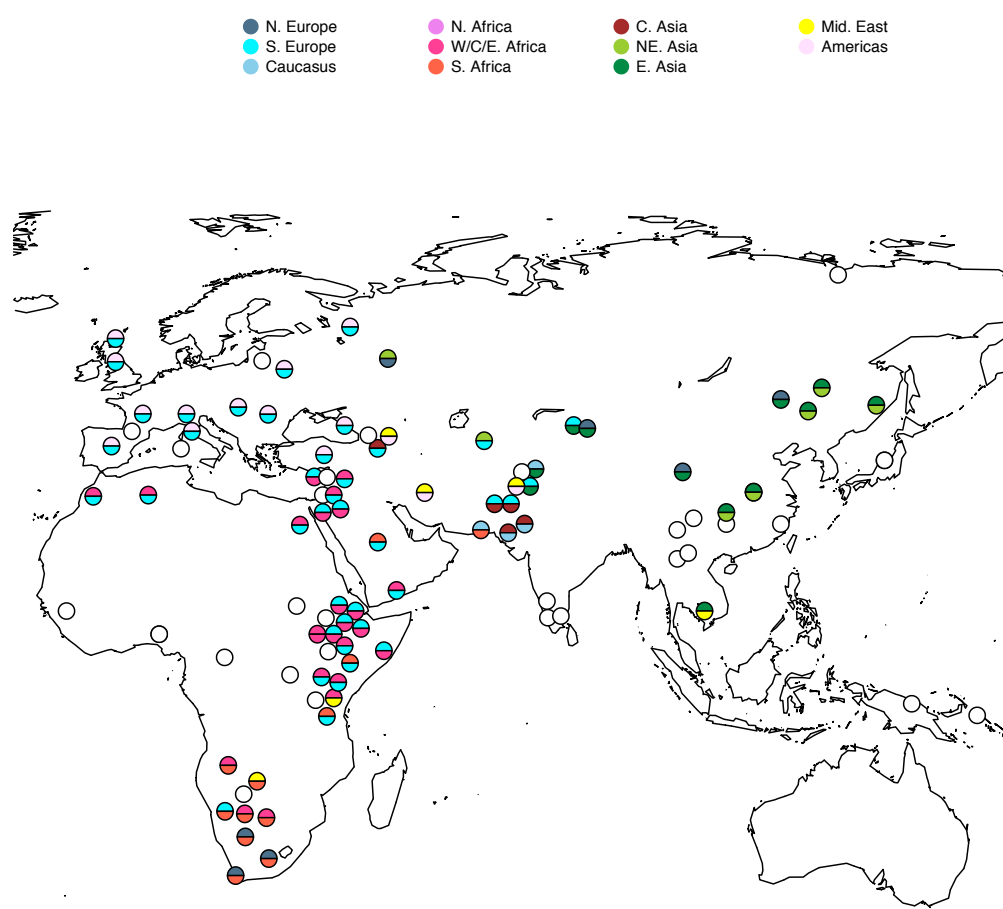

Figure 2: A birds-eye view of admixture in human populations. We performed a three-population test for admixture (Reich et al., 2009) on 103 worldwide populations. Circles show the approximate current geographic locations of all tested populations (except for populations in the Americas, which are not plotted for ease of display). Filled circles represent populations identified as admixed, and the colors represent the current geographic labels of the inferred admixing populations. Empty circles represent populations with no statistically-significant evidence for admixture in this test.

8

The results are visually startling. Figure 2 shows the geographic locations of all the populations, along with the locations of the best present-day proxies of their ancestral populations. Admixture between populations related to ones that are now geographically distant is evident in most populations of the world. For example, Native American-related ancestry is present throughout Europe (Patterson et al., 2012), likely reflecting the genetic input from the Ancient Northern Eurasian population related to Upper Paleolithic Siberians (Lazaridis et al., 2013; Patterson et al., 2012) both into the Americas (most likely prior to 15,000 years ago) and into Europe (Lazaridis et al., 2013; Raghavan et al., 2013). Also, Near Eastern-related ancestry is found in Cambodia (Pickrell and Pritchard, 2012), likely due to mixture from an ancestral South Asian population that was itself an admixed population containing ancestry related to present-day Near Easterners (Hellenthal et al., 2014). The test we use as the basis for Figure 2 detects only the largest signals of admixture in the data, and cannot detect complete population replacement. The true population history is likely to have been even more complex.

These examples show that the parsimonious assumption often made in population genetics—that the populations in a given region today are largely descended from the inhabitants of that geographic region in the distant past—is rarely true. Combined with the simulations presented in Figure 1, these results demonstrate that the serial founder effect model is no longer a reasonable null hypothesis for modeling the ancient spread of anatomically modern humans around the globe. This motivates a search for alternative models that may better fit the data.
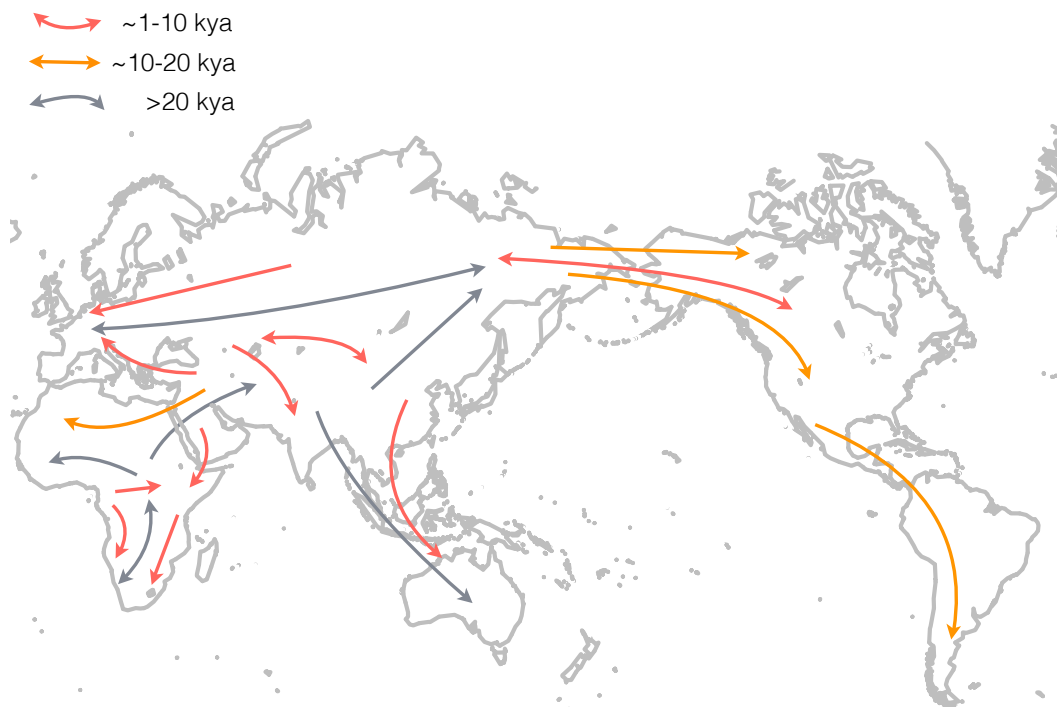

Figure 3: A rough guide to genetically documented population movements in the history of anatomically modern humans. Arrows represent population movements, and the color of each arrow represents the approximate time of the migration.

