

## Evolutionary Patterns and Processes: Lessons from Ancient DNA

MICHELA LEONARDI<sup>1</sup>, PABLO LIBRADO<sup>1</sup>, CLIO DER SARKISSIAN<sup>1</sup>, MIKKEL SCHUBERT<sup>1</sup>,  
AHMED H. ALFARHAN<sup>2</sup>, SALEH A. ALQURAISHI<sup>2</sup>, KHALED A. S. AL-RASHEID<sup>2</sup>, CRISTINA GAMBA<sup>1</sup>,  
ESKE WILLERSLEV<sup>1,3</sup>, AND LUDOVIC ORLANDO<sup>1,4</sup>

<sup>1</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350K Copenhagen, Denmark; <sup>2</sup>Zoology Department, College of Science, King Saud University, Riyadh 11451, Saudi Arabia; <sup>3</sup>Department of Zoology, University of Cambridge, Downing St, Cambridge CB2 3EJ, UK; and <sup>4</sup>Université de Toulouse, University Paul Sabatier (UPS), Laboratoire AMIS, CNRS UMR 5288, 37 Allées Jules Guesde, 31000 Toulouse, France

\*Correspondence to be sent to: Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350K Copenhagen, Denmark; E-mail: [Lorlando@snm.ku.dk](mailto:Lorlando@snm.ku.dk)

Received 26 March 2016; reviews returned 4 June 2016; accepted 6 June 2016

Associate Editor: Olivier Gascuel

**Abstract.**—Ever since its emergence in 1984, the field of ancient DNA has struggled to overcome the challenges related to the decay of DNA molecules in the fossil record. With the recent development of high-throughput DNA sequencing technologies and molecular techniques tailored to ultra-damaged templates, it has now come of age, merging together approaches in phylogenomics, population genomics, epigenomics, and metagenomics. Leveraging on complete temporal sample series, ancient DNA provides direct access to the most important dimension in evolution—time, allowing a wealth of fundamental evolutionary processes to be addressed at unprecedented resolution. This review taps into the most recent findings in ancient DNA research to present analyses of ancient genomic and metagenomic data. [Ancient DNA; metagenomics; palaeogenomics; population genomics; temporal sampling.]

The first DNA sequence from an extinct organism was reported in 1984, after Russell Higuchi and colleagues successfully extracted DNA from a quagga zebra museum specimen prepared in 1883 (Higuchi et al. 1984). With only 229 nucleotides from the mitochondrial genome, the amount of genetic information they characterized was admittedly limited. However, the authors anticipated that “if the long-term survival of DNA proves to be a general phenomenon, several fields including palaeontology, evolutionary biology, archaeology and forensic science may benefit” (Higuchi et al. 1984). They were proven right. Thirty years later, the complete genome of this extinct group of zebras was sequenced to an ~8× average depth of coverage (Jónsson et al. 2014).

### *A Wealth Of Ancient genomes Has Now Been Characterized*

Ancient DNA research has made massive progress in its rather short history, extending greatly beyond the temporal range covered by museum specimens. In 2008, the first draft of the complete genome of an extinct species—the woolly mammoth—was characterized, benefitting from exceptionally well-preserved DNA material present in hairs kept frozen for ~40 kyr in the Siberian permafrost (Miller et al. 2008). This unleashed the field of palaeogenomics, focusing on the genome characterization of extinct species and ancient individuals. Four high-quality genomes of mammoth specimens have now been characterized at ~11×, 20× coverage, unveiling the genetic makeup of this fascinating elephantid lineage, including signatures of

adaptation to the cold Arctic (Lynch et al. 2015), as well as the dynamics of its demographic expansions and collapses over the last 4 Myr (Palkopoulou et al. 2015).

The first ancient human genome was sequenced at high-coverage (~20×) in 2010 from a ~4 kyr-old hair shaft of a palaeo-Eskimo Greenlander (Rasmussen et al. 2010). Up to that time, the general understanding was that no reliable DNA sequence could be retrieved from anatomically modern human remains, due to problems with contamination, as exemplified in Green and colleagues (Green et al. 2009): “...for Cro-Magnons and other modern humans, the problems (of contamination) are so severe that over the past 15 years we have been pessimistic over the prospects of ever reliably determining such DNA sequences.” Yet, at the middle of 2016, more than 500 ancient humans have been analyzed for genome-wide sequence variation, including Upper Palaeolithic individuals (Fu et al. 2014, 2016; Raghavan et al. 2014b; Seguin-Orlando et al. 2014; Jones et al. 2015), pre-historical and historical Amerindians (Malaspina et al. 2014; Rasmussen et al. 2014, 2015a; Raghavan et al. 2015), Mesolithic hunter-gatherers (Lazaridis et al. 2014; Olalde et al. 2014; Günther et al. 2015; Jones et al. 2015; Fu et al. 2016), Neolithic hunter-gatherers and farmers (Gamba et al. 2014; Lazaridis et al. 2014; Günther et al. 2015; Haak et al. 2015; Llorente et al. 2015; Mathieson et al. 2015; Olalde et al. 2015; Cassidy et al. 2016), Bronze Age Eurasians (Gamba et al. 2014; Allentoft et al. 2015; Haak et al. 2015; Cassidy et al. 2016), and historical samples (Schroeder et al. 2015; Martiniano et al. 2016; Schiffels et al. 2016).

Besides Anatomically Modern Humans (AMH), the first draft of the Neanderthal genome was released in

2010, opening ancient genomics to the study of archaic hominins (Green et al. 2010). Soon after, a new species, otherwise totally unknown to physical anthropologists, was discovered based on the divergent genetic signature from a phalanx bone: the so-called “Denisovans,” named after the cave where the small bone fragment was excavated (Reich et al. 2010; Sawyer et al. 2015). Since then, high-quality genomes from both Neanderthals and Denisovans have been generated (Meyer et al. 2012; Fu et al. 2014) together with draft genomes of additional Neanderthal individuals (Prüfer et al. 2014), providing—as closer relatives than the chimpanzee—insights into the origins and genetic makeup of our own species (Pääbo 2014, 2015), and evidence of ancient admixture between different hominin lineages (Meyer et al. 2012; Fu et al. 2014, 2015; Prüfer et al. 2014; Kuhlwilm et al. 2016; Vernot et al. 2016). At the time these lines are written (spring 2016), nearly 80 ancient human genomes (with an average depth of coverage above 1.0×) have been characterized, and this number is expected to more than double within the forthcoming months (Table 1).

Until then limited to the Late Pleistocene, the temporal range of palaeogenomics was pushed back to the Middle Pleistocene with the sequencing of a draft genome from a 560 to 780 kyr-old permafrost-preserved horse bone (Orlando et al. 2013), the complete mitochondrial genome of a ~400 kyr-old cave bear (Dabney et al. 2013) and archaic hominin (Meyer et al. 2014) from Spain, and nuclear sequence data from the latter sample (Meyer et al. 2016). Yet, only a few years ago, ancient DNA researchers were unanimous in thinking that “*barring some unimaginable technical advance, diachronical studies of DNA sequences are confined to the past one million years, and more probably to the past 100,000 years.*” (Hofreiter et al. 2001b). Clearly, this “unimaginable” technical advance has been realized.

#### *A Technological Revolution*

The advent of high-throughput sequencing (HTS) has been the main driver behind the emergence of palaeogenomics, significantly reducing the amount of sample, time, and costs involved. With HTS, ancient DNA fragments are unspecifically built into DNA libraries, which are amplified by PCR and massively sequenced in parallel, to up to several billion sequencing reads in few days. Such sequencing capacities made the characterization of ancient genomes feasible, as ancient DNA libraries often contain extremely low amounts of endogenous DNA, with most templates deriving from the environment and from microbes colonizing fossil material after death (Der Sarkissian et al. 2015a).

In addition to brute force shotgun sequencing, a range of molecular techniques has been developed to improve access to ancient DNA molecules, starting from the extraction stage, with methods tailored to the

recovery of ultra-short (~25–50 bp) DNA fragments (Dabney et al. 2013; Allentoft et al. 2015; Gamba et al. 2016). Some fossil material, such as hair (Gilbert et al. 2004), the petrous bone (Gamba et al. 2014; Rasmussen et al. 2014), and tooth cementum (Adler et al. 2011; Damgaard et al. 2015), are now recognized for their superior DNA preservation, often reducing the cost of ancient genome sequencing. Partial digestion of the bone matrix (Ginolhac et al. 2012; Der Sarkissian et al. 2014; Damgaard et al. 2015; Gamba et al. 2016), bleach treatment of the bone powder (Korlević et al. 2015), and immuno-precipitation of methylated DNA (Seguin-Orlando et al. 2015) also offer novel alternatives to reduce the microbial fraction present in ancient DNA extracts. Additionally, innovative DNA library construction methods can improve access even to the most damaged DNA templates (Meyer et al. 2012; Gansauge and Meyer 2013, 2014), and with targeted enrichment, the regions required to address a particular biological question, which can range from a limited number of loci (Fabre et al. 2014) to several millions genome-wide markers (Haak et al. 2015; Fu et al. 2016) and even the entire genome (Carpenter et al. 2013), can be characterized. Finally, computational tools have also improved significantly. Read alignment parameters optimized for ancient DNA data have been identified (Green et al. 2009; Schubert et al. 2012) and alignment procedures building on the specificities of ancient DNA nucleotide mis-incorporations have been developed to maximize alignment specificity and accuracy (Kerpedjiev et al. 2014).

#### *Towards Integrative Ancient “-omic” Approaches*

In addition to the genome of their hosts, ancient DNA studies have successfully characterized the genome of a number of ancient pathogens, including some of the deadliest in human history such as those responsible for the Spanish flu of 1917–1918 (Taubenberger et al. 2001) and the Black Death of 1348–1350 (Bos et al. 2011). They have even revealed plague to be widespread some 3,000 years before any written recording (Rasmussen et al. 2015). The recent discovery of excellent microbial DNA preservation in ancient dental calculus (calcified dental plaque) (Adler et al. 2013; Warinner et al. 2014) also allows for the reconstruction of ancient oral microbiomes. This provides a temporal dimension to the analysis of biological systems that play an important role in the development of diseases and cognitive disorders (Burokas et al. 2015; Rosenbaum et al. 2015).

