Journal of Human Evolution 79 (2015) 21–34

Contents lists available at ScienceDirect

Journal of Human Evolution

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

Ancient human genomics: the methodology behind reconstructing evolutionary pathways

Stephanie Marciniak [a, *], Jennifer Klunk [a, b], Alison Devault [a, c], Jacob Enk [a, b, c], Hendrik N. Poinar [a, b, d, *]

[a] McMaster Ancient DNA Centre, Department of Anthropology, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L9, Canada
[b] Department of Biology, McMaster University, Hamilton, ON, Canada
[c] MYcroarray, Ann Arbor, MI, USA
[d] Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON, Canada

ARTICLE INFO

Article history:
Received 15 May 2014
Accepted 7 November 2014
Available online 15 January 2015

Keywords:
Ancient DNA
High-throughput sequencing
Human evolution

ABSTRACT

High-throughput sequencing (HTS) has radically altered approaches to human evolutionary research. Recent contributions highlight that HTS is able to reach depths of the human lineage previously thought to be impossible. In this paper, we outline the methodological advances afforded by recent developments in DNA recovery, data output, scalability, speed, and resolution of the current sequencing technology. We review and critically evaluate the 'DNA pipeline' for ancient samples: from DNA extraction, to constructing immortalized sequence libraries, to enrichment strategies (e.g., polymerase chain reaction [PCR] and hybridization capture), and finally, to bioinformatic analyses of sequence data. We argue that continued evaluations and improvements to this process are essential to ensure sequence data validity. Also, we highlight the role of contamination and authentication in ancient DNA-HTS, which is particularly relevant to ancient human genomics, since sequencing the genomes of hominins such as *Homo erectus* and *Homo heidelbergensis* may soon be within the realm of possibility.

Introduction

Beginning with the first Neandertal mitochondrial DNA sequences reported (Krings et al., 1997, 1999), the study of ancient human DNA has recently culminated in the sequencing of entire archaic hominin genomes for familiar species such as the Neandertal (Green et al., 2010) and novel 'species' like the Denisovans (Reich et al., 2010; Krause et al., 2010a). This has been made possible with the advent of high-throughput sequencing (HTS) technology, revolutionizing the reconstruction of human evolution and demographic history. Previously, it was argued that the retrieval of DNA from specimens older than 100,000 years would be difficult or impossible (Pääbo and Wilson, 1991; Lindahl, 1993; Krings et al., 1997). However, shifts in sequencing technology and bioinformatic approaches to ancient DNA sequence analysis demonstrate the feasibility of delving deep into the human lineage, as realized with a 400,000 year old mitochondrial genome retrieved from an unknown hominin that shared a common ancestor with the Denisovan mitochondrial lineage (Meyer et al., 2014). The insights gleaned from the Neandertal and Denisovan genomes illuminate how HTS technology can be used to impact our understandings of human origins and complement the paleontological and archaeological records.

Here we review advances in ancient DNA methods that have facilitated these contributions and discuss the challenges of retrieving high quality DNA from hominin fossil remains. We also explore how genomic data have been used to address hypotheses of human evolution, as well as the types of research questions best suited to the diverse sequencing strategies. Finally, we consider where further technological advancements in DNA retrieval and sequencing might take human evolutionary research in the future.

Ancient DNA: degradation and contamination

DNA in ancient samples is typically a combination of endogenous and contaminant sequences degraded into short fragments, often averaging 40–60 base pairs (bp), and of lower quantity than the DNA typical of modern biological samples (Pääbo, 1989;

* Corresponding authors.
E-mail addresses: marcis@mcmaster.ca (S. Marciniak), poinarh@mcmaster.ca (H.N. Poinar).

http://dx.doi.org/10.1016/j.jhevol.2014.11.003
0047-2484/

O'Rourke et al., 2000; Hofreiter et al., 2001; Millar et al., 2008; Stoneking and Krause, 2011; Fulton, 2012). It is critical to understand the impact of post-mortem DNA damage and contamination (e.g., pre- or post-laboratory) on retrieving and authenticating ancient DNA, as such factors may limit genomic analysis. A number of reviews have already examined the impact of ancient DNA degradation on obtaining authentic sequence data (e.g., O'Rourke et al., 2000; Hofreiter et al., 2001; Pääbo et al., 2004).

DNA preservation in ancient samples can vary dramatically as post-mortem degradation is an idiosyncratic process where often, very little or no endogenous DNA will survive and if it does, preservation is highly variable (Hofreiter et al., 2001; Pääbo et al., 2004; Willerslev and Cooper, 2005; Rohland and Hofreiter, 2007a; Stoneking and Krause, 2011). For example, of 21 bones screened for Neandertal-specific mitochondrial DNA by polymerase chain reaction (PCR) in Green et al. (2010), only three were selected for additional analysis, while seven of 15 Neandertal bones contained amplifiable mtDNA in Krings et al. (2000). This variability of DNA survival is due to not only rapidly occurring molecular damage (of unpredictable rates in complex systems) but the physical environment where the geochemistry of a site (e.g., soil pH, organic and inorganic organisms) facilitates or inhibits the subsequent retrieval of highly fragmented endogenous DNA. Broadly, the expectation is for the surviving fraction of endogenous DNA to constitute less than 1% and not exceed 5% of a sample, with 95%–99% representing contamination (e.g., environmental, post-mortem processes; Burbano et al., 2010; Green et al., 2010; Reich et al., 2010; Stoneking and Krause, 2011; Meyer et al., 2012; Carpenter et al., 2013; Fig. 1). Notable exceptions include permafrost remains, where well-preserved genomic data of a greater time-depth is retrievable from diverse specimens, including plants (e.g., Willerslev et al., 2003), mammals (e.g., mammoth [Hoss et al., 1994]; bison [Shapiro et al., 2004]; horse [Orlando et al., 2013]), microbes (e.g., Bellemain et al., 2013), or viruses (e.g., Legendre et al., 2014).

