Journal of Human Evolution 79 (2015) 64–72

Contents lists available at ScienceDirect

Journal of Human Evolution

journal homepage: www.elsevier.com/locate/jhevol

The utility of ancient human DNA for improving allele age estimates, with implications for demographic models and tests of natural selection

Aaron J. Sams [a, b, *], John Hawks [a, 1], Alon Keinan [b, 1]

[a] Department of Anthropology, University of Wisconsin-Madison, Madison, WI, USA
[b] Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY, USA

ARTICLE INFO

Article history:
Received 14 January 2014
Accepted 22 October 2014
Available online 15 November 2014

Keywords:
Ötzi
Neandertals
Europe

ABSTRACT

The age of polymorphic alleles in humans is often estimated from population genetic patterns in extant human populations, such as allele frequencies, linkage disequilibrium, and rate of mutations. Ancient DNA can improve the accuracy of such estimates, as well as facilitate testing the validity of demographic models underlying many population genetic methods. Specifically, the presence of an allele in a genome derived from an ancient sample testifies that the allele is at least as old as that sample. In this study, we consider a common method for estimating allele age based on allele frequency as applied to variants from the US National Institutes of Health (NIH) Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project. We view these estimates in the context of the presence or absence of each allele in the genomes of the 5300 year old Tyrolean Iceman, Ötzi, and of the 50,000 year old Altai Neandertal. Our results illuminate the accuracy of these estimates and their sensitivity to demographic events that were not included in the model underlying age estimation. Specifically, allele presence in the Iceman genome provides a good fit of allele age estimates to the expectation based on the age of that specimen. The equivalent based on the Neandertal genome leads to a poorer fit. This is likely due in part to the older age of the Neandertal and the older time of the split between modern humans and Neandertals, but also due to gene flow from Neandertals to modern humans not being considered in the underlying demographic model. Thus, the incorporation of ancient DNA can improve allele age estimation, demographic modeling, and tests of natural selection. Our results also point to the importance of considering a more diverse set of ancient samples for understanding the geographic and temporal range of individual alleles.

Introduction

The quality of ancient human DNA (aDNA) sequencing has steadily improved during the past decade, culminating recently in the high-coverage sequencing of two archaic hominin genomes from human skeletal remains pre-dating 40,000 years ago (Meyer et al., 2012; Prüfer et al., 2014). Additionally, Meyer et al. (2013) recently sequenced a mitochondrial genome from the Middle Pleistocene site of Sima de los Huesos in Spain, pushing the oldest sequenced human DNA beyond 300,000 years. In addition to providing information about human demographic history, the growing sample of aDNA is useful for understanding the age of mutations that segregate in extant populations and, therefore, the timing of natural selection that has shaped present-day human populations. Herein, we illustrate the importance of aDNA for addressing questions of allele age and timing of selection by first briefly reviewing related literature, and by analyzing allele age in the context of two ancient genomes from different time periods and different phylogenetic distance to present-day Europeans. By comparing with allele age estimates based on allele frequency in an extant population, we conclude that consideration of whether an allele is present or absent in aDNA can provide important information about allele age and improve on the former type of estimates.

Presence of recently selected alleles in ancient European specimens

A straightforward means of testing selective hypotheses, e.g., based on long-range haplotype methods (Sabeti et al., 2002; Voight et al., 2006), is to analyze putatively selected haplotypes and

* Corresponding author.
E-mail address: as2847@cornell.edu (A.J. Sams).
[1] Co-project leaders.

http://dx.doi.org/10.1016/j.jhevol.2014.10.009
0047-2484/

single-nucleotide variants (SNVs) in ancient human specimens. Perhaps the most successful application to date of aDNA to a potential case of recent natural selection pertains to the genetic variants underlying lactase persistence. The ability to digest lactose, a disaccharide found in milk, typically slows in most mammals around the time of weaning. In many humans the ability to produce the enzyme lactase, which digests lactose into glucose and galactose sugars, continues into adulthood, a phenotype known as 'lactase persistence'. The timing and evolution of lactase persistence was documented first in Europeans, where the genomic region surrounding *LCT*, the gene encoding lactase, has been consistently reported as one of the most extreme long-range haplotype based examples of recent selection in Europe (Enattah et al., 2002, 2007, 2008; Bersaglieri et al., 2004; Tishkoff et al., 2006). The selection has been shown to be in the nearby *MCM6* gene and results in downregulation of the cessation of lactase production after weaning (Enattah et al., 2002). Similar disruptive changes to the *MCM6* gene have convergently evolved in both African (Tishkoff et al., 2006) and Middle Eastern (Enattah et al., 2007) populations. Selection for lactase persistence shows the importance of comparing genetic data to known cultural changes in the past, such as the timing and geographic distribution of cattle and camel pastoralism and milk consumption (Holden and Mace, 1997; Gerbault et al., 2009). The age of the mutation and subsequent beginning of the selective sweep underlying lactase persistence in Europeans (C/T-13910) has been estimated between 3000 and 12,000 years, which seems to coincide with the presence of domesticated cattle (Bollongino et al., 2006) and a record of increasing pastoralism and dairying in several human populations, particularly in northern Europe. For example, Tishkoff and colleagues (2006) estimated the age, using a coalescent simulation model that incorporated selection and recombination, at approximately 8000 to 9000 years depending on the degree of dominance assumed for the allele. While consistent with the anthropological record, the confidence intervals spanning 2000 to 19,000 years points to the large uncertainty in the estimates. This is consistent with the large range of variation in coalescence times (Slatkin and Rannala, 2000).