Beyond genomic sequences, ancient epigenomes have started to be characterized, with the first genome-wide methylation and nucleosome maps reconstructed from *post-mortem* patterns of DNA degradation along the genome (Gokhman et al. 2014; Pedersen et al. 2014) or from direct enrichment for methylated CpG sites (Seguin-Orlando et al. 2015). This promises

TABLE 1. List of published ancient whole genomes above 1× depth of coverage

References	Organism	Date	Number of genomes above 1× depth of coverage	Average coverage
Reich et al. (2010); Meyer et al. (2012)	Archaic hominin—Denisovan	40,000 yrs BP	1	30
Green et al. (2010)	Archaic hominin—Neanderthal	44,450–38,310 yrs BP	1	1.3
Prüfer et al. (2014)	Archaic hominin—Neanderthal	50,000 yrs BP	1	52
Rasmussen et al. (2010)	Human	4000 yrs BP	1	20
Rasmussen et al. (2011)	Human	200 yrs BP	1	6.4
Keller et al. (2012)	Human	5300 yrs BP	1	7.6
Fu et al. (2014)	Human	45,000 yrs BP	1	42
Gamba et al. (2014)	Human	7680–2830 yrs BP	8	1–22.1
Lazaridis et al. (2014)	Human	8000–7000 yrs BP	3	2.4–22
Malaspinas et al. (2014)	Human	500–300 yrs BP	2	1.2–1.5
Olalde et al. (2014)	Human	7000 yrs BP	1	3.4
Raghavan et al. (2014b)	Human	24,000 yrs BP	1	1
Rasmussen et al. (2014)	Human	10,705 yrs BP	1	14.4
Seguin-Orlando et al. (2014)	Human	36,200 yrs BP	1	2.4
Skoglund et al. (2014a)	Human	5050–4,600 yrs BP	2	1.3–2.2
Allentoft et al. (2015)	Human	5400–1400 yrs BP	19	1.1–7.4
Llorente et al. (2015)	Human	4500 yrs BP	1	12.5
Günther et al. (2015)	Human	5200–4,600 yrs BP	4	1.1–4.1
Jones et al. (2015)	Human	13,700–9700 yrs BP	3	1.4–15.4
Olalde et al. (2015)	Human	7400 yrs BP	1	1.1
Raghavan et al. (2015)	Human	200 yrs BP	3	1.0–1.7
Rasmussen et al. (2015a)	Human	9200–8340 yrs BP	1	1
Cassidy et al. (2016)	Human	3300–1500 yrs BP	2	1.5–10.3
Martiniano et al. (2016)	Human	1900–1100 yrs BP	6	1.1–1.7
Schiffels et al. (2016)	Human	2360–1230 yrs BP	9	1.3–11.8
Park et al. (2015)	Mammal—Aurochs	6750 yrs BP	1	6.2
Orlando et al. (2013)	Mammal—Horse	780,000–43,000 yrs BP	2	1.1–1.8
Schubert et al. (2014b)	Mammal—Horse	43,000–16,000 yrs BP	2	7.4–24.3
Librado et al. (2015)	Mammal—Horse	5200–100 yrs BP	2	18.3–20.2
Lynch et al. (2015)	Mammal—Mammoth	30,000–20,000 yrs BP	2	20
Palkopoulou et al. (2015)	Mammal—Mammoth	44,800–4300 yrs BP	2	11.1–17.2
Der Sarkissian et al. (2015b)	Mammal—Przewalski horse	200 yrs BP	3	1.4–4.3
Jónsson et al. (2014)	Mammal—Quagga	100 yrs BP	1	7.9
Skoglund et al. (2015)	Mammal—Wolf	35,000 yrs BP	1	1
Maixner et al. (2016)	Pathogen— <i>Helicobacter pylori</i>	5300 yrs BP	1	18.5
Schuenemann et al. (2013)	Pathogen— <i>Mycobacterium leprae</i>	1023–644 yrs BP	5	10.1–105.1
Librado et al. (2014)	Pathogen— <i>Mycobacterium leprae</i>	1050–950 yrs BP	2	6.7–44.8
Bos et al. (2014)	Pathogen— <i>Mycobacterium tuberculosis</i>	1000–700 yrs BP	3	20.5–31.4
Martin et al. (2013)	Pathogen— <i>Phytophthora infestans</i>	150 yrs BP	5	3–22
Yoshida et al. (2013)	Pathogen— <i>Phytophthora infestans</i>	150 yrs BP	11	~1–~28
Bos et al. (2011v)	Pathogen— <i>Yersinia pestis</i>	650 yrs BP	1	30
Rasmussen et al. (2015b)	Pathogen— <i>Yersinia pestis</i>	4887–3626 yrs BP	2	8.7–29.5
Bos et al. (2016)	Pathogen— <i>Yersinia pestis</i>	300 yrs BP	5	84.8–92.8

to deliver important biological insights on ancient gene expression levels and on the contribution of regulatory changes in the evolutionary processes of adaptation and species divergence (Orlando and Willerslev 2014).

In this review, we will not present the molecular and computational methods underlying the characterization of ancient genomes and epigenomes, which have recently been covered elsewhere (Heintzman et al. 2015; Orlando et al. 2015). We will instead focus on describing what can be learnt from ancient genomes when using phylogenetic approaches and methods in statistical and population genomics. The underlying concepts and limits respective to each method will be presented and illustrated with the results from a diversity of studies.

#### PHYLOGEOGRAPHIC INFERENCE

Phylogeographic methods provided a functional framework for early genetic analyses (Avise et al. 1987), exploring the association of geographic patterns with the gene genealogy of the surveyed individuals using phylogenetic trees. Integrating ancient data into such trees has often revealed that population histories are much more complex than could be inferred from the genetic variation present in extant populations, and has led to a better understanding of the true dynamics of speciation processes, past population movements, and expansions and declines (Table 2). Regarding animal domestication, for example, the phylogenetic tree based on mitochondrial DNA (mtDNA) sequences from modern pig and their wild progenitor, the wild boar, showed a clear geographic pattern suggesting

TABLE 2. Examples of possible applications of ancient DNA analyses to different fields

Field of application	Method	Organism(s) or population	Major finding	References
Vicariant speciation	Phylogenetics (1)	Hutias	Colonization of the Antilles via rafting from South America.	<a href="#">Fabre et al. (2014)</a>
		Ratites	No continental vicariance, but dispersal by flying ancestors.	<a href="#">Mitchell et al. (2014)</a>
Population migration and replacement	Phylogenetics (1)	Pigs and wild boars	Separate pig domestication in Europe and then population replacement by Near-eastern domesticates.	<a href="#">Larson et al. (2005, 2007)</a> ; <a href="#">Krause-Kyora et al. (2013)</a>
	Serial Coalescent Simulations and ABC model fitting (3), PCA (6), Clustering methods (6), $f_3$ -statistic (6), $f_4$ -statistic (6)	Palaeo-Mesolithic hunter-gatherers, Neolithic farmers, present-day individuals	Genetic discontinuity associated with Neolithic transition in Europe supporting a demic diffusion from the Near East.	<a href="#">Bramanti et al. (2009)</a> ; <a href="#">Malmström et al. (2009)</a> ; <a href="#">Haak et al. (2010)</a> ; <a href="#">Gamba et al. (2012)</a> ; <a href="#">Skoglund et al. (2012)</a> ; <a href="#">Lazaridis et al. (2014)</a> ; <a href="#">Seguin-Orlando et al. (2014)</a>
	Clustering methods (6), $f_4$ -statistic (6)	Neolithic to Bronze Age individuals	At least three main genetic components in modern-day Europeans.	<a href="#">Allentoft et al. (2015)</a> ; <a href="#">Haak et al. (2015)</a>
	MSMC (3), clustering methods (6), $f_3$ -statistics (6), admixture graph (6)	Ancient and modern-day Native Americans and Siberians	Reconstruction of the peopling of Northern Siberia and America.	<a href="#">Raghavan et al. (2014b, 2015)</a>
Estimation of substitution rates	PMSC (3)	Humans, mammoth	Estimation of substitution rates.	<a href="#">Fu et al. (2014)</a> ; <a href="#">Palkopoulou et al. (2015)</a>
Calibration	Branch shortening method (2), Shotgun sequencing-based metagenomics (8)	<i>Yersinia pestis</i> , <i>Mycobacterium tuberculosis</i>	Date and genetic mechanisms underlying the emergence of the different strains.	<a href="#">Bos et al. (2011, 2014)</a> ; <a href="#">Cui et al. (2013)</a> ; <a href="#">Kay et al. (2015)</a> ; <a href="#">Rasmussen et al. (2015b)</a>
Molecular dating	Branch shortening method (2)	Denisovans, <i>Ursus deningeri</i>	Matching dates around 400 kyr ago from two different organisms from the same site.	<a href="#">Dabney et al. (2013)</a> ; <a href="#">Meyer et al. (2014)</a>
Changes in effective population size	Bayesian skyline plots (3), PMSC (3), Serial Coalescent Simulations and ABC model fitting (3)	Bison, mammoth, woolly rhino, horse, musk-ox.	Population expansions/collapses match global climatic changes.	<a href="#">Shapiro et al. (2004)</a> ; <a href="#">Drummond et al. (2005)</a> ; <a href="#">Campos et al. (2010)</a> ; <a href="#">Lorenzen et al. (2011)</a> ; <a href="#">Orlando et al. (2013)</a> ; <a href="#">Metcalf et al. (2014)</a> ; <a href="#">Schubert et al. (2014b)</a> ; <a href="#">Palkopoulou et al. (2015)</a>
	SMC (3)	Neanderthals, Denisovans	Demographic profiles for both species.	<a href="#">Prüfer et al. (2014)</a>
	Serial Coalescent Simulations and ABC model fitting (3)	Patagonian rodents	Present-day low diversity is the result of a severe bottleneck that happened 3 kyr ago.	<a href="#">Chan et al. (2006)</a>

(continued)

TABLE 2. Continued

Field of application	Method	Organism(s) or population	Major finding	References
Testing for ancestry	Ancestry test (6)	Kennewick man, Anzick infant, modern Native Americans	Revealed the true relationships of the samples with Native Americans.	Rasmussen et al. (2014, 2015a)
	$f_3$ -statistic (6)	Mota, modern-day Ethiopians	Genetic continuity in East Africa since 4.5 kyr ago.	Llorente et al. (2015)
	PCA (6)	Palaeo-Eskimo	Closer affinities with Siberians than with Greenlanders.	Rasmussen et al. (2010)
Admixture between different species	$f_3$ -statistic (6), $D$ -statistics (6), enhanced $D$ -statistics (6), admixture graphs (6), local ancestry and admixture dates (6)	Neanderthals, Denisovans, Anatomically Modern Humans	Admixture between different hominin species.	Green et al. (2010); Reich et al. 2010; Meyer et al. 2012; Fu et al. 2014; Prüfer et al. (2014); Seguin-Orlando et al. 2014; Vernot et al. 2016
Ancient signatures of adaptation	Reconstructing allele frequencies through time (7)	Horses	Selection signatures for specific phenotypic traits in horses.	Ludwig et al. (2009, 2014)
		Ancient and modern Eastern Europeans	Selection signatures on light pigmentation-related genes in humans.	Wilde et al. (2014)
Recognition of selected loci	Comparing genomes of closely related species (7)	Horses	Selection signatures associated with horse domestication led to changes in behavioral traits, circulatory system and physical structure.	Schubert et al. (2014b)
		Neanderthals, modern humans	Selective sweeps in humans more likely related to regulatory changes than protein-coding genes.	Green et al. (2010); Prüfer et al. (2014); Pääbo (2015)
Tracking ecosystem changes associated with climatic and/or anthropic factors	Metabarcoding (8)	Permafrost sediments from the Arctic	Ecosystem changes associated with the Last Glacial Maximum.	Willerslev et al. (2014)
		Alpine lake sediments	Reconstruction of changes in the distribution of plants and domestic animal species.	Giguët-Covex et al. (2014); Pansu et al. 2015
		Dental calculus from ancient and modern humans	Changes in oral microbiota associated with important dietary shifts during the Neolithic and Industrial age.	Adler et al. (2013)
Detecting contamination	SourceTracker (8)	Ancient coprolites, ancient dental calculus	Contamination estimates in different samples.	Tito et al. (2012); Adler et al. 2013

In brackets the section where each method is described. Legend: 1: phylogeographic inferences; 2: calibration and divergence estimates; 3: population demography; 4: genotype likelihood; 5: genetic continuity over time; 6: population structure and admixture; 7: tracking genetic signatures of selection; 8: ancient metagenomics.

that European pigs were domesticated *in situ*, while archaeological evidence suggested that they were introduced from the Near East (Larson et al. 2005). However, ancient mtDNA data from animals spread across the whole temporal range of domestication supported a much more complex scenario. In this model, pigs were first domesticated within the Near East, and were later introduced into Europe, before being ultimately replaced by incorporation of local wild European boars into the domestic pool (Larson et al. 2007; Krause-Kyora et al. 2013).