9

The type of model that is emerging from the data is presented in Figure 3. In this model, the major land masses of the world have all been peopled multiple times, with population replacement and/or admixture happening periodically throughout history. As new data become available, this model will likely be refined and completely re-written in parts.

Ancient DNA: a transformative source of information about the past

The types of models that are most useful for making sense of human history are ones that specify geographic as well as temporal information; that is, those that make statements of the form "a population from location X moved to location Y during time period Z". For example, one might wish to test whether the first pastoralists in southern Africa arrived as migrants from eastern Africa after around 2,500 years ago (the time when evidence of pastoralism in southern Africa appears in the archaeological record), against the alternative hypothesis that a pastoralist lifestyle was adopted by indigenous people who learned it through cultural transmission. Testing this hypothesis using DNA from individuals today involves assuming that populations in southern and eastern Africa today are representative of the populations in southern and eastern Africa at the times of interest. As discussed above, this assumption is often untrue.

Studies of DNA from ancient human remains offer a way around this limitation. Rather than studying the past by the traces it has left in present-day people—which is problematic since human individuals and even whole populations are capable of migrating hundreds or thousands of kilometers in a lifetime—ancient DNA offers the ability to analyze the genetic patterns that existed at a particular time at a geographical location. This allows direct inference about the relationships of historical populations to each other and to populations living today. Below we list some surprising findings about human history based on ancient DNA, which could not have been obtained without this source of information:

- There was archaic Neanderthal gene flow into the ancestors of all anatomically modern humans outside African 37,000-85,000 years ago (Sankararaman et al., 2012). This gene flow contributed on the order of 2% of the genetic ancestry of non-Africans (Green et al., 2010; Prüfer et al., 2014). This finding was suggested by genetic analysis prior to the use of ancient DNA to document this admixture event (Plagnol and Wall, 2006). However, the consensus about whether admixture occurred only changed after ancient DNA evidence showed that the deeply diverged segments in present-day non-Africans are related to Neanderthals (Green et al., 2010).

- A previously unknown archaic population that was neither Neanderthal nor modern human was present in Siberia before 50,000 years ago (Krause et al., 2010; Reich et al., 2010). The discovery of the "Denisovans" shows how

10

ancient DNA can reveal "genomes in search of a fossil"; populations that were not expected at all based on the archaeological or fossil record.

- There was gene flow from a population related to the Denisovans into the ancestors of present-day aboriginal people from New Guinea, Australia, and the Philippines (Cooper and Stringer, 2013; Meyer et al., 2012; Prüfer et al., 2014; Reich et al., 2010, 2011). It is notable that the populations today that contain the largest proportion of Denisovan ancestry are in Oceania, far from where the Denisovan bone was found in Siberia, a finding that again would not have been expected in the absence of ancient DNA data.

- There was admixture between an Ancestral North Eurasian population and a population related to present-day East Asians in the ancestry of Native Americans prior to the diversification of Native American populations in the New World (Raghavan et al., 2013).

- There were at least two important changes in population structure in central Europe over the course of the last 9,000 years (Bramanti et al., 2009; Brandt et al., 2013; Haak et al., 2005, 2010; Keller et al., 2012; Lazaridis et al., 2013; Malmström et al., 2009; Skoglund et al., 2012).

- Populations in the Americas have experienced multiple episodes of population admixture or turnover, such that an individual living in Greenland around 4,000 years ago is more closely related to populations currently in Siberia than to present-day Greenland Inuit populations (Rasmussen et al., 2010), and an individual living in North America around 12,000 years ago is more closely related to present-day Native South American populations than to Native North American populations (Rasmussen et al., 2014).

Ancient DNA results are so regularly surprising that almost any measurement is interesting: new historical discoveries have been made in virtually every ancient DNA study that has been carried out. The reason why ancient DNA studies are so informative is that the technology provides a tool to measure quantities that were previously unmeasurable. In this sense, the value of ancient DNA technology as a window into ancient migrations is analogous to the 17th century invention of the light microscope as a window into the world of microbes and cells.

Scientific opportunities for ancient DNA studies

Moving forward, ancient DNA studies afford major opportunities in two areas: studies of population history, and studies of natural selection.

The literature on ancient mitochondrial DNA contains a number of promising study designs. For example, one might sample multiple individuals from different archaeological cultures at a single time point (Malmström et al., 2009). Such a

"horizontal time slice" allows a snapshot of population structure over a broad geographic region, which can then be compared to the relatively complete picture of population structure today. An alternative study design (like that used by Brandt et al., 2013) is to take a single geographic location and sample individuals from multiple time points. This "vertical time slice" allows direct quantification of changes in population composition over time.

Mitochondrial DNA studies have major drawbacks compared with analysis of the whole genome, however (Ballard and Whitlock, 2004). First, mtDNA is inherited maternally, and thus does not capture any information about the history of males (which may differ from that of females due to sex-biased demographic processes). More importantly, a study of a single locus (or two loci if the Y chromosome is included to avoid the first limitation) has less statistical resolution for studies of history than do studies of the nuclear genome. The reason for this is that whole genome studies of an individual obtain information about hundreds of thousands of that individual's ancestors, not just those on a single lineage. It is thus important that the more advanced study designs of the mtDNA studies be combined with analysis of the more informative autosomal DNA.

What outstanding questions in population history can be addressed with these study designs? One important question is whether changes in populations over time are typically gradual—due to consistent, low-level gene flow between neighboring populations—or punctate, with migration events rapidly altering the population genetic composition of a region. One line of work on modeling human history explicitly assumes the latter (Hellenthal et al., 2014; Lipson et al., 2013; Loh et al., 2013; Patterson et al., 2012; Pickrell and Pritchard, 2012; Reich et al., 2009). If this assumption is unfounded, however, other models that accommodate continuous gene flow (e.g. Excoffier et al., 2013; Gutenkunst et al., 2009) may be more appropriate. In principle, distinguishing between these possibilities with a time series of ancient DNA is straightforward; in Figure 4A we show how the predictions differ.

Ancient DNA is also a promising way to address ongoing historical debates about the origins of different populations. Do linguistic isolates like the Basques in Europe have more ancestry from the pre-farming Mesolithic populations of the region than do their neighbors (Behar et al., 2012)? When did the west Eurasian-related population(s) that admixed with most Indian populations first appear in South Asia (Reich et al., 2009)? Did the first modern human inhabitants of East Asia descend from an earlier out-of-Africa migration than the populations living in East Asia today (Rasmussen et al., 2011; Reich et al., 2011)? Answering these questions will require a time series of snapshots of human genetic structure, combining the two types of study design mentioned above. If experience is a guide, this type of information will uncover additional unexpected aspects of human history.