Estimates of allele age and timing of selection based on population genetic patterns observed in extant humans depend heavily on assumptions about the demographic history of human populations and are often associated with large ranges of error (as illustrated above for the timing of C/T-13910). By testing whether specific genetic variants were absent or present in an ancient sample, aDNA can be used to test hypotheses about the timing of selective changes in past human populations (Burger et al., 2007; Malmström et al., 2010; Plantinga et al., 2012). This can lead to much more precise time estimates, though these depend on the ability to accurately date ancient skeletal materials. For example, the derived allele (-13910*T) that underlies lactase persistence in Northern Europeans was found in only one copy out of 20 (~5%) in a 5000 year old skeletal sample from Sweden (Malmström et al., 2010), in ~27% of a sample of 26 Basque individuals dating between 4500 and 5000 years ago (Plantinga et al., 2012), and was completely absent from a skeletal sample of nine individuals from eastern Europe dating between 5000 and 5800 years ago (Burger et al., 2007).

Holocene demography of Europe and ancient DNA

Archaeological evidence suggests that the transition from a hunting and gathering lifestyle to a more sedentary agricultural 'Neolithic' lifestyle, which began in the Near East by 10,000 years ago, spread across Europe between 8000 and 4000 years ago (Price, 2000). Archaeological and genetic evidence has traditionally been divided between two viewpoints regarding the spread of agricultural lifestyles from the Near East. The earlier and more popular viewpoint argues that this transition is characterized by a large amount of *demic diffusion* involving a large influx of agricultural populations across Europe (Childe, 1958). Others view this transition as being dominated by *cultural diffusion*, such that Mesolithic hunter-gatherers in Europe embraced Neolithic lifestyles with little genetic input from Near Eastern farmers (Zvelebil and Dolukhanov, 1991).

Considering the two viewpoints, using phylogeographic analyses of present human DNA samples has led to contrasting results (Sokal and Menozzi, 1982; Ammerman and Cavalli-Sforza, 1984; Sokal et al., 1991; Puit et al., 1994; Richards et al., 2000; Rosser et al., 2000; Simoni et al., 2000; Belle et al., 2006; Balaresque et al., 2010). It is important to note that estimates based solely on phylogeographic analysis of genetic variation in extant humans can be biased by several factors. Importantly, recent demographic events, such as gene-flow and back-migration between Europe and the Near East can result in the geography of extant human populations misrepresenting that of their perceived ancestors (Richards et al., 2000; Balaresque et al., 2010). Additionally, when considering a limited number of genetic loci, clines of genetic variation can be sensitive to stochastic processes such as isolation by distance (Novembre and Stephens, 2008).

Ancient DNA is an ideal data source to circumvent these issues and settle this debate because genetic differences between skeletal samples associated with Mesolithic and Neolithic technologies can be directly assayed and correlated with cultural differences. A model of cultural diffusion predicts little to no major genetic differences associated with culture change in Holocene samples, while the demic diffusion model predicts substantial influx of novel genetic variation. The majority of aDNA studies of Europe have involved mtDNA analysis, primarily from hunter-gatherer and farming populations of north and central Europe. These studies have revealed that approximately ~83% of pre-Neolithic peoples of Europe carried mtDNA haplogroup U and none belong to haplogroup H, a composition that is markedly different from present samples in which haplogroup H is dominant (Haak et al., 2005, 2008, 2010; Bramanti, 2008; Bramanti et al., 2009; Guba et al., 2011; Fu et al., 2012; Lee et al., 2012; Nikitin et al., 2012). On the other hand, ~12% from early farming populations belong to haplogroup U, while haplogroup H is present in between 25 and 37% of mtDNA from early farming samples, both consistent with the haplotype frequencies of extant Europeans. Combined, these results from ancient mtDNA analyses suggest that pre-Neolithic hunter-gatherers contributed at most 20% to the mtDNA genetic composition of present European populations (Fu et al., 2012). These ancient mtDNA results were recently combined with a dataset of 1151 complete mtDNAs from across Europe (Fu et al., 2012). Fu and colleagues (2012) found evidence for a population expansion between 15,000 and 10,000 years ago in mtDNA typical of pre-Neolithic hunter-gatherers and a subsequent contraction of these haplotypes between 10,000 and 5000 years ago, consistent with the expansion of mtDNA from agricultural populations from the Near East.

In addition to ancient mtDNA data, results from whole genome sequencing of ancient Holocene-aged individuals have been applied to the question of Neolithic population replacement in Europe. Within the past two years, substantial amounts of autosomal DNA have been reported for seven ancient individuals in Europe. These include Ötzi (the Tyrolean Iceman) dating to roughly 5300 years ago (Keller et al., 2012), three hunter-gatherers associated with the Pitted Ware culture and one farmer associated with the Funnel Beaker culture from Scandinavia dating to approximately 5000 years ago (Skoglund et al., 2012), and two 7000 year old Iberian hunter-gatherers (Sánchez-Quinto et al., 2012).