Phylogenetic analyses based on ancient DNA have also been widely used to study the dispersal of now extinct taxa and vicariant speciation, the process leading to the development of new species due to geographic fragmentation. Vicariant speciation in the Cretaceous, following the separation of the main Gondwana supercontinent, was for example proposed to explain the evolutionary origin of non-flying ratite birds (ostrich, rhea, kiwi, emu, cassowary, and the extinct elephant bird and moa). Their mtDNA phylogeny and divergence times, including data from two different genera of extinct elephant birds, ruled out the hypothesis of continental vicariance (Mitchell et al. 2014). The results not only supported an initial dispersal by flying ancestors, but also provided a possible explanation for the disparity in size between lineages living in the same region. The first genus to arrive on each landmass is indeed proposed to have monopolized the available niche space for large flightless herbivores and omnivores, forcing subsequent newcomers to adopt different life history traits, including smaller sizes (Mitchell et al. 2014).

Beside classical biases associated with phylogenetic reconstructions (e.g., Hagen et al. 2015), additional problems can arise when considering temporally sampled DNA data. As a common assumption in phylogenetics is that all samples are tips, direct ancestry relationships are not allowed, which can affect parameter estimates if not properly modelled (Gavryushkina et al. 2014). In addition, inflated sequence error rates can derive from extensive DNA chemical reactions taking place after death (Briggs et al. 2007; Sawyer et al. 2012), resulting in substantial overestimates of the substitution rate, distorting branch lengths, and potentially the true topology of the phylogenetic tree (Fig. 1). Other issues are related to the fact that ancient DNA studies have traditionally focused on haploid and uniparentally inherited loci, such as the Y-chromosome and, even more frequently, mtDNA due to its high copy number per cell (Pakendorf and Stoneking 2005). These markers can be extremely useful to investigate sex-specific processes, such as the respective contributions of mares and stallions during horse domestication (Vila et al. 2001; Lippold et al. 2011). However, they also face a number of limitations. First, they may not always contain enough information to build well-resolved phylogenies or retrace the full evolutionary history of species or populations (Nordborg 1998; Goldstein and Chikhi 2002). Second, not only is the gene tree of a single marker affected by the stochastic nature underlying evolutionary processes,

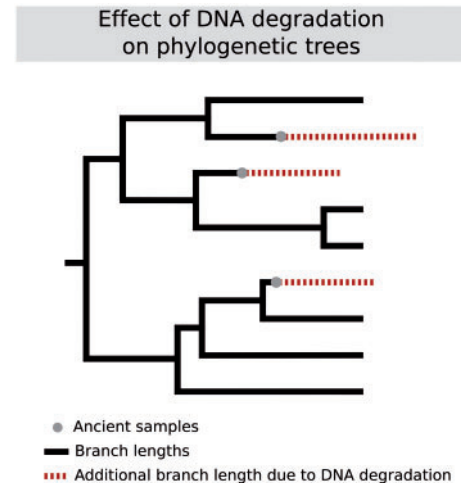


FIGURE 1. Effect of ancient DNA degradation on phylogenetic reconstruction, when including modern and ancient sequences. The continuous lines show the hypothetical tree, the points mark the true chronological location of the ancient samples, while the dotted lines represent the branch length artificially increased by DNA degradation-related mutations.

but also different population histories can also lead to the same tree (Nielsen and Beaumont 2009). Finally, phylogeography assumes that gene genealogies are only shaped by past demographic events, but natural selection can create similar molecular patterns (Currat et al. 2006). The Demography section below describes some methods for leveraging the power of multiple loci to reconstruct past demographic histories.

#### CALIBRATION AND DIVERGENCE ESTIMATES

Phylogenetic methods can exploit and provide insights into the time of divergence between species and populations with the “molecular clock” hypothesis, allowing the direct conversion of the substitution rate estimates calculated from genetic data into absolute time scales (see Ho and Duchêne 2014 for an extensive review). Calibration points are essential to this conversion, as they anchor samples (tip-dating) and/or population splits (internal node calibration) to specific dates, which enables extrapolation of the time scale for the remaining samples and nodes (Shapiro et al. 2011). Traditionally, calibration points are found within the fossil record, and/or biogeographic (e.g. vicariance) and bio-cultural dates (Ho et al. 2015).

Clock calibration is usually performed when reconstructing phylogenetic trees in a Bayesian statistical framework (Ho and Duchêne 2014), for example, with the software BEAST (Drummond et al. 2012). None of the clock calibration methods are, however, devoid of limitations (reviewed in Ho et al. 2011a), despite important methodological improvements for handling uneven evolutionary rates across sites (Soubrier et al. 2012) or clades (the so-called “relaxed” clocks). The initial Bayesian implementation

underlying the “relaxed” clock enabled accommodation of uncertainties in the diversifying process through arbitrary prior distributions, both for the calibration nodes and for the coupled birth-and-death process of lineage speciation and extinction, which is ultimately informative about the age of the uncalibrated nodes. By considering heterochronous taxa as part of the same macroevolutionary process, the recently developed “fossilized birth–death” (FBD) model unifies the stochastic nature of lineage diversification (Heath et al. 2014). Although this model was mainly proposed for the analysis of infectious diseases, it is also specially suitable for temporal samples, including ancient DNA data (e.g., Stadler and Yang 2013).

Recently, Pairwise Sequential Markov Coalescent (PSMC) demographic profiles (Li and Durbin 2011) (see Sequential Markov Coalescent (SMC) section) have offered a novel way to directly estimate substitution rates. This was first implemented by Fu et al. (2014), who compared the demographic trajectories as inferred from a PSMC analysis of eleven non-African human genomes, with that obtained from the genome of a ~45-kyr-old AMH from Ust'-Ishim, Russian Federation. Here, genome-wide substitution rates were estimated by maximizing the overlap between ancient and modern demographic trajectories, assuming that both modern and ancient groups showed similar recombination rates and generation times. The substitution rate estimated was close to that derived from familial pedigrees in humans, representing about half that calculated when using 6 Myr for a divergence date between the chimpanzee and the human lineage (Sally and Durbin 2012).

Substitution rate estimates can also be derived from the branch length differences observed along the tree. Under the so-called branch-shortening method, a clock model is assumed, and any branch leading to ancient samples is expected to be shorter than that leading to extant organisms due to the fact that their evolution stopped at the time the individual died. Relating the extent of branch shortening to their age (e.g., as estimated from radiocarbon dating) thus provides a direct substitution rate estimate. This approach is particularly useful in the case of bacteria, for which external calibration points are not available. For *Yersinia pestis*, the agent of plague, Bos and colleagues made use of the genome of the Black Death strain to calibrate a genome-wide molecular clock, and propose 668–729 years ago as a date for the emergence of pathogenic *Y. pestis* (Bos et al. 2011). This date has been recently pushed back in time to 1,505–6,409 years ago on the basis of genomes from isolates from different Asian locations (Cui et al. 2013) and finally to 5,021–7,022 years ago following the discovery of early plague in Bronze Age humans (Rasmussen et al. 2015b).

Molecular clocks calibrated with ancient DNA have also been used to estimate the age of samples for which direct dates cannot be obtained, such as biological material beyond the range of radiocarbon dating. Phylogenetic trees are reconstructed within a Bayesian

framework, and the ages of internal nodes are treated as unknown parameters. The ages of all sampled external nodes are known (e.g., from radiocarbon dating and/or cultural context) except the one that needs to be estimated, which represents an additional model parameter (Shapiro et al. 2011). This method has been used to date the Sima de los Huesos hominin to about 400 kyr ago (confidence range = 150–640 kyr) (Meyer et al. 2014), a result in strong agreement with the age recovered with the same method using the mtDNA sequence of the contemporary *Ursus deningeri* material found at the same site (Dabney et al. 2013). Recently, the original method has been modified to also take into account uncertainties in radiocarbon dating of the samples used for tip dating (Molak et al. 2015).

The disadvantage of this method is the high uncertainties in the mutation rate estimates (Ségurel et al. 2014) and, given the short time frame of interest compared to the total human–chimpanzee divergence, even low errors can substantially bias the results (Moorjani et al. 2016). For this reason a recent study (Moorjani et al. 2016) proposes a variant of the branch shortening method based on the recombination clock (Hinch et al. 2011). The proof-of-concept was developed for dating ancient non-African individuals, leveraging on the size of Neanderthal blocks present in their genome. The basic idea is to estimate the date of Neanderthal introgression separately for ancient and modern non-African genome(s), by jointly modeling recombination rates across the genome and the decay of Neanderthal ancestry through time. It is then possible to calculate the age of the ancient specimen as the difference between those two values. Tested on five ancient Eurasian genomes older than 10,000 years ago, this method provided age estimates largely consistent with radiocarbon dates. It is worth noting that this method gives more precise estimates as the age of the individual is closer to the admixture event, but performs poorly towards more recent times, making it complementary to radiocarbon dating (Moorjani et al. 2016).

#### POPULATION DEMOGRAPHY

The main strength of ancient DNA is to open access to genetic variation in past populations, thereby providing maps of allelic frequencies through space and time. This information can significantly improve the statistical power to detect past population processes, for example, demographic expansions and declines (Ramakrishnan and Hadly 2009; Mourier et al. 2012), population splits (Skoglund et al. 2014b), etc. This is particularly relevant in cases where some historical events, for example, demographic collapses, have erased most of the genetic variation existing in the past and where our ability to resolve past population dynamics from modern DNA data alone is limited (Heled and Drummond 2008, see also Fig. 2b). However, the additional power gained by including ancient DNA depends on the summary statistics used, the effective population size ( $N_e$ ), the