The admixture and population replacements identified by ancient DNA also have implications for studies of natural selection. It is often assumed that populations

have been in their current geographic locations long enough to adapt to their local conditions. This is explicit in approaches that look for correlations between environmental variables and allele frequencies (e.g. Coop et al., 2010; Hancock et al., 2010a; Huerta-Sánchez et al., 2013) and implicit in studies that interpret selected loci in terms of the current locations of populations (e.g. Simonson et al., 2010).

To what extent are the geographic distributions of selected alleles today indicative of the geographic distributions of selective pressures? (Note that the answer to this question may depend on the selective pressure in question). In individual cases, these two distributions are highly correlated; the classic example is the correlation between malaria incidence and disorders of hemoglobin (Kwiatkowski, 2005). For other cases, the correlation is imperfect. For example, alleles causing light skin pigmentation are at high frequency in northern Africa (Norton et al., 2007). If the selective pressure causing light skin is indeed a relative lack of ultraviolet radiation (Jablonski and Chaplin, 2000), it seems reasonable to expect that people living around the Sahara desert will be unaffected by this selection pressure. It is thus likely that the high frequency of alleles causing light skin pigmentation in north Africa was caused by the arrival of lightly-pigmented people (Henn et al., 2012b).

More generally, it has been observed that allele frequencies at loci under natural selection in humans tend to track neutral population structure rather than any obvious geographical variation in selection pressures (Coop et al., 2009; Granka et al., 2012). One factor contributing to this observation may be that selection pressures on individual loci are relatively weak due to the quantitative nature of phenotypes (Hancock et al., 2010b; Hernandez et al., 2011; Pritchard et al., 2010; Turchin et al., 2012). Another contributing factor may be that population movements over the last several thousand years have to some extent decoupled the geographic distributions of selected alleles from the geographic distributions of selective pressures. From the point of view of an individual allele, the movement of populations acts as a form of fluctuating selection; as populations move from environment to environment, the selection coefficient on an allele may change in both sign and magnitude (assuming a fixed selection coefficient in a given environment). This means that the environment that dominated the allele frequency trajectory may not be the environment that the allele is found in today.

Ancient DNA is also a potentially transformative tool for understanding human adaptation more generally. Nearly all methods for detecting positive selection at the genetic level are based around the principle that a selected allele changes frequency more quickly than a neutral locus. Tracking the trajectories of individuals over time allows direct access to this information, and makes it possible to infer precisely when in time (and with more certainty where geographically) genetic changes began to arise (Wilde et al., 2014). In fact, in the presence of admixture it is simple to create scenarios where the trajectory of the selected allele allows one to identify selection despite there being no net change in allele frequency (Figure 4B). To date, studies of selection over time have been limited either to small sample sizes (Lazaridis et al., 2013; Olalde et al., 2014) or small numbers of sites (Krüttli et al.,

2014; Malmström et al., 2010; Plantinga et al., 2012; Wilde et al., 2014). However, whole genome technologies should make it possible to interrogate many thousands of phenotypically-relevant variants simultaneously.
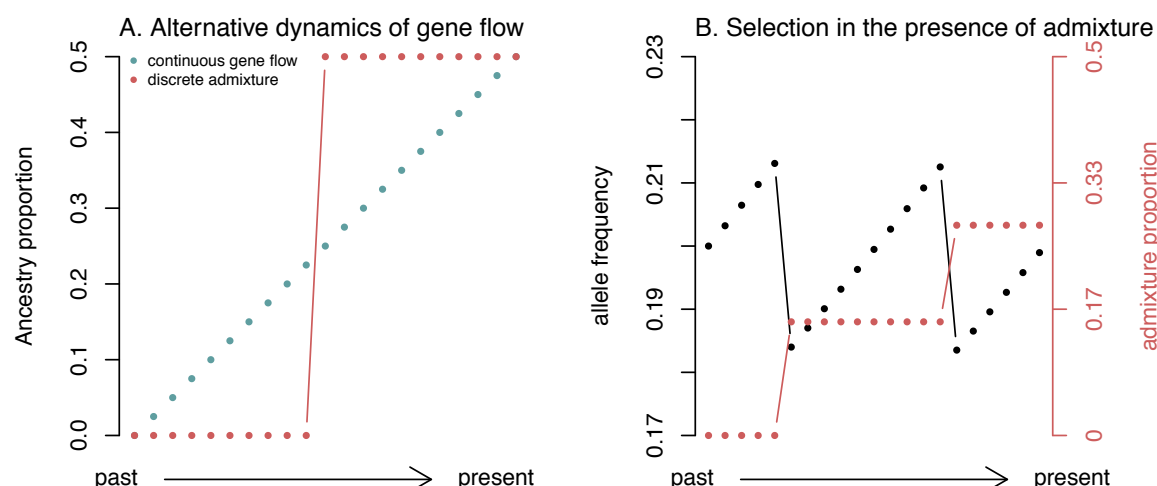

Figure 4: Time series allow tests of different models about human history. A. Different scenarios for the dynamics of gene flow. We show idealized time series of admixture proportions in a population under models of continuous gene flow and discrete admixture. B. Selection in the presence of admixture. We show idealized time series of the allele frequencies at a selected allele and admixture proportions in a population. Note that despite selection the allele does not show any net change in allele frequency.

Taming the wild west of ancient DNA

Realizing the potential of ancient DNA studies will require systematic approaches. However, ancient DNA studies today are often more spectacular than systematic.

The paradigmatic approach to ancient DNA research in the era of high-throughput sequencing involves identifying a "golden" archaeological sample that yields usable DNA and obtaining a complete or partial genome sequence from it (e.g. Green et al., 2010; Keller et al., 2012; Meyer et al., 2012; Prüfer et al., 2014; Raghavan et al., 2013; Rasmussen et al., 2010). To some extent, the discovery of analyzable samples has been the main driver of the scientific questions asked. However, this opportunistic approach will not be driving ancient DNA research in a few years' time. The future of ancient DNA research will involve a return to a more systematic approach: hypothesis-driven sampling across time and space, and analysis of much larger sample sizes.

Ancient DNA research has three major experimental challenges and two computational challenges that need to be jointly addressed by researchers who wish to access this transformative technology.