A theoretical limit of DNA preservation has been estimated between 100,000 and 1,000,000 years (Pääbo and Wilson, 1991; Lindahl, 1993; Willerslev and Cooper, 2005; Fulton, 2012), but the retrieval of ancient DNA is not temporally-bound, as a specimen's age is not linearly correlated to the amount of surviving DNA, which is characteristic of the inherent variability of DNA degradation and percentages of surviving DNA across specimens and spatio-temporal contexts (Tuross, 1994; O'Rourke et al., 2000; Pääbo et al., 2004; Millar et al., 2008; Allentoft et al., 2012) (Table 1).

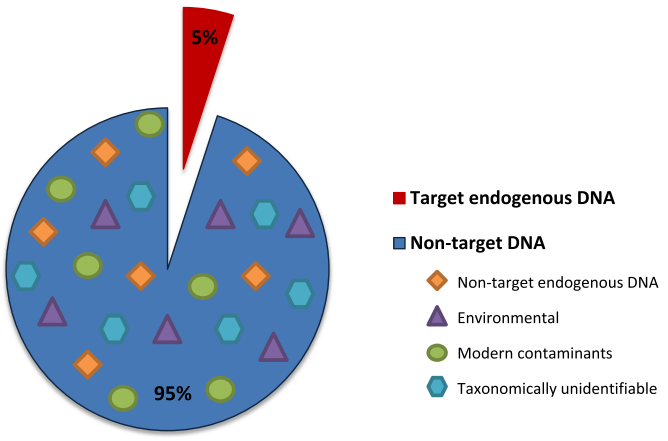

Figure 1. Non-target DNA (approximately 95%) comprises the majority of surviving DNA in ancient samples, whereas the desired or targeted endogenous DNA is only a fraction (approximately 0–5%) of the overall constituents.

For example, depositional environments, whether the same or different sites, affect the fraction of preserved endogenous DNA in an unpredictable manner, such as a Denisovan phalanx containing 70% endogenous DNA recovered from a cave (southern Siberia) by Reich et al. (2010), in comparison to the 0.01%–0.03% of endogenous DNA recovered from an early modern human at the Tianyuan Cave site (northern China) by Fu et al. (2013a). Additionally, intra-site variability or differential tissue preservation may also exist, such as 0.17% endogenous DNA characterizing the Denisovan tooth compared with 70% endogenous DNA of the phalanx (Reich et al., 2010). The advantage of HTS technology in light of unpredictable preservation is maximizing detection of those short surviving endogenous DNA fragments from the overwhelming pool of non-specific sequences.

The characteristics of DNA in ancient samples, primarily the low quality and quantity of endogenous DNA with variable preservation, are the result of dynamic post-mortem molecular degradation processes due to strand breaks, baseless sites, miscoding lesions, and cross-links, to name a few (Pääbo, 1989; Lindahl, 1993; Hoss et al., 1996; Hofreiter et al., 2001; Pääbo et al., 2004; Willerslev and Cooper, 2005; Fulton, 2012). The cellular and biomolecular processes that support DNA integrity during life cease functioning after death, causing cellular degradation by endogenous nucleases and proteases with associated infiltrations of exogenous bacteria, fungi, or other organisms that further digest and non-specifically fragment the DNA (Hofreiter et al., 2001; Pääbo et al., 2004; Molak and Ho, 2011). Molecular DNA damage is broadly categorized as: 1) shortening lesions that reduce the size of DNA, which prevent extension of the polymerase during PCR, such as strand breaks, condensation, cross-links, and oxidative damage; and 2) miscoding lesions due to hydrolytic damage, particularly deamination, that produce incorrect sequence reads, such as C to T (replacement of cytosine with uracil) and G to A (opposite strand) transitions (Hoss et al., 1996; O'Rourke et al., 2000; Hofreiter et al., 2001; Pääbo et al., 2004; Molak and Ho, 2011; Fulton, 2012; Table 2). Proposed resolutions to counteract the effects of such damage include the use of N-phenylacyl thiazolium bromide (PTB) to break Maillard products, uracil-DNA-glycosylase to remove cytosine deamination, and overlapping amplifications of short PCR fragments alongside multiple extractions to overcome contaminating effects of damaged nucleotides (Pääbo et al., 2004; Stoneking and Krause, 2011; Fulton, 2012). However, debate exists on how useful these approaches are to modify post-mortem damage (e.g., Hofreiter et al., 2001; Malmstrom et al., 2005; Rohland and Hofreiter, 2007b).

Aside from molecular damage to ancient DNA, exogenous DNA contamination of samples may also occur due to cross-contamination between samples or experiments, laboratory equipment or reagents, and/or through laboratory personnel, including post-excavation specimen handling (Pääbo et al., 2004; Brown and Brown, 2011; Fulton, 2012; Barta et al., 2013). Cross-contamination in the ancient DNA laboratory from previous PCR preparations (not PCR amplifications, which are restricted to the modern clean rooms) or related DNA work, such as preparing samples for extraction, library preparation, or indexing, is a critical issue that has not been fully addressed. In this situation, DNA-containing molecules remain airborne and/or adhere to lab surfaces and equipment, which transfer contaminants to proceeding experiments (Brown and Brown, 2011; Fulton, 2012; Barta et al., 2013). This type of contamination is often undetectable, as negative controls employed in the experiment design may be unaffected, and is only revealed when sequence data are obtained that are contrary to the expectations within a genomic study (e.g., not phylogenetically sound; Handt et al., 1996; Brown and Brown, 2011; Fulton, 2012).