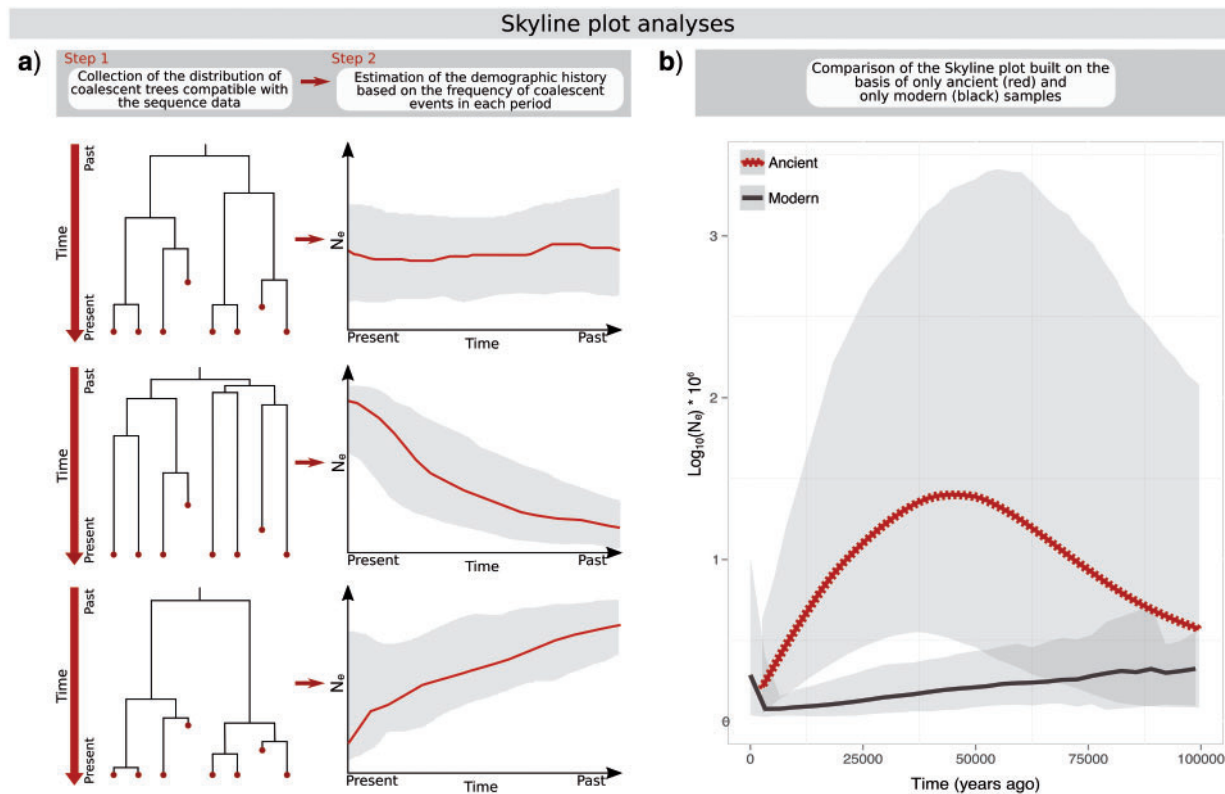


FIGURE 2. Skyline-plot methods. a) Overview on how skyline-plot methods are used to reconstruct the variation of effective population size through time. After the reconstruction of the gene genealogy (step 1), the effective population size is estimated over time, on the basis of the density of coalescent events present in each corresponding time interval (step 2). The thick solid line is the median estimate, and the coloured area shows the 95% bounds of the highest posterior density (HPD). b) Bayesian skyline plot reconstructed from mitochondrial genome data of 34 ancient (dotted line) and 105 modern (continuous line) horse samples. (Table S1 available as Supplementary material in Dryad as <http://dx.doi.org/10.5061/dryad.tt78r>), calibrated by adding the donkey (Orlando et al. 2013) as outgroup, considering a split time of 2.75 Myrs  $\pm$  650 kyr BP based on the most credible range estimate for the date when caballine (horse) and non-caballine (donkeys, asses, and zebras) equine lineages became genetically isolated (Jónsson et al. 2014). Bayesian phylogenetic inference was performed with BEAST 1.8.2 (Drummond et al. 2012), considering 5 site categories treated as unlinked partitions, following (Achilli et al. 2012). Divergence was estimated using a log-uncorrelated molecular clock model, including dating of ancient samples as calibration tips. The phylogenetic inference was run for 200 million generations, sampling the chain every 30,000, with 10% burn-in. Convergence was checked with Tracer v1.6.0 (Rambaut et al. 2014), also used for reconstructing the Bayesian skyline demographic profile, plotted with ggplot2 package (Wickham 2009), R (R et al. 2015).

mutation rate and the bottleneck intensity, and, to a lesser extent, on the sample size (Ramakrishnan et al. 2005). The sampling scheme is also crucial. For example, collecting genetic information from individuals just pre-dating and post-dating bottlenecks is essential to detect such events from mtDNA markers, especially when small population sizes are concerned (Mourier et al. 2012). In contrast, nuclear Single Nucleotide Polymorphisms (SNPs) from independent loci provide much higher power to detect bottlenecks with fast recoveries, although moderate bottlenecks generally require an increasing number of loci to be detected (Mourier et al. 2012), and wide, rather than clustered, temporal sampling appears to be more powerful (Skoglund et al. 2014a).

If not properly handled, damage-derived errors in ancient DNA data can lead to flawed demographic reconstructions, by either creating false relationships between samples, or by inflating the number of singleton mutations observed (Fig. 1). Simulations have shown that the latter can lead to spurious demographic

inference, where an initial phase of growth is followed by a relatively recent decline, even though the actual population size was kept constant (Axelsson et al. 2008). To overcome these problems, mutational models that include the specific chemical features of ancient DNA damage have been developed (Ho et al. 2007; Rambaut et al. 2009), as have experimental methods for removing *post-mortem* damage, including independent amplification, multiple sequencing, and enzymatic treatment of ancient DNA extracts before amplification (Hofreiter et al. 2001a; Briggs et al. 2010), or amplification with DNA polymerases that cannot bypass DNA damage (Pedersen et al. 2014).

#### The Serial Coalescent

As previously discussed, phylogenetic methods face important limitations when trying to reconstruct population histories, especially from single genetic markers. Rather than searching for exact historical reconstructions, coalescent theory provides a statistical



framework to compare and fit population demographic models (Kingman 1982; Hudson 1990).

For demographic inference, neutral markers that do not affect reproductive success are generally preferred. Their mutational and genealogical processes can be considered separately (Nordborg 2008), first by building a possible genealogy, then by spreading mutations along the branches according to a given substitution model. Temporal sampling, however, affects the shape of the coalescent tree (Depaulis et al. 2009), and consequently the expected patterns of genetic diversity (Skoglund et al. 2014a), as different samples show different branch lengths. Simulations have shown that the extent of the bias is directly proportional to the temporal range and the sample size, and inversely proportional to the generation time and effective population size (Depaulis et al. 2009). Serial coalescent approaches (Anderson et al. 2005; Excoffier and Foll 2011), which extend the coalescent framework to allow the incorporation of sequentially sampled sequences, are thus recommended. The serial coalescent was originally formulated in 1999 (Rodrigo and Felsenstein 1999) and several programs have been developed to apply this method, following two main frameworks, namely the skyline-plot methods (and their numerous derivatives) and direct coalescent simulations.

#### *Skyline-Plot Methods*

Skyline plots are an application of coalescent theory aimed at the reconstruction of historical population size trajectories, without the need for *a priori* restrictions on possible demographic models (Pybus et al. 2000; Ho and Shapiro 2011). They can be applied to a single locus or multiple independent ones, if aimed at increasing statistical power (Ho and Shapiro 2011), and work in two steps. The first one involves the reconstruction of the gene genealogy, including the tree topology and the branch lengths, often carried out under a Monte Carlo Markov Chain (MCMC) sampling scheme. In the second step, the profile of variations in effective population size over time is estimated, exploiting the density of coalescent events observed in each corresponding time interval. Since there is a relatively simple relationship between effective population size and the probability of coalescence (i.e., the larger the population size, the lower the probability that two sampled alleles are inherited from the same ancestor), this step only depends on the timing of the coalescent events and not on the exact genealogical relationships among the sequences (Fig. 2a) (Pybus et al. 2000). Yet, unaccounted-for population structure (Heller et al. 2013), sequencing errors (Heled and Drummond 2008) and, more specifically, *post-mortem* DNA damage can lead to spurious variations in the effective population size (Axelsson et al. 2008).

The software most commonly used within this framework is BEAST (Drummond et al. 2012), and since it allows serial sampling over different time

periods, it has been used in a variety of ancient—mainly mitochondrial—DNA studies (Ramakrishnan and Hadly 2009). BEAST implements a Bayesian framework to jointly infer any parameter of the associated coalescent, including the tree topology, the dates of the internal nodes, the rate of evolution, and substitution models (Drummond et al. 2002). The latest version of the software, BEAST 2 (Bouckaert et al. 2014) contains an add-on (Gavryushkina et al. 2014) allowing implementation of the above-mentioned FBD model (see section on Calibration and Divergence estimates) (Stadler et al. 2013).

The first and perhaps most iconic study so far analyzed more than four hundred ancient bison mtDNA sequences from northern and eastern Asia and northern America. Bayesian skyline-plot reconstructions show that the bison genetic diversity started decreasing between 45 and 30 kyr ago, concomitant with the onset of the last glacial cycle, and increased again after 10 kyr ago (Shapiro et al. 2004; Drummond et al. 2005). Similar changes in effective population sizes matching glacial cycles have been observed in a range of species sampled for mtDNA across their whole geographic range (see Table 2). The analysis of ancient data allows reconstruction of demographic events for which signals may be absent in modern-day samples due to recent evolutionary histories (e.g., domestication) sweeping away any signatures of earlier expansions/collapses (Fig. 2b).

#### *Sequential Markov Coalescent*

In recent years, another framework for inferring complex population histories has been developed, the so-called SMC (McVean and Cardin 2005; Marjoram and Wall 2006, but see also Wiuf and Hein 1999). The SMC framework is a tractable and accurate simplification of the Ancestral Recombination Graph (ARG), (Hudson 1983), an extended representation of gene genealogies that fully captures the coalescent process with recombination, and enables the study of unlinked markers with different evolutionary histories along the sequence. Integrating over all these markers has increased the statistical power for population inferences, even with a limited number of diploid genome sequences—generally not more than four (Schiffels and Durbin 2014).

SMC has received much interest in ancient DNA research especially through the PSMC, which allows for demographic inferences from a single diploid genome (Li and Durbin 2011). As a proxy for changes in effective population size over time, and similarly to Bayesian skyline plots, the PSMC framework estimates the density of coalescent events over a range of algorithmically selected time intervals. The applicability of the Bayesian skyline plots is, however, limited to unlinked loci without intragenic recombination, while PSMC integrates all the information along the genome. This method was first applied on ancient genomes to estimate the

demographic profile of archaic hominin populations, such as Neanderthals and Denisovans (Prüfer et al. 2014), which revealed extremely reduced population sizes, respectively, about seven or eight times lower than estimated for anatomically modern humans. It has been since applied to a range of other species, for example, revealing in horses and woolly mammoths periods of demographic expansion and collapses concomitant with major climatic changes (Table 2).

Other SMC methods, such as CoalHMM (Mailund et al. 2011), use Hidden Markov Models (HMMs) to partition pairwise or multiple sequence alignments into fragments with different evolutionary histories, from which they evaluate whether population/species pairs diverged in allopatry (isolation) or sympatry (isolation with migration). Such models hold great potential for ancient DNA analyses as they allow for an accurate inference of ancestral population sizes and divergence times, both at the time when population split and when migration stopped. However, they have until now received much less attention (with some exceptions, e.g., Jónsson et al. 2014).

#### *ABC-like Model-Fitting and Serial Coalescent Simulations*

Likelihood-based population models are often analytically complex, and require large computational resources. Complex scenarios are, thus, often analytically intractable. Approximate Bayesian Computation (ABC) is a likelihood-free framework for model selection and estimation of corresponding parameters. When applied to genetic data it allows researchers to circumvent the mathematical and computational problems associated with complex evolutionary scenarios (Beaumont 2010). ABC selects models and parameters minimizing the difference between observed and expected levels and patterns of genetic diversity, as summarized in a set of highly informative statistics. With temporal sampling, the expected summary statistics are simulated using the serial coalescent (Anderson et al. 2005; Excoffier and Foll 2011).

A recent study examining the population history of the American bison integrated serial coalescent simulations based on heterochronous DNA sequences with palaeoclimatology, bioclimatic envelope modeling, and spatio-temporal fossil information (Metcalf et al. 2014). The best-fit scenario supported a history where both climate and human hunting pressure drove population declines. Lorenzen and colleagues (2011) also used ABC coupled with serial coalescent simulations on data from woolly mammoths, reindeer, musk ox, bison, and woolly rhino. They characterized population dynamics in a more detailed fashion than when they applied skyline plots. All species considered except the bison showed an increase in population size in the Late Pleistocene. The time for these expansions varied across species, most being concomitant with climatic changes. Moreover, the comparison of the geographical ranges of humans and the different species analyzed suggested that the extinction of past megafauna was not only

linked to human activity, but that climate also played an important role.