The first challenge in ancient DNA research is the danger of contamination from archaeologists or lab researchers who have handled the sample. This is a recurring

14

problem (Cooper and Poinar, 2000; Green et al., 2006, 2009; Hedges and Schweitzer, 1995; Pääbo, 1985; Pääbo et al., 2004; Wall and Kim, 2007; Willerslev et al., 2004). While experimental guidelines (e.g. Gilbert et al., 2005; Hofreiter et al., 2001; Pääbo et al., 2004) as well as laboratory methods for reducing the possibility of contamination (Green et al., 2010) and empirically assessing the authenticity of ancient DNA sequences (Sawyer et al., 2012) have greatly improved the situation, in practice this concern will never disappear. The importance of controlling for the possibility of contamination is the single most important reason why, to date, convincing ancient DNA research has been dominated by a small number of specialist laboratories with the expertise and facilities to control contamination. All laboratories that carry out this type of work will need to maintain the same level of vigilance currently maintained at specialist laboratories. To reduce contamination at the source, it is also important that archaeologists excavating new samples become trained in handing samples in ways that minimizes contamination. Valuable measures include wearing sterile gloves and a protective suit while excavating remains, not washing the remains, and immediately placing the remains in a sterile plastic bag and refrigerating them prior to shipment to an ancient DNA lab.

A second experimental challenge is the difficulty of identifying human remains that contain preserved DNA. This often requires screening dozens of individuals from multiple sites, with success rates in DNA extraction depending strongly on the conditions experienced by the sample: its age, temperature, humidity, acidity, the part of the body from which it derives, the speed with which it dried after death, whether or not it is from a sample that was rapidly defleshed, and likely other factors that are not currently understood. Modern methods have increased the efficiency of DNA extraction by capturing the short molecules that make up the great majority of any ancient sample (Dabney et al., 2013; Rohland, 2012). In addition, improved library preparation methods have increased the fraction of molecules in an extract that are amenable to sequencing (Fu et al., 2013; Meyer et al., 2012). Nevertheless, it is still the case that there is great variability in success. The secret of successful ancient DNA research is not luck, but hard work and years of experience: screening many carefully chosen and prepared samples until a subset are identified that perform well.

A third experimental challenge is the difficulty of finding a sample that has a sufficiently high proportion of DNA from the bone itself to be economically analyzed. Concretely, the challenge is that for many ancient DNA extracts, the proportion of endogenous DNA is very low, on the order of one percent or less. For example, a recent study on a ~40,000 year-old sample from the Tianyuan site of Northern China had an endogenous DNA proportion of 0.02% (Fu et al., 2013). For such samples, important as they are, it is not economical to simply carry out brute-force sequencing of random genomic fragments and expect to obtain sufficient coverage to make meaningful inferences.

The challenges of ancient DNA research do not end once the wet lab sample preparation is complete, as the datasets that are generated pose formidable computational and analytical challenges.

The first computational challenge is that once useable samples are obtained and sequenced, the data need to be processed. Ancient DNA laboratories, which traditionally have their strongest expertise in archaeology, physical anthropology, or biochemistry, often lack the bioinformatics expertise, data processing power and data storage solutions necessary to handle the millions or even billions of sequences that are generated by modern ancient DNA studies. Moreover, ancient DNA data also requires tailored bioinformatics tools for handling the short sequences that are characteristic of old samples (Kircher, 2012); one cannot simply use existing tools such as SAMtools (Li et al., 2009) or the Genome Analysis Toolkit (DePristo et al., 2011) with the default settings. For example, the sequenced DNA fragments are usually short and degraded, and are expected to have C→T and G→A errors at the ends of the molecules due to cytosine deamination. The data additionally need to be computationally assessed for evidence of contamination, for example by checking the rate of molecules that do not map to the mitochondrial DNA consensus sequence obtained for the sample (Fu et al., 2013). If a sample is determined to be contaminated, the characteristic ancient DNA errors can be leveraged to reduce the level of contamination (assuming that the contamination is not old) by restricting to sequences that contain typical ancient DNA degradations (Meyer et al., 2013; Skoglund et al., 2014).

A second computational challenge for ancient DNA research is that once the data are processed and a sample is determined as likely to be authentic, the data also need to be analyzed using statistical methods that infer population history. While many methods that have been developed to make inferences about population relationships are not biased by the fact that ancient DNA is old and error-prone (Green et al., 2010; Reich et al., 2010; Meyer et al., 2012; Patterson et al., 2012; Prüfer et al., 2014), these methods must be implemented carefully to produce meaningful results. In addition, the methods need to handle complications of ancient DNA, such as the fact that there is rarely deep sequencing data, so that unambiguous determination of genotypes at each position in the genome is often unreliable.

At present, few laboratories have the experimental and computational expertise to address all these challenges simultaneously. As a result, ancient DNA research has been dominated by a few vertically integrated and well-funded laboratories that combine all these skills under the same roof. The high barriers to entry have meant that the full potential of ancient DNA has been largely untapped. In what follows, we sketch out a way forward that we expect will make ancient DNA analysis more accessible to the broad research community, including to archaeologists.

Democratization of ancient DNA technology

16

The usual paradigm in ancient DNA analysis of the nuclear genome has been to identify a sample that has a high enough proportion of DNA to deeply sequence. Such golden samples are rare, and are typically identified only after laborious screening of many dozens of samples that have low proportions of human DNA. Once a sample is found that has an appreciable proportion of human DNA, it is typically sequenced to as high a coverage as possible (sometimes the limited number of starting molecules in many ancient DNA libraries is the main factor limiting sequencing depth). All of this is very expensive, and has been an important barrier to entry into ancient DNA work for less well-funded laboratories.

Whole genome sequences, however, are not required for most historical inferences. In a paper that was important not only for what it showed about history but also for what it showed about the quantity of information that can be extracted from small amounts of data, Skoglund et al., (2012) showed that even low levels of genome coverage per sample – 1–5% of genomic bases covered in their case – was sufficient to support profound historical inferences.

Following the example of Skoglund et al. (2012), we propose that a way to make ancient DNA analysis accessible to a much larger number of laboratories is to use targeted capture approaches that enrich a sample for human DNA. Enrichment strategies have already been shown to be extremely effective for ancient DNA analysis. In one approach to target enrichment, Carpenter et al., (2013) showed that it is possible to take RNA baits obtained by transcribing a whole human genome, and hybridize them in solution to an ancient DNA library, thus increasing the fraction of human DNA to be sequenced by several orders of magnitude. Fu et al., (2013) showed that it is possible to synthesize oligonucleotide DNA baits targeted at a specified subset of the genome—the entire mitochondrial DNA sequence, or the coding sequences of all genes—and to hybridize to ancient DNA libraries in solution to enrich a library for molecules from the targeted subset of the genome. This strategy has been used to obtain a high quality mitochondrial genome sequence from a ~400,000 year-old archaic human (Meyer et al., 2013), as well as approximately 2-fold redundant coverage of chromosome 21 from the ~40,000 year old Tianyuan sample from northern China (Fu et al., 2013).

We are particularly enthusiastic about the possibility of adapting a technology like that of Fu et al., (2013) to enrich human samples for panels of several hundred thousand single nucleotide polymorphisms (SNPs) that have already been genotyped on large panels of present-day samples. This is a sufficient number of SNPs that it would allow for high-resolution analysis of how an ancient sample relates to present-day as well as other ancient samples. The strategy has two major advantages. First, through enrichment, it allows analysis of samples with much less than 10% human DNA that are not economical for whole genome sequencing studies. Second, it requires about two orders of magnitude less sequencing per sample to saturate all its targets (Table 1). We caution that there are some questions—for example, estimation of population divergence times—that rely on identification of sample-specific mutations and may be better addressed with whole

17

genome sequencing than by a SNP capture experiment. Nevertheless, we believe that a capture experiment can answer the substantial majority of questions related to population history or natural selection that are addressable with genetic data, while allowing larger numbers of samples to be analyzed for the same cost.

Table 1: Comparison of two strategies for ancient DNA studies of history

	Whole genome sequencing	Capture and sequencing of 300,000 SNPs
Number of samples that need to be screened*	1	1/10
Amount of sequencing that needs to be performed*	1	1/100
High resolution inference of pop. relationships	Yes	Yes
Allows studies of selection	Yes	Yes
Variants specific to sample can be discovered	Yes	No
Works for samples with low % of human DNA	No	Yes

* These are rough qualitative estimates based on discussion with colleagues. The estimated amount of work required for capture 300,000 SNPs is expressed as a fraction of that required for whole-genome sequencing.

An important goal for the coming years should be to make DNA a tool that will become fully accessible not just to smaller labs but also to archaeologists. In this regard, it is important to encourage interaction and collaboration between the genetic and archaeologist communities to develop standards for the interpretation and use of genetic data in answering questions relevant to archaeology.

Members of the archaeologist community are already sophisticated users of other scientific technologies for analyzing ancient biological remains such as carbon 14 analysis (for date estimation) and stable isotope analysis (for making inferences about diet). Currently, when archaeologists have a sample that they wish to analyze, they send it to specialist (sometimes commercial) laboratories, which then provide a carbon date and/or an isotope analysis, along with a report giving interpretation. We envision a future in which ancient DNA analysis services could be provided in a similar way. Concretely, archaeologists would be able to send skeletal material to a specialist laboratory for analysis, where it can be tested for ancient DNA using up-to-date protocols. If ancient DNA is detected, a whole mitochondrial genome sequence can be produced and a sex can be determined. For ancient samples that produce uncontaminated DNA, whole genome data can be produced.

To make such a future possible, it will be necessary to build up a database of present-day human and ancient DNA samples all genotyped at the same set of SNPs to which any new sample can be compared. In addition, it will be necessary to write software to automatically compare a new sample to samples in the database (Alexander et al., 2009; Patterson et al., 2012; Price et al., 2006; Raj et al., 2013), which will be useful for producing a report for archaeologists on how the sample relates to diverse present-day humans, and to other ancient samples. Because of the subtleties of interpreting genetic data, it will important be necessary for

18

archaeologists to work collaboratively with population geneticists in interpreting the results of such studies. Such a report would, however, be a useful tool for giving archaeologists a first impression of the biological ancestry of their samples.

We particularly wish to highlight the potential of ancient DNA as a direct tool for archaeological analysis, which will be useful for elucidating the population structure and patterns of relatedness within a particular archaeological site. If many individuals from an archaeological culture can be analyzed in a way that correlates the genetic information—sex, patterns of relatedness, and ethnic affiliation—to archaeological information derived from grave goods, status symbols, and other objects, it seems likely to provide new insights. A particular area of opportunity is the identification of outlier individuals. Genetic data should make it easy to detect individuals that are outliers in terms of their ancestry relative to others at nearby sites, which could be informative about migration rates and sex biases in these rates.

Conclusions

We have argued that it will likely to be fruitful to re-examine many aspects of human population history and the history of natural selection from a perspective in which population movements play a central role. We have further argued that advances in ancient DNA are likely to have an important impact on how we think about using genetic data to address questions about human history.

We conclude by presenting three questions that we believe can be answered in the coming years by genetic studies of human history and variation; this list is not comprehensive, and many other important questions surely can also be addressed:

1. What is the genetic legacy of the inhabitants of eastern Africa before the arrival of Bantu speakers? The population genetic structure of sub-Saharan Africa has been transformed by the expansion of Bantu-speaking agriculturalists over the last 3,000 years. As a concrete example, the population structure of eastern Africa prior to this expansion remains controversial (Morris, 2003; Schepartz, 1988). Ancient DNA from samples prior to the Bantu expansion could settle this debate. While the climate in much, but not all, of sub-Saharan Africa is less favorable for preservation of DNA than that of northern Eurasia, it is not clear whether enough African samples have been tested to determine whether or not ancient DNA analysis works. Moreover, the time scale of the Bantu expansion is only several thousand years rather than tens of thousands of years. In light of continuing technological progress in ancient DNA extraction (including the successful extraction of DNA from a ~400,000 year old sample in Spain (Meyer et al., 2013)), we are hopeful that ancient DNA studies of some African samples within the last few thousand years may become possible.

2. When were the geographic patterns at loci under positive selection established? Loci under positive selection in humans have allele frequencies with characteristic geographic patterns, which have been interpreted as "west

Eurasian", "east Asian" and "non-African" selective sweeps (Coop et al., 2009). Were these patterns established tens of thousands of years ago during the initial human peopling of the globe by anatomically modern humans? Or were they established after later population movements? Resolving these questions will become possible once substantial numbers of ancient samples are genotyped. Work in this area is only just beginning (Wilde et al., 2014).

3. Was the spread of Indo-European languages caused by large-scale migrations of people or did language shifts occur without extensive population replacement? For times prior to the invention of writing, it will never be possible to directly relate the language that people spoke to their remains. Nevertheless, genetic data may still informative about the historical events that accompany linguistic expansions. The origins of the Indo-European languages—spoken across Europe and West, Central, and South Asia—is a particularly important question (Anthony, 2007; Mallory, 1997). These languages likely spread across Eurasia and diversified within the last 5-10 thousand years. Was this spread of languages caused by a large-scale movement of people? It now may be possible to obtain ancient DNA from cultures known to have spoken Indo-European languages—for example, Hittites and Tocharians—and to compare these populations with their neighbors to determine whether they harbor a genetic signature specific to the Indo-European speakers. It may also be possible to search for the spread of (currently hypothetical) Indo-European genetic signatures in Europe and India at the times when various hypotheses have suggested that they may have arrived.

In conclusion, genetic studies of the human past are at the beginning of an era in which genome-wide studies of modern and ancient DNA samples will make it possible to ask and answer questions about human history and natural selection that could not be addressed before. We expect that these approaches will provide new insights into important and still controversial issues.

Acknowledgements

References

Abecasis, G., Altshuler, D., Auton, A., Brooks, L., Durbin, R., Gibbs, R.A., Hurles, M.E., McVean, G.A., Bentley, D., Chakravarti, A., et al. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res *19*, 1655–1664.

Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Bonnen, P.E., de Bakker, P., Deloukas, P., Gabriel, S.B., et al. (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467*, 52.