#### GENETIC CONTINUITY OVER TIME

Populations can experience rapid expansions and collapses in effective size, for example through serial founder events at the front wave of expansion. The colonization of new environments not only provides opportunities for the emergence of local adaptations, but may also entail the full or partial replacement of native populations previously inhabiting the region, resulting in a local genetic discontinuity over the whole time scale.

Several methods have been developed to test for genetic discontinuity. Present-day genetic data can be instrumental to date and quantify serial founder events leading to isolation-by-distance patterns. However, they cannot inform us about fully replaced populations, as they went extinct leaving no modern descendants. Serial sampling encompassing the time for the potential population replacements provides a framework for directly testing the hypothesis of genetic continuity over time. Direct ancestry can, for instance, be visually explored with genome projections (Yang and Slatkin 2015), which summarize the differences in derived allele frequencies (DAF) in a test genome relative to a reference panel from a single population (Yang et al. 2014). The difference in DAF is calculated separately for each frequency bin of the unfolded site frequency spectrum (SFS), reporting a pattern that in the case of direct ancestry is unique and distinct from that generated by potentially confounding evolutionary scenarios. The observed pattern consists of a deficit of derived alleles in the ancestral genomes, reflecting the additional evolutionary time leading to descending populations.

As useful as genome projections can be to formulate ancestry hypotheses, they do not represent a formal test for direct ancestry, in contrast to the more widely used serial coalescent simulations within an ABC framework (see the section on ABC-like model-fitting and serial coalescent simulations). Regarding the process of agricultural diffusion in Europe during the Neolithic, a range of studies, first based on ancient mtDNA data (Bramanti et al. 2009; Malmström et al. 2009; Haak et al. 2010; Gamba et al. 2012), but now based on full-genome sequencing of diachronic time series (Skoglund et al. 2012; Gamba et al. 2014; Lazaridis et al. 2014), have highlighted a genetic discontinuity between the Neolithic farmers from different regions and the local Mesolithic hunter-gatherers, supporting a demic diffusion from the Near East (Ammerman and Cavalli-Sforza 1984).

Unlike ABC simulations, formal tests of direct ancestry explicitly model genetic drift, which is often quantified as changes in allele frequency between temporal series. On the basis of this concept, Rasmussen et al. (2014) developed a simple likelihood ratio test demonstrating that the 10.7-kyr-old Anzick-1 infant, excavated in Montana (USA), belonged to the ancestral population of

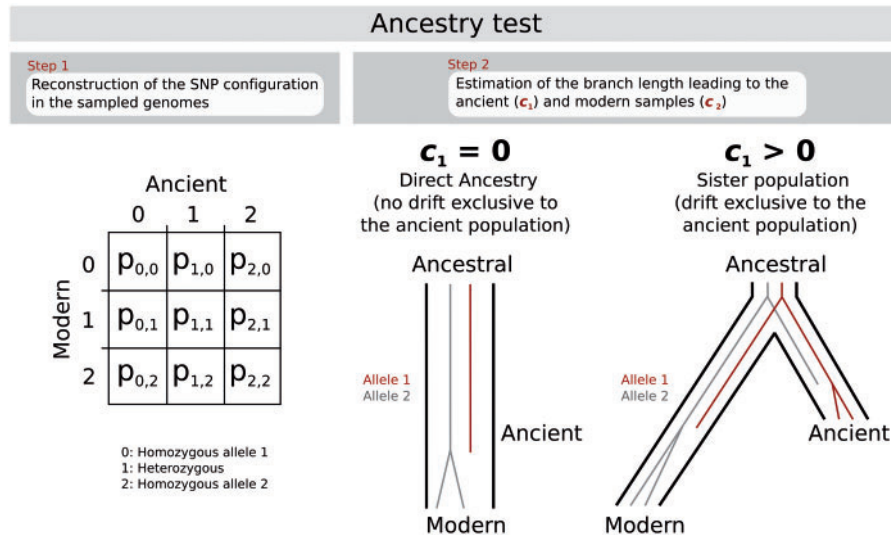


FIGURE 3. Likelihood Ratio Test for direct ancestry implemented in [Rasmussen et al. \(2014\)](#). The alternative and null model differ in a single parameter,  $c_1$ , representing the genetic drift leading to the ancient lineage. The method fully models the 2D-SFS of the observed samples to infer the most likely parameter values, including  $c_1$ . The null model, considering the ancient sample belonging to the population directly ancestral to the modern one, is a particular case of the alternative one, where  $c_1$  is equal to 0.

modern Native Americans ([Rasmussen et al. 2014](#)). The authors first estimated the so-called two-dimensional SFS (2D-SFS), a matrix that in its  $(i, j)$ th entry contains the number of positions with  $i$  alternative alleles in the ancient Anzick-1 genome and  $j$  in the genome of a present-day Native American ( $p_{i,j}$  in Fig. 3). Then, two competing models were fit to the 2D-SFS (Fig. 3), (1) a null model assuming that Anzick-1 belonged to the direct ancestors of Native Americans (Direct Ancestry model) and (2) an alternative model in which the population Anzick-1 belonged to experienced drift since its divergence from the ancestors of the present-day Native Americans (Sister Population model). The data showed a significantly better fit to the Direct Ancestry model, supporting the Anzick-1 population as a direct ancestor to Native Americans. Sjödin and colleagues have also developed a similar method to assess the maximum modern genetic legacy of ancient populations ([Sjödin et al. 2014](#)). This approach considers that an ancient sample belongs to a population that is directly ancestral to a modern one, while allowing for gene flow from an unknown source into either the descending population (the so-called “genetic contribution”) or the parental one (“demographic contribution”). A mathematical framework to quantify the amount of genetic drift expected under these scenarios is derived, and subsequently used within a classical null hypothesis test to reject the scenarios that are incompatible with the observed allele frequency changes.

#### POPULATION STRUCTURE AND ADMIXTURE

The advent of genome-wide population data, especially human genomes, has stimulated the development of robust methods for accurately detecting

and quantifying population structure and admixture. In the following sections, we will first review the so-called “global” methods, estimating the most likely number of ancestral populations (population structure) and their admixing proportions in modern individuals, and then “local” methods, aiming at detecting chromosome segments that introgressed from external population sources ([Alexander et al. 2009](#)).

#### *Multivariate Principal component analysis Approaches*

Principal component analysis (PCA) is part of the classical multivariate statistical tool kit for describing the genetic structure of populations, and was already applied to genetic data some 40 years ago ([Menozzi et al. 1978](#)). More recently, the method was applied to part of the POPRES data set ([Nelson et al. 2008](#)) representing hundreds of thousands of SNPs spread across the genome, revealing a pattern in which the first two principal components mirrored the geographic map of the European continent ([Novembre et al. 2008](#)). Many similar data sets have been produced since, such as the Human Genome Diversity Project (HGPD, [Li et al. 2008](#)) and the Affymetrix Human Origins Panel ([Patterson et al. 2012](#)), the latter designed with no population ascertainment bias.

Interestingly, placing ancient data sets within modern-data PCA plots does not require complete genome sequencing. [Skoglund et al. \(2012\)](#) recovered informative PCA positioning of four ancient Scandinavian samples (three hunter-gatherers and one farmer) with genome coverage of only 0.01–0.035 $\times$ . At biallelic positions in modern-day populations, the authors randomly sampled one allele per individual, transforming all diploid genotypes into pseudo-homozygous genotypes,

and first drew individual PCAs for each of the ancient specimens. They then applied the Procrustes transformation to the two components, to position all four ancient individuals within the same PCA space (Skoglund et al. 2012). This revealed that the three hunter-gatherers were outliers to the present-day variability, while the farmer clustered with southern Europeans, a pattern compatible with population discontinuity between hunter-gatherers and early Scandinavian farmers.

PCAs provide first-exploratory approximations of population affinities, but their interpretation is limited by their non-parametric nature. Suggested population affinities can largely be mistaken due to differential levels of genetic drift, especially when information about the populations' demographic history is missing (Skoglund et al. 2014a). The temporal structure within ancient DNA data sets is also known to impact PCAs, with individuals from the same population background appearing increasingly distant as their temporal differences increases (Skoglund et al. 2014a). Ancestry components and population affinities between individuals are thus better identified with more robust statistical procedures relying on a restricted number of parameters, such as clustering methods.

#### Clustering Approaches

Clustering methods assign individuals to their corresponding ancestral populations, based on their allele frequencies. Briefly, the model assumes  $K$  clusters, representing  $K$  putative ancestral populations with different allele frequencies, and the individuals tested are probabilistically assigned to one or more of clusters based on their genotypes, and assuming Hardy-Weinberg equilibrium at each locus analyzed (Pritchard et al. 2000). Several model-based clustering approaches (i.e., STRUCTURE, Pritchard et al. 2000 and ADMIXTURE, Alexander et al. 2009) provide, for each individual tested, estimates of the genetic fraction descending from each of the  $K$  ancestral populations. The value of  $K$  is set *a priori* and can be selected based on historical information, although more complex scenarios are usually also investigated. When the number of clusters is unknown, it can be inferred as the smallest number of  $K$  (the simplest model) within a credible range, giving the best fit to the data. Such a cross-validation procedure is provided within the program ADMIXTURE (Alexander et al. 2009). Clustering approaches can be directly applied to genotype calls, pseudo-homozygous genotype calls (see Multivariate PCA Approaches), and genotype likelihoods. The latter is particularly convenient when low-coverage (ancient) genomes are examined and is implemented in the NGSAdmix software from the ANGSD suite (Skotte et al. 2013).

Clustering approaches have been extensively used in studies of ancient specimens, including horses (Der Sarkissian et al. 2015b; Librado et al. 2015), maize (da Fonseca et al. 2015), and cattle (Park et al. 2015), but

most frequently to place ancient individuals within the present-day human diversity. The three largest studies so far have generated genome-wide data from Neolithic and Bronze Age Eurasian individuals, either through extensive genotyping of 69 (Haak et al. 2015) and 163 (Mathieson et al. 2015) individuals or through whole-genome sequencing of 101 individuals (Allentoft et al. 2015). These studies provided a diachronic picture of European ancestors and identified a significant ancestry linked to the massive migration from the steppe of the Yamnaya culture, in the late Copper/early Bronze Age (around 5 kyr ago).

Clustering methods do not rely on explicit population models and always assume a defined number of ancestral clusters, which can be ill-suited to complex admixture histories (Patterson et al. 2012). Moreover, global ancestry approaches are generally prone to sampling biases, as uneven representation of individuals across populations can distort the relative frequencies of the different ancestry components (McVean 2009). This is particularly relevant when comparing modern and ancient sequence data sets, as modern genomes are generally overrepresented. For all these reasons, a complement of formal tests for detecting population admixture is required for confirming population affinities as reflected in global analyses.

#### Summary Statistics

Patterson and colleagues (Patterson et al. 2012) formalized into the ADMIXTOOLS package a suite of tests for quantifying ancestral admixture from modern and ancient genomic data. These methods were introduced earlier in a series of publications (Reich et al. 2009; Green et al. 2010; Durand et al. 2011; Moorjani et al. 2011), and include a three-population test (or  $f_3$ -statistic) and two four-population tests, namely the  $D$ -statistics (or ABBA-BABA test), and the  $f_4$ -statistic (the latter also allows the calculation of the  $F_4$ -ratio; see below). These tests have become extremely popular in ancient and modern genomic studies and measure the shared genetic drift between populations and individuals, which is interpreted as shared ancestry. Under this framework, genetic drift is quantified as its impact on the allele frequencies of two related populations,  $H_1$  and  $H_2$ , and denoted as  $F_2(H_1, H_2)$  for theoretical expression, and as  $f_2(H_1, H_2)$  for explicit data quantification (Patterson et al. 2012).