Alves, I., Hanulová, A. \vSrámková, Foll, M., and Excoffier, L. (2012). Genomic data reveal a complex making of humans. PLoS Genetics *8*, e1002837.

Amos, W., and Hoffman, J.I. (2010). Evidence that two main bottleneck events shaped modern human genetic diversity. Proc Biol Sci *277*, 131–137.

Anthony, D.W. (2007). The horse, the wheel, and language: how bronze-age riders from the Eurasian steppes shaped the modern world (Princeton, N.J.: Princeton University Press).

Ashraf, Q., and Galor, O. (2013). The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development. The American Economic Review *103*.

Atkinson, Q.D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. Science *332*, 346–349.

Ballard, J.W.O., and Whitlock, M.C. (2004). The incomplete natural history of mitochondria. Mol Ecol *13*, 729–744.

Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., et al. (2010). The genome-wide structure of the Jewish people. Nature *466*, 238–242.

Behar, D.M., Harmant, C., Manry, J., van Oven, M., Haak, W., Martinez-Cruz, B., Salaberria, J., Oyharçabal, B., Bauduer, F., Comas, D., et al. (2012). The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times. Am J Hum Genet *90*, 486–493.

Betti, L., Cramon-Taubadel, N.V., and Lycett, S.J. (2012). Human pelvis and long bones reveal differential preservation of ancient population history and migration out of Africa. Hum. Biol. *84*, 139–152.

Bramanti, B., Thomas, M.G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M.N., Jankauskas, R., Kind, C.-J., et al. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science *326*, 137–140.

Brandt, G., Haak, W., Adler, C.J., Roth, C., Szécsényi-Nagy, A., Karimnia, S., Möller-Rieker, S., Meller, H., Ganslmeier, R., Friederich, S., et al. (2013). Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. Science *342*, 257–261.

Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., Hammer, M., Bustamante, C.D., and Ostrer, H. (2010). Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. Proc Natl Acad Sci U S A *107 Suppl 2*, 8954–8961.

Carpenter, M.L., Buenrostro, J.D., Valdiosera, C., Schroeder, H., Allentoft, M.E., Sikora, M., Rasmussen, M., Gravel, S., Guillen, S., Nekhrizov, G., et al. (2013). Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. The American Journal of Human Genetics *93*, 852–864.

Cavalli-Sforza, L.L.L., Menozzi, P., and Piazza, A. (1994). The history and geography of human genes (Princeton university press).

Colonna, V., Pagani, L., Xue, Y., and Tyler-Smith, C. (2011). A world in a grain of sand: human history from genetic data. Genome Biol *12*, 234.

Coop, G., Pickrell, J.K., Novembre, J., Kudaravalli, S., Li, J., Absher, D., Myers, R.M., Cavalli-Sforza, L.L., Feldman, M.W., and Pritchard, J.K. (2009). The role of geography in human adaptation. PLoS Genet *5*, e1000500.

Coop, G., Witonsky, D., Di Rienzo, A., and Pritchard, J.K. (2010). Using environmental correlations to identify loci underlying local adaptation. Genetics *185*, 1411–1423.

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

Cooper, A., and Stringer, C.B. (2013). Paleontology. Did the Denisovans cross Wallace's Line? Science *342*, 321–323.

Von Cramon-Taubadel, N., and Lycett, S.J. (2008). Brief communication: human cranial variation fits iterative founder effect model with African origin. Am J Phys Anthropol *136*, 108–113.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., et al. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. U.S.A. *110*, 15758–15763.

22

Davies, N. (1998). Europe: a history (New York: HarperPerennial).

DeGiorgio, M., Jakobsson, M., and Rosenberg, N.A. (2009). Out of Africa: modern human origins special feature: explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. Proc Natl Acad Sci U S A *106*, 16057–16062.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics *43*, 491–498.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., and Foll, M. (2013). Robust demographic inference from genomic and SNP data. PLoS Genet *9*, e1003905.

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proc Natl Acad Sci U S A *110*, 2223–2227.

Gilbert, M.T.P., Bandelt, H.-J., Hofreiter, M., and Barnes, I. (2005). Assessing ancient DNA studies. Trends Ecol. Evol. (Amst.) *20*, 541–544.

Granka, J.M., Henn, B.M., Gignoux, C.R., Kidd, J.M., Bustamante, C.D., and Feldman, M.W. (2012). Limited evidence for classic selective sweeps in African populations. Genetics *192*, 1049–1064.

Gravel, S., Zakharia, F., Byrnes, J.K., Muzzio, M., Moreno-Estrada, A., Rodriguez-Flores, J.L., Kenny, E.E., Gignoux, C.R., Maples, B.K., Guiblet, W., et al. (2013). Reconstructing Native American Migrations from Whole-genome and Whole-exome Data. arXiv Preprint arXiv:1306.4021.

Green, R.E., Krause, J., Ptak, S.E., Briggs, A.W., Ronan, M.T., Simons, J.F., Du, L., Egholm, M., Rothberg, J.M., Paunovic, M., et al. (2006). Analysis of one million base pairs of Neanderthal DNA. Nature *444*, 330–336.

Green, R.E., Briggs, A.W., Krause, J., Prüfer, K., Burbano, H.A., Siebauer, M., Lachmann, M., and Pääbo, S. (2009). The Neandertal genome and ancient DNA authenticity. The EMBO Journal *28*, 2494–2502.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., et al. (2010). A draft sequence of the Neandertal genome. Science *328*, 710–722.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., and Bustamante, C.D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet *5*, e1000695.

Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K.W., et al. (2005). Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. Science *310*, 1016–1018.

Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., Der Sarkissian, C.S.I., Brandt, G., Schwarz, C., Nicklisch, N., et al. (2010). Ancient DNA from European early neolithic farmers reveals their near eastern affinities. PLoS Biol *8*, e1000536.

Hancock, A.M., Witonsky, D.B., Ehler, E., Alkorta-Aranburu, G., Beall, C., Gebremedhin, A., Sukernik, R., Utermann, G., Pritchard, J., Coop, G., et al. (2010a). Colloquium paper: human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. Proc Natl Acad Sci U S A *107 Suppl 2*, 8924–8930.

Hancock, A.M., Alkorta-Aranburu, G., Witonsky, D.B., and Di Rienzo, A. (2010b). Adaptations to new environments in humans: the role of subtle allele frequency shifts. Philos Trans R Soc Lond B Biol Sci *365*, 2459–2468.

Hanihara, T. (2008). Morphological variation of major human populations based on nonmetric dental traits. Am. J. Phys. Anthropol. *136*, 169–182.

Harpending, H., and Rogers, A. (2000). Genetic perspectives on human origins and differentiation. Annu Rev Genomics Hum Genet *1*, 361–385.

Harpending, H.C., and Eller, E. (2000). Human diversity and its history. In The Biology of Biodiversity, (Springer), pp. 301–314.