*$f_3$ -statistic.*—This three-population statistic is designed to test whether a population  $H_3$  descends from the admixture of two populations, referred as  $H_1$  and  $H_2$  (Fig. 4). The genetic drift occurring after the population mix, however, gradually erodes the molecular signature of admixture, which limits its applicability to relatively recent episodes (up to 500 generations ago, i.e., approximately 15 kyr for humans; Patterson et al. 2012). Mathematically, the  $F_3$ -statistic can be defined as  $F_3(H_3; H_1, H_2) = F_2(H_3, h_3) - \alpha\beta F_2(h_1, h_2)$ , where  $h_1, h_2, h_3$  refer

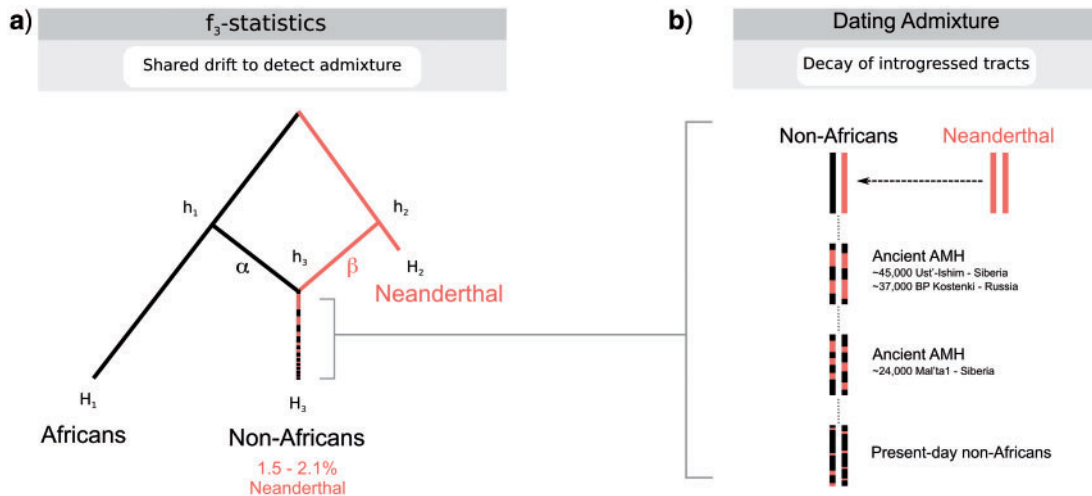


FIGURE 4. Global and local ancestry methods. a)  $F_3$ -statistics correspond to a global ancestry method testing for shared drift. They assess if a test population ( $H_3$ ) descends from two parental populations ( $H_1$  and  $H_2$ ). Here we represent the test carried out for the Neanderthal interbreeding with non-African populations, which estimated 1.5 to 2.1% Neanderthal contribution to the present non-Africans (Prüfer et al. 2014). b) Local ancestry methods infer the date of admixture based on the length distribution of introgressed tracts. Here we show the Neanderthal introgression into the modern non-African population. Three ancient AMH helped dating the admixture to 50–60 kyr ago (Seguin-Orlando et al. 2014; Fu et al. 2016): Ust'-Ishim, a ~45-kyr-old AMH from Siberia (Fu et al. 2014); Kostenki K14, a ~37-kyr-old individual from Russian Federation (Seguin-Orlando et al. 2014), and; Mal'ta 1, a ~24-kyr-old individual from Siberia (Raghavan et al. 2014b). Similar estimates (40–65 kyrs ago) have been found with a different method based on recombination rates (Sankararaman et al. 2012; Moorjani et al. 2016) (see sections on Calibration and Divergence Estimates, and Local Ancestry and Admixture).

to internal nodes,  $F_2(H_3, h_3)$  the drift experienced by  $H_3$  after the admixture event, and  $\alpha$  and  $\beta$  the relative contribution of  $h_1$  and  $h_2$  to the foundation of  $h_3$  (Fig. 4). Whenever  $H_3$  is related to only one population (i.e.,  $\alpha = 1$  and  $\beta = 0$ , or  $\alpha = 0$  and  $\beta = 1$ ) the second term  $\alpha\beta F_2(h_1, h_2)$  is equal to zero and the  $F_3$ -statistic is positive. Conversely, in case of admixture, the  $F_3$ -statistic is negative as long as the amount of shared genetic drift pre-admixture is greater than post-admixture. Therefore, the signature of really ancient admixture events might be lost, and in such cases, analyzing ancient genomes from individuals that lived closer to the time of admixture can help detect signals that would have been missed otherwise. This mathematical definition is however impracticable, because the allele frequencies in the  $h_1$ ,  $h_2$ , and  $h_3$  ancestral populations are often unknown. Operationally,  $f_3$ -statistic values are calculated as a linear combination of  $f_2$ -statistics,  $f_3(H_3; H_1, H_2) = (f_2(H_3, H_1) + f_2(H_3, H_2) - f_2(H_1, H_2))/2$ .

Many examples of negative  $f_3$ -statistic can be found in the ancient DNA literature, revealing that population admixture between ancient human groups was commonplace. For example, interbreeding between Neanderthals and the ancestors of non-African populations was identified using the  $f_3$ -statistic and is estimated to have contributed 1.5–2.1% of the present non-African genomes (Prüfer et al. 2014) (Fig. 4).

In cases where an outgroup is used as test population  $H_3$ , the  $f_3$ -statistic provides a direct estimate for the shared genetic drift between the reference populations  $H_1$  and  $H_2$ , thus reflecting their genetic proximity (Patterson et al. 2012). African genomes are often used as outgroups to estimate the shared genetic drift of

populations outside Africa. For instance, the “outgroup”  $f_3$ -statistic was maximized when pairing the genome of the ~36 kyr-old European from Kostenki (Russian Federation) with modern Europeans, suggesting that the genetic differentiation of Europeans and Asians pre-dated the radiocarbon age of this individual (Seguin-Orlando et al. 2014). The “outgroup”  $f_3$ -statistic was also key in identifying genetic continuity in eastern Africa, between the population represented by the first ancient African genome ever sequenced (Mota; 4,5 kyr ago), and an isolated group of Ethiopian Omotic speakers (Llorente et al. 2015).

*D-statistics and  $f_4$ -statistic.*—The  $D$ -statistics, also known as the ABBA-BABA test, is a four-population test based on evaluating the asymmetries in allele frequencies in a three-population tree topology, given a closely related outgroup (Patterson et al. 2012).

The test was described for the first time by the Neanderthal Genome consortium (Green et al. 2010; Durand et al. 2011) when assessing the presence of archaic hominin ancestry in modern human populations, and works as follows. Let us assume that the phylogenetic relationships between three populations (hereafter referred to as  $H_1$ ,  $H_2$ , and  $H_3$ ) and an outgroup  $O$  is known (e.g.,  $((H_1, H_2), H_3), O$ ) (Fig. 5a). The most common site pattern along the genome is in the form of BBBA, where  $A$  and  $B$ , respectively, represent the ancestral and derived allele states. The second most represented site pattern is BBAA, grouping together  $H_1$  and  $H_2$ . In contrast, ABBA and BABA patterns represent sites not supporting the tree topology (Fig. 5a). In absence of gene flow and structure

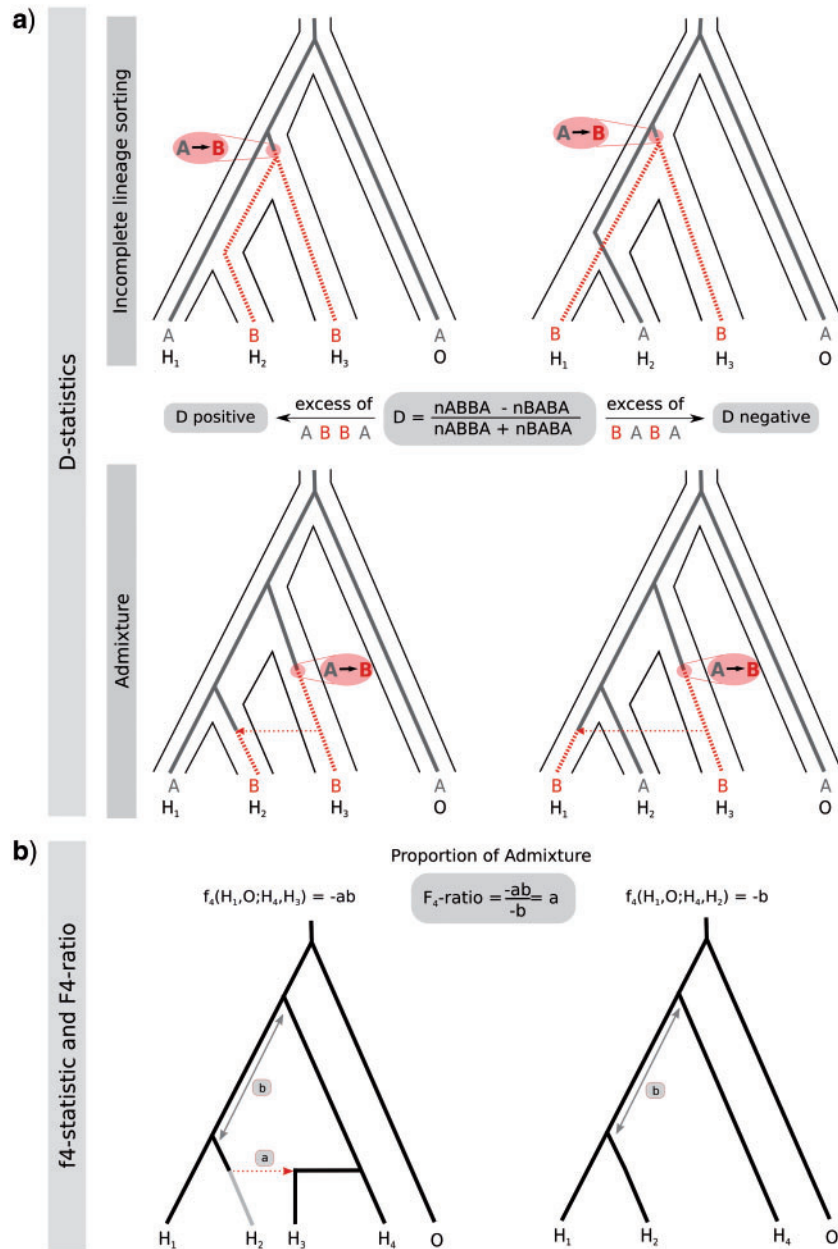


FIGURE 5. *D*-statistics. *D*-statistics are built on loci showing patterns of incomplete lineage sorting (a). A *D*-statistics value deviating from zero indicates an excess of ABBA or BABA events, pointing to a shared ancestry or admixture event with one of the populations tested ( $H_1$  and  $H_2$ ). In case of admixture events, the  $f_4$ -statistic and  $F_4$ -ratio (b) allow quantification of the admixture proportions.

within the population ancestral to  $H_1$ ,  $H_2$ , and  $H_3$ , such patterns are by-products of incomplete lineage sorting, and are expected to appear in balanced proportions. The *D*-statistics measures the difference between the number of sites in ABBA and BABA topologies over their sum, and is therefore expected to be equal to zero. In contrast, positive (negative) *D*-statistics point to an overrepresentation of either the ABBA (BABA) patterns, which suggests gene flow between  $H_3$  and  $H_2$ , (between  $H_3$  and  $H_1$ ). To account for the standard error, the *D*-statistics are transformed into Z-scores, that is, their value is transformed into the number of standard

deviations, with  $|Z| > 3$  indicating significant levels of admixture (Rasmussen et al. 2011; Orlando et al. 2013).

Green and colleagues (Green et al. 2010) retrieved positive  $D((\text{Non-African}, \text{African}), \text{Neanderthal}, \text{Chimpanzee})$  values for all pairs of present-day African and non-African populations, suggesting Neanderthal affinities to non-Africans, probably linked to interbreeding between Neanderthal and the ancestors of Eurasians (confirming  $f_3$ -statistic results; Fig. 4). The detection of gene flow between Neanderthals and Anatomically Modern Humans (AMHs) is, however, challenged. Possible ancient population structure in

Africa, where Neanderthals and AMHs might originate from a similar population background, could also generate non-null  $D$ -statistics in absence of gene flow (Green et al. 2010; Eriksson and Manica 2012). However, non-African populations appear to share more derived alleles with the Neanderthal specimen from the Caucasus than the specimens from the Altai Mountains or Croatia, which supports a population scenario where particular Neanderthal subgroups admixed with non-African humans (Prüfer et al. 2014).

Meyer et al. (2012) also tested for gene flow between archaic hominins ( $H_3$ ) and non-African human populations ( $H_1$  and  $H_2$ ), using the chimpanzee as outgroup. To remove potential spurious signals due to substructure in the ancestral population leading to  $H_1$  and  $H_2$ , Meyer et al. (2012) introduced the enhanced  $D$ -statistics, conditioned to sites where the ancestral population of  $H_1$  and  $H_2$  (represented by sub-Saharan Africans) have the same allele as the chimpanzee. This revealed the presence of Denisovan ancestry in Papuans (Meyer et al. 2012), a result that was further supported by the results of a TreeMix analysis (see the section on admixture graphs).

The  $F_4$ -statistic, an analogue of the  $D$ -statistics, has been developed to detect gene flow and estimate the admixture proportions (Reich et al. 2009; Patterson et al. 2012; Sankararaman et al. 2012). Assuming a  $((H_1, H_2), (H_3, H_4), O)$  topology, where  $H_3$  represents the population potentially admixed with  $H_2$  (Fig. 5b),  $f_4(H_1, O; H_4, H_3)$  can be estimated, similarly to  $f_3$ , from the population allele frequencies. Particularly as  $f_4(H_1, O; H_4, H_3) = (p_1 - p_O)(p_3 - p_4)$ , which intuitively corresponds to the overlap between the paths going from  $H_1$  and  $O$ , and from  $H_2$  to  $H_3$ . This is represented as  $-ab$  (Fig. 5b), where  $b$  is the amount of drift exclusively shared by  $H_1$  and  $H_2$ , and  $a$  the fraction introgressed between  $H_2$  and  $H_3$ . The benefit of this is that  $f_4(H_1, O; H_4, H_2)$  is simply equal to  $-b$  because it does not include the test population  $H_3$ . Therefore, the ratio computed between the two  $F_4$ -statistics  $F_4(H_1, O; H_4, H_3)$  and  $F_4(H_1, O; H_4, H_2)$ , called the  $F_4$ -ratio, equals  $a$ , providing a direct estimate of the admixture proportion between  $H_3$  and  $H_2$  (Fig. 5b).

While  $F$ -statistics should ideally be restricted to robust population allele sampling,  $D$ -statistics are more flexible, as they can be implemented even when only one sequence per taxon is available (Pease and Hahn 2015). Given the robustness of both tests to various SNP ascertainment biases (Patterson et al. 2012), they have been used in a number of cases involving ancient populations, including AMHs (Fu et al. 2014, 2015; Seguin-Orlando et al. 2014), archaic hominins (Green et al. 2010; Meyer et al. 2012; Prüfer et al. 2014), horses (Schubert et al. 2014b; Der Sarkissian et al. 2015b; Librado et al. 2015), aurochs and cattle (Park et al. 2015).

#### Admixture Graphs

Admixture graphs represent natural extensions of phylogenetic trees to accommodate exchanges of genetic

material resulting from secondary contacts between diverged populations. The most popular method for building admixture graphs builds on the drift-based  $F$ -statistics described above, generalizing a framework that has been efficiently implemented in several tools scalable up to a moderate number of populations. MixMapper, for example, implements a semi-automated method that first builds a neighbor-joining tree from the  $F_2$  distances calculated between clearly non-admixed populations, and subsequently adds the remaining populations as two- or three-way mixtures of the non-admixed populations (Lipson et al. 2013, 2014). TreeMix implements a conceptually similar approach, though with multiple operational differences that enable the full automation of the admixture graph estimation, given a user-fixed number of gene flow events (or migration edges) (Pickrell and Pritchard 2012). This number should be carefully selected, as forcing inadequate migration event edges may induce spurious changes in the tree topology.

The scalability of such programs, especially TreeMix, has been key in ancient DNA studies, as they allow rapidly and jointly screening for genetic affinities and mixture events between ancient samples and present-day populations. MixMapper has been used in a number of ancient DNA studies, showing that a significant fraction of Native American ancestry originates from pre-Last Glacial Maximum (LGM) east Asian populations (Raghavan et al. 2014b) and that Denisovans have contributed to the genomes of modern Melanesians, Australian aborigines, and other Southeast Asian islanders (Meyer et al. 2012). More recently, TreeMix has been applied to study a range of population histories, including the peopling of the New World Arctic (Raghavan et al. 2014a), and revealed genetic affinities between aurochs and local cattle landraces from the British isles (Park et al. 2015).

#### Local Ancestry and Admixture Dates

Admixture among sexually reproducing populations results in mosaic genomic regions of different ancestries. Identifying such regions is fundamental not only for the characterization of migration and cross-species gene flow (Green et al. 2010; Reich et al. 2010; Prüfer et al. 2014), but also for the identification of recombination patterns and local adaptations (Kim and Rothschild 2014). This is instrumental to uncover the genetic architecture of reproductive isolation, to detect the so-called “speciation genes” (Wu and Ting 2004; Nosil and Schluter 2011; Burri et al. 2015), and can be applied to the management of endangered populations (Der Sarkissian et al. 2015b).

Some approximations for mapping the ancestry of genomic tracts build on genome-wide sliding window analyses. In one popular method,  $D$ -statistics are calculated within sliding windows of a given size, and candidates for introgression are identified in blocks showing significantly non-null  $D$ -statistics (Kronforst et al. 2013; Smith and Kronforst 2013). However, the variance of the  $D$ -statistics is extreme in small windows,

mainly due to the presence of a rather limited amount of unlinked information (Martin et al. 2015). This increases the rate of non-null statistics, and creates spurious correlations with recombination rates. To avoid this, the minimum size of a sliding windows should be dynamically defined according to  $L=1000/(N_e*r)$ , where  $N_e$  is the effective population size and  $r$  the recombination rate per nucleotide and generation (Pease and Hahn 2015). This restriction typically requires using regions on the order of megabases, a magnitude that is obviously unsuitable for fine-scale local ancestry inference.

More dedicated implementations of the problem of local ancestry inference have been thoroughly reviewed in Padhukasahasram (2014). These methods generally use HMM to define the genomic boundaries of introgressed regions inherited from parental populations (Moorjani et al. 2011; Patterson et al. 2012; Harris and Nielsen 2013). The length of admixture-induced LD segments is also extremely valuable to date admixture events, as recombination breaks down introgressed tracts, producing an exponential decay of genomic blocks with the number of generations, which can effectively operate as a "recombination clock." Although extremely powerful, these methods are not without caveats, and typically require specification of source populations, which are often unknown, extinct, or at best surrogates of the parental ones. Additionally, the asymptotic decay of the recombination clock compromises its applicability to date admixture events after a few hundreds to thousands of generations (Loh et al. 2013).

Using ancient genomes from individuals temporally closer to the date of the gene flow event can minimize or even circumvent these limitations (Fig. 4) whether they belong to the source (parental) or to the descending (admixed) population. For example, the genomes of 36–45 kyr-old modern humans were found to show much longer tracts of Neanderthal ancestry than present-day modern humans, which helped refine the time estimates of Neanderthal admixture to ~55 kyr ago (Fu et al. 2014; Seguin-Orlando et al. 2014), later corrected to ~50–60 kyrs ago (Fu et al. 2016). This value is in accordance with the ones calculated based on recombination rates (see section on Calibration and Divergence Estimates): ~47–65 kyr ago (Sankararaman et al. 2012) and ~40–55 kyr ago (Moorjani et al. 2016).

#### TRACKING GENETIC SIGNATURES OF ADAPTATION

Detecting signatures of positive selection is of major interest in evolution, as it is the only evolutionary force that can explain environmental adaptation. Such analyses are generally carried out by comparing the levels and patterns of genetic variability within and/or between different species, which enables detection of episodes of selection at various time scales (Sabeti et al. 2006).

By charting allele frequencies through time, ancient DNA provides an opportunity to identify the pace of adaptation, and thereby determine the age and selection coefficients of an adaptive mutation (Bollback et al. 2008; Malaspinas et al. 2012; Malaspinas 2016; Schraiber et al. 2016). This approach uncovered the tempo and mode of the selective sweep to an unprecedented resolution. For example SNPs at six genes associated with coat color were tracked through time in 89 ancient horses, revealing an explosion in allelic diversity following the early Bronze Age, just after the onset of horse domestication (Ludwig et al. 2009). The dun and bay horses no longer dominated the Eurasian steppes, as black, white, and leopard-spotted horses started reaching significant frequencies in bone assemblages. Further work on the allele associated with leopard spotting revealed dynamic selection regimes, with herders apparently favouring this mutation in the early Bronze Age, but selected against it in the late Bronze Age (Ludwig et al. 2014). As the leopard-spotting associated allele is also a recessive mutation responsible for congenital stationary night blindness in horses, it shows how human preference for a particular phenotype can artificially select and maintain polymorphisms that would otherwise be purged from the population.

Another example utilizing the same approach concerns the origins of lactase persistence (LP) in the human gut in adulthood, conferring the ability to digest milk after weaning. In Europe LP is highly associated with a specific SNP located ~14k bp upstream of the *LCT* gene, and shows a strong selection signature in modern European populations, reaching almost fixation frequencies in Scandinavia (Leonardi et al. 2012). It has long been debated if LP-associated alleles spread before the introduction of milk consumption in adulthood, or if the cultural behavior of milking started before the ability to fully digest it, entailing a selective pressure favoring the mentioned variants. Ancient DNA data from Europe show that the Neolithic European population was not able to digest fresh milk (Burger et al. 2007; Malmstrom et al. 2010; Lacan et al. 2011a, 2011b; Plantinga et al. 2012; Gamba et al. 2014; Lazaridis et al. 2014; Sverrisdóttir et al. 2014; Olalde et al. 2015) supporting the hypothesis that the increase of the LP trait was a consequence, rather than a cause, of milking behavior. Furthermore the frequency of LP in Bronze Age Europeans was recently revealed to be low (<10%) (Allentoft et al. 2015; Mathieson et al. 2015), suggesting that selection for this allele took place later than the 7.5 kyr estimated by Itan et al. (2009). Additionally the highest frequency of the LP-associated allele was found amongst the Bronze Age steppe people, particularly the Yamnaya. This suggests that European LP possibly originated on the steppes (Allentoft et al. 2015), and not from Central Europe, as previously proposed based on patterns of genetic variation observed in living European populations (Itan et al. 2009).

Beyond extensive analyses of the temporal allele dynamics at key loci, the characterization of single genomes from a closely related extinct species (or



lineage) can provide invaluable information regarding selective regimes within present-day populations. This is illustrated by the Neanderthal genome, which illuminated the genetic makeup of anatomically modern humans (Green et al. 2010; Prüfer et al. 2014; Pääbo 2015), and in horses, where a now extinct, but closely related lineage helped better understand the genetic foundation of domesticated horses (Schubert et al. 2014b). Briefly, such studies screened for regions where the genomes of modern species (lineages) lost significant amounts of genetic variability, as this represents the molecular signature of selective sweeps. In practice, these authors used SMC to implement a HMM that categorized the sequence alignment in regions compatible or incompatible with genealogies resulting from a sweep in the focal lineage. To avoid the confounding effect of decreased variability from purifying selection, it was furthermore required that the genetic variability had largely recovered through subsequent neutral evolution, resulting in the accumulation of alleles that are not shared with their close relatives. Green et al. (2010) applied this method to the Neanderthal genome and identified 212 human-specific selective sweeps, spanning large genetic distances, and including genes involved in diet metabolism, morphology, and cognition. Five out of the 20 clearest signals of positive selection did not include, however, any known protein-coding gene, suggesting that recent selective sweeps in humans may be associated with regulatory changes rather than changes in protein-coding genes (Green et al. 2010).

This approach was also implemented to unravel domestication genes in horses (Schubert et al. 2014b), together with a number of complementary scans to minimize the number of false positives. One such scan relied on the ratio of the Watterson estimator ( $\theta_w$ ) (Watterson 1975) calculated for a set of domestic horses and for a pair of pre-domestic horses. By using this ratio, regions in which the modern population showed a decrease in genetic diversity could be identified, predicated on a corresponding deviation from neutrality detected using the Tajima's  $D$  statistic (Tajima 1989). The resulting set of 125 candidate domestication genes contained many related to the nervous system and behavioral traits, including learning and fear responses, in addition to genes involved in the circulatory system, metabolism, and muscular and skeletal structure.

Fumagalli et al. 2013). The greater uncertainty is partly due to the presence of sequencing errors, and in the case of ancient DNA, molecular damage-related misincorporations, both potentially leading to heterozygous SNPs being called at homozygous sites. This can clearly bias the estimates of the SFS, which summarizes the frequencies of segregating variants within a population. To solve this problem, Nielsen et al. (2012) proposed to jointly infer the SFS of the sampled individuals from the genotype likelihoods, given the probability of each possible genotype at a given position, rather than from their discrete genotype calls (Nielsen et al. 2012; Vieira et al. 2013). Fumagalli et al. (2013) subsequently proposed a Bayesian framework for estimating  $F_{ST}$  coefficients from posterior probabilities of the SFS. This method was found to perform substantially better than methods based on genotype calling, as examined using simulated and empirical data. Vieira et al. (2013) further expanded on this work by developing methods for calculating per-individual ( $F_{ind}$ ) and per-site ( $F_{site}$ ) estimates of inbreeding coefficients directly using genotype likelihoods from low-coverage HTS data.

Tools allowing a range of analyses on low-coverage genomes in the genotype-likelihood framework are available in the open-source software Analysis of Next Generation Sequencing Data (ANGSD) (Korneliusson et al. 2014). Analyses that can be performed in ANGS range from estimation of population and individual statistics (SFS, Watterson's  $\theta_w$ , Tajima's  $D$ , inbreeding coefficients, etc.), to population tests (ABBA-BABA, PCA, etc.—see above). This software has, therefore, been used in a number of ancient DNA studies, including for calculating the genome-wide distribution of heterozygosity and Tajima's  $D$  within user-defined genomic regions (Orlando et al. 2013; Jónsson et al. 2014; Schubert et al. 2014b), allele frequencies, inbreeding and  $F_{ST}$  coefficients in modern and museum specimens of honey bees (Mikheyev et al. 2015), maize (da Fonseca et al. 2015) and horses (Der Sarkissian et al. 2015b; Mikheyev et al. 2015), and estimating error and contamination rates (Raghavan et al. 2015).

While support for likelihood-based genotyping methods and associated downstream analyses is, as of yet, limited, this methodology promises to further increase the reach of ancient genomics, allowing for the study of older and lower quality samples.

## GENOTYPING ANCIENT GENOMES

Genomic analyses often rely on the ability to accurately call genotypes for the samples being studied. However, ancient genomics is mostly carried out using HTS data at low to medium depths of coverage ( $<20\times$ ), with a few notable exceptions (e.g., Meyer et al. 2012; Fu et al. 2014; Gamba et al. 2014; Lazaridis et al. 2014; Prüfer et al. 2014; Schubert et al. 2014b; Librado et al. 2015). With lower coverage comes greater statistical uncertainty associated with genotype calling, which can impact downstream inferences (Nielsen et al. 2012;

## ANCIENT METAGENOMICS

### *Ancient Metabarcoding*

Ancient remains and environmental samples often contain DNA from multiple organisms, such as microbes, plants, or animals, either arising from the ecosystem or infecting the analyzed specimens. This biodiversity can be revealed through sequence analyses based on strict identity or similarity in short, but informative, genetic markers, or "metabarcodes." These are recovered by PCR amplification of DNA fragments

selected to optimize the resolution of taxonomic assignment.

Metabarcoding has been applied to a variety of ancient environmental samples collected from diverse locations (see Pedersen et al. 2015 for a review) and dated to up to 400 kyr ago (Willerslev 2003). These encompass permafrost soils (Willerslev 2003; Haile et al. 2009; Epp et al. 2012; Jørgensen et al. 2012; Hou et al. 2014; Willerslev et al. 2014), cave deposits (Hofreiter et al. 2003; Haile et al. 2007; Murray et al. 2013; Haouchar et al. 2014; Grealy et al. 2015), ice cores (Willerslev et al. 1999, 2007), plant remains (Gould et al. 2010), lake and fjord sediments (Matisoo-Smith et al. 2008; Anderson-Carpenter et al. 2011; Parducci et al. 2012, 2013; Pedersen et al. 2013; Giguet-Covex et al. 2014; Pawłowska et al. 2014; Pansu et al. 2015). DNA in these samples derives from the deposition of plants, eukaryotes and prokaryotes, as well as from egesta (i.e., feces, urine) of animals in the past ecosystem.

The power of this approach is perhaps best illustrated with ancient sediments, where it revealed the dynamic interplay between climate, human activity, and the environment. For example, 50 kyr of vegetation history could be reconstructed through plant metabarcoding of 42 dated permafrost sediments spanning 21 sites across the Arctic (Willerslev et al. 2014). The analyses revealed forbs-dominated ecosystems before and during the LGM, and a turnover in favor of graminoids post-LGM. Similarly, ancient metabarcoding applied to the sediments from the subalpine Lake Anterne, in the North French Alps, helped reveal the temporal distribution of plants and common domestic herbivorous species (Giguet-Covex et al. 2014; Pansu et al. 2015), showing an intensification of cattle and sheep herding at the end of the Iron Age and in the early Roman period (2.4–1.9 kyr ago) and during the Medieval-Modern period (150–1200 years ago) (Giguet-Covex et al. 2014). Such analyses are not limited to classical vertebrate, plant, and bacterial metabarcodes as illustrated by D'Costa and colleagues (2011), who could detect variants of the antibiotic resistance genes *vanX*, *tetM*, and *bla* in 30 kyr-old permafrost sediments from Beringia, showing that resistance predates by millennia the clinical use of antibiotics (D'Costa et al. 2011).

Ancient metabarcoding has also been used on ancient and modern dental calculus to evaluate whether important dietary changes in human evolution have influenced the structure of human oral microbiomes, and potentially impacted human health (Adler et al. 2013). The analyses revealed a significantly higher incidence of species responsible for caries and periodontal diseases in the Neolithic, probably linked to a shift toward a diet richer in carbohydrates. This study also reported a drastic decrease in oral microbial biodiversity in the modern sample, paralleled by the emergence of oral microbial taxa known to cause chronic diseases. This was attributed to dietary changes during the 18th–19th century Industrial Revolution, which introduced refined grains and concentrated sugars (Adler et al. 2013).

#### *Ancient Shotgun Sequencing-Based Metagenomics*

With ever decreasing sequencing costs and growing sequencing capacities, shotgun sequencing approaches are increasingly applied to the metagenomic analyses of ancient samples, potentially revealing the whole population of DNA molecules present in the DNA extract.

Applied to ancient human samples, such approaches have helped detect the *Mycobacterium tuberculosis* pathogen in 26 ancient human remains found in a Hungarian church crypt dated to the 18th century CE (Kay et al. 2015), and previously identified as victims of tuberculosis (Fletcher et al. 2003; Chan et al. 2013). The analyses revealed 12 different strains falling within the modern diversity of the European-American *M. tuberculosis* Lineage 4. Consistent times to the Most Recent Common Ancestor of 470 and 396 CE were estimated, placing the origins of Lineage 4 in the Late Roman Period. Bayesian skyline plots of this lineage showed an increase in effective population size until the beginning of the 20th century, when the population declined. This study also revealed infectious patterns indicating that multiple strains could simultaneously infect a single individual, which is rarely reported today. Another work by Rasmussen et al. (2015b) focused on *Y. pestis*, the etiological agent of plague. They detected the bacteria in 7 Asian and European human specimens ranging in age from 2.8 to 5.0 kyr, that is, three millennia before the earliest historical record of infection. Rasmussen et al. (2015b) provided several lines of evidence for the authenticity of their *Y. pestis* sequences: (1) the presence of damage patterns characteristic of ancient DNA; (2) the presence of three plasmids of *Y. pestis* (pCD1, pMT1, and pPCP1); (3) uniform coverage along the *Y. pestis* reference genome; and (4) closer read to reference edit distances with the reference genome of *Y. pestis* than any of its closest relative, *Y. pseudotuberculosis*. The authenticated sequences eventually allowed identification of temporal patterns for the acquisition of the genes responsible for the high pathogenicity of *Y. pestis* and flea transmission, and for its ability to escape the mammalian innate immune system.

Even though the latter studies focused on the genetic characterization of single microorganisms, shotgun metagenomics can also be applied to investigate the entire microbial diversity preserved in ancient samples as shown by Warinner et al. (2014), who identified bacteria typical of the human oral microbiome in medieval dental calculus from two adult individuals (Warinner et al. 2014).

#### *Analytical Tools for Ancient Metabarcoding*

A plethora of programs have been developed to perform metagenomic analyses of both metabarcoding and shotgun sequencing data sets (reviewed in e.g., Sharpton 2014; Hoffman and Matsen 2015). A family of taxonomic assignment methods is

based on DNA sequence similarity between the query and taxonomically annotated sequences from comprehensive reference databases. One such method, the Basic Local Alignment Search Tool (BLAST; Altschul et al. 1990, 1997) has been widely applied in ancient metagenomic studies, with subsequent analyses of taxonomic profiles sometimes carried out using MEGAN (Huson et al. 2007). BLAST-based taxonomic assignment, however, suffers from a number of limitations (Munch et al. 2008b), including species mis-identification, and mapping scores based on local alignments that do not directly measure the statistical significance of the assignment. Assignment methods based on phylogenetic criteria, and