Hedges, S.B., and Schweitzer, M.H. (1995). Detecting dinosaur DNA. Science *268*, 1191–1192; author reply 1194.

Hellenthal, G., Auton, A., and Falush, D. (2008). Inferring human colonization history using a copying model. PLoS Genet *4*, e1000078.

Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., and Myers, S. (2014). A genetic atlas of human admixture history. Science *343*, 747–751.

Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A., et al. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc Natl Acad Sci U S A *108*, 5154–5162.

Henn, B.M., Cavalli-Sforza, L.L., and Feldman, M.W. (2012a). The great human expansion. Proc Natl Acad Sci U S A *109*, 17758–17764.

Henn, B.M., Botigué, L.R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J.K., Fadhlaoui-Zid, K., Zalloua, P.A., Moreno-Estrada, A., Bertranpetit, J., et al. (2012b). Genomic

24

Ancestry of North Africans Supports Back-to-Africa Migrations. PLoS Genet *8*, e1002397.

Hernandez, R.D., Kelley, J.L., Elyashiv, E., Melton, S.C., Auton, A., McVean, G., Project, 1000 Genomes, Sella, G., and Przeworski, M. (2011). Classic selective sweeps were rare in recent human evolution. Science *331*, 920–924.

Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., and Pääbo, S. (2001). Ancient DNA. Nat. Rev. Genet. *2*, 353–359.

Hudson, R.R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics *18*, 337–338.

Huerta-Sánchez, E., Degiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., Cardona, A., Montgomery, H.E., Cavalleri, G.L., Robbins, P.A., et al. (2013). Genetic signatures reveal high-altitude adaptation in a set of ethiopian populations. Mol Biol Evol *30*, 1877–1888.

Jablonski, N.G., and Chaplin, G. (2000). The evolution of human skin coloration. J Hum Evol *39*, 57–106.

Johnson, N.A., Coram, M.A., Shriver, M.D., Romieu, I., Barsh, G.S., London, S.J., and Tang, H. (2011). Ancestral components of admixed genomes in a Mexican cohort. PLoS Genet *7*, e1002410.

Keller, A., Graefen, A., Ball, M., Matzas, M., Boisguerin, V., Maixner, F., Leidinger, P., Backes, C., Khairat, R., Forster, M., et al. (2012). New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nat Commun *3*, 698.

Kircher, M. (2012). Analysis of high-throughput ancient DNA sequencing data. Methods Mol. Biol. *840*, 197–228.

Krause, J., Fu, Q., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P., and Pääbo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature *464*, 894–897.

Krüttli, A., Bouwman, A., Akgül, G., Della Casa, P., Rühli, F., and Warinner, C. (2014). Ancient DNA analysis reveals high frequency of European lactase persistence allele (T-13910) in medieval central europe. PLoS ONE *9*, e86251.

Kwiatkowski, D.P. (2005). How malaria has affected the human genome and what human genetics can teach us about malaria. Am. J. Hum. Genet. *77*, 171–192.

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Sudmant, P.H., Schraiber, J.G., Castellano, S., Kirsanow, K., Economou, C., et al. (2013). Ancient

human genomes suggest three ancestral populations for present-day Europeans. arXiv Preprint arXiv:1312.6639.

Lell, J.T., Sukernik, R.I., Starikovskaya, Y.B., Su, B., Jin, L., Schurr, T.G., Underhill, P.A., and Wallace, D.C. (2002). The dual origin and Siberian affinities of Native American Y chromosomes. Am J Hum Genet *70*, 192–206.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., et al. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science *319*, 1100–1104.

Lipson, M., Loh, P.-R., Levin, A., Reich, D., Patterson, N., and Berger, B. (2013). Efficient moment-based inference of admixture parameters and sources of gene flow. Mol Biol Evol *30*, 1788–1802.

Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. Genetics *193*, 1233–1254.

Mallory, J. (1997). The homelands of the Indo-Europeans. In Archaeology and Language, R. Blench, and M. Spriggs, eds. (London ; New York: Routledge),.

Malmström, H., Gilbert, M.T.P., Thomas, M.G., Brandström, M., Stor\aa, J., Molnar, P., Andersen, P.K., Bendixen, C., Holmlund, G., Götherström, A., et al. (2009). Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. Curr Biol *19*, 1758–1762.

Malmström, H., Linderholm, A., Lidén, K., Storå, J., Molnar, P., Holmlund, G., Jakobsson, M., and Götherström, A. (2010). High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. BMC Evol. Biol. *10*, 89.

Manica, A., Amos, W., Balloux, F., and Hanihara, T. (2007). The effect of ancient population bottlenecks on human phenotypic variation. Nature *448*, 346–348.

McEvoy, B.P., Lind, J.M., Wang, E.T., Moyzis, R.K., Visscher, P.M., van Holst Pellekaan, S.M., and Wilton, A.N. (2010). Whole-genome genetic diversity in a sample of Australians with deep Aboriginal ancestry. Am. J. Hum. Genet. *87*, 297–305.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science.

26

Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J.-L., Martínez, I., Gracia, A., de Castro, J.M.B., Carbonell, E., et al. (2013). A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature.

Moorjani, P., Patterson, N., Hirschhorn, J.N., Keinan, A., Hao, L., Atzmon, G., Burns, E., Ostrer, H., Price, A.L., and Reich, D. (2011). The history of African gene flow into Southern Europeans, Levantines, and Jews. PLoS Genet *7*, e1001373.

Moorjani, P., Thangaraj, K., Patterson, N., Lipson, M., Loh, P.-R., Govindaraj, P., Berger, B., Reich, D., and Singh, L. (2013). Genetic evidence for recent population mixture in India. Am J Hum Genet *93*, 422–438.

Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J.L., Byrnes, J.K., Gignoux, C.R., Ortiz-Tello, P.A., Martínez, R.J., Hedges, D.J., Morris, R.W., et al. (2013). Reconstructing the population genetic history of the Caribbean. PLoS Genet *9*, e1003925.

Morris, A.G. (2003). The myth of the East African'Bushmen'. The South African Archaeological Bulletin 85–90.

Norton, H.L., Kittles, R.A., Parra, E., McKeigue, P., Mao, X., Cheng, K., Canfield, V.A., Bradley, D.G., McEvoy, B., and Shriver, M.D. (2007). Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Mol Biol Evol *24*, 710–722.

Novembre, J., and Ramachandran, S. (2011). Perspectives on human population structure at the cusp of the sequencing era. Annu Rev Genomics Hum Genet *12*, 245–274.

Olalde, I., Allentoft, M.E., Sánchez-Quinto, F., Santpere, G., Chiang, C.W.K., Degiorgio, M., Prado-Martinez, J., Rodríguez, J.A., Rasmussen, S., Quilez, J., et al. (2014). Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature.

Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. Nature *314*, 644–645.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., and Hofreiter, M. (2004). Genetic analyses from ancient DNA. Annu. Rev. Genet. *38*, 645–679.

Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S.Q., Thomas, M.G., Luiselli, D., et al. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am J Hum Genet *91*, 83–96.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.

Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet *8*, e1002967.

Pickrell, J.K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., Kure, B., Mpoloka, S.W., Nakagawa, H., Naumann, C., et al. (2012). The genetic prehistory of southern Africa. Nat Commun *3*, 1143.

Pickrell, J.K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., and Reich, D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. Proc. Natl. Acad. Sci. U.S.A. *111*, 2632–2637.

Pierron, D., Razafindrazaka, H., Pagani, L., Ricaut, F.-X., Antao, T., Capredon, M., Sambo, C., Radimilahy, C., Rakotoarisoa, J.-A., Blench, R.M., et al. (2014). Genome-wide evidence of Austronesian-Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. Proc. Natl. Acad. Sci. U.S.A. *111*, 936–941.

Pinhasi, R., Thomas, M.G., Hofreiter, M., Currat, M., and Burger, J. (2012). The genetic history of Europeans. Trends Genet *28*, 496–505.

Plagnol, V., and Wall, J.D. (2006). Possible ancestral structure in human populations. PLoS Genet. *2*, e105.

Plantinga, T.S., Alonso, S., Izagirre, N., Hervella, M., Fregel, R., van der Meer, J.W.M., Netea, M.G., and de la Rúa, C. (2012). Low prevalence of lactase persistence in Neolithic South-West Europe. Eur. J. Hum. Genet. *20*, 778–782.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet *38*, 904–909.

Price, A.L., Patterson, N., Yu, F., Cox, D.R., Waliszewska, A., McDonald, G.J., Tandon, A., Schirmer, C., Neubauer, J., Bedoya, G., et al. (2007). A Genomewide Admixture Map for Latino Populations. The American Journal of Human Genetics *80*, 1024–1036.

Pritchard, J.K., Pickrell, J.K., and Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol *20*, R208–15.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43–49.

Prugnolle, F., Manica, A., and Balloux, F. (2005). Geography predicts neutral genetic diversity of human populations. Curr Biol *15*, R159–60.

Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford Jr, T.W., Orlando, L., Metspalu, E., et al. (2013). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature.

Raj, A., Stephens, M., and Pritchard, J.K. (2013). Variational Inference of Population Structure in Large SNP Datasets. bioRxiv.

Ralph, P., and Coop, G. (2013). The geography of recent genetic ancestry across Europe. PLoS Biol *11*, e1001555.

Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-Sforza, L.L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proceedings of the National Academy of Sciences of the United States of America *102*, 15942.

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature *463*, 757–762.

Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., et al. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. Science *334*, 94–98.

Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., DeGiorgio, M., Stafford, T.W., Jr, Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S.M., et al. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature *506*, 225–229.

Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian population history. Nature *461*, 489–494.

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature *468*, 1053–1060.

Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M.R., Pugach, I., Ko, A.M.-S., Ko, Y.-C., Jinam, T.A., Phipps, M.E., et al. (2011). Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. Am J Hum Genet *89*, 516–528.

Renfrew, C., and Bahn, P.G. (1996). Archaeology: theories, methods, and practice (London: Thames and Hudson).

Rogers, D.S., Feldman, M.W., and Ehrlich, P.R. (2009). Inferring population histories using cultural data. Proc. Biol. Sci. *276*, 3835–3843.

29

Rohland, N. (2012). DNA extraction of ancient animal hard tissue samples via adsorption to silica particles. Methods Mol. Biol. *840*, 21–28.

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., and Feldman, M.W. (2002). Genetic structure of human populations. Science *298*, 2381–2385.

Sankararaman, S., Patterson, N., Li, H., Pääbo, S., and Reich, D. (2012). The date of interbreeding between Neandertals and modern humans. PLoS Genet. *8*, e1002947.

Santos, F.R., Pandya, A., Tyler-Smith, C., Pena, S.D., Schanfield, M., Leonard, W.R., Osipova, L., Crawford, M.H., and Mitchell, R.J. (1999). The central Siberian origin for native American Y chromosomes. Am J Hum Genet *64*, 619–628.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS ONE *7*, e34131.

Schepartz, L.A. (1988). Who were the latter Pleistocene eastern Africans? African Archaeological Review *6*, 57–72.

Schlebusch, C.M., Skoglund, P., Sjödin, P., Gattepaille, L.M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A., Blum, M.G.B., et al. (2012). Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science *338*, 374–379.

Simonson, T.S., Yang, Y., Huff, C.D., Yun, H., Qin, G., Witherspoon, D.J., Bai, Z., Lorenzo, F.R., Xing, J., Jorde, L.B., et al. (2010). Genetic evidence for high-altitude adaptation in Tibet. Science *329*, 72–75.

Skoglund, P., Malmström, H., Raghavan, M., Stor\aa, J., Hall, P., Willerslev, E., Gilbert, M.T.P., Götherström, A., and Jakobsson, M. (2012). Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. Science *336*, 466–469.

Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. U.S.A. *111*, 2229–2234.

Starikovskaya, E.B., Sukernik, R.I., Derbeneva, O.A., Volodko, N.V., Ruiz-Pesini, E., Torroni, A., Brown, M.D., Lott, M.T., Hosseini, S.H., Huoponen, K., et al. (2005). Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. Ann Hum Genet *69*, 67–89.

Stoneking, M., and Krause, J. (2011). Learning about human population history from ancient and modern genomes. Nat Rev Genet *12*, 603–614.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.-M., Doumbo, O., et al. (2009). The genetic structure and history of Africans and African Americans. Science *324*, 1035–1044.

Turchin, M.C., Chiang, C.W.K., Palmer, C.D., Sankararaman, S., Reich, D., Consortium, G.I. of An.T. (GIANT), and Hirschhorn, J.N. (2012). Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nat Genet *44*, 1015–1019.

Veeramah, K.R., and Hammer, M.F. (2014). The impact of whole-genome sequencing on the reconstruction of human population history. Nature Reviews Genetics *15*, 149–162.

Wall, J.D., and Kim, S.K. (2007). Inconsistencies in Neanderthal genomic DNA sequences. PLoS Genet. *3*, 1862–1866.

Wall, J.D., and Slatkin, M. (2012). Paleopopulation genetics. Annu Rev Genet *46*, 635–649.

Wang, S., Lewis, C.M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., Rojas, W., Parra, M.V., Molina, J.A., Gallo, C., et al. (2007). Genetic variation and population structure in native Americans. PLoS Genet *3*, e185.

Wilde, S., Timpson, A., Kirsanow, K., Kaiser, E., Kayser, M., Unterländer, M., Hollfelder, N., Potekhina, I.D., Schier, W., Thomas, M.G., et al. (2014). Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. Proc. Natl. Acad. Sci. U.S.A.

Willerslev, E., Hansen, A.J., and Poinar, H.N. (2004). Isolation of nucleic acids and cultures from fossil ice and permafrost. Trends Ecol. Evol. (Amst.) *19*, 141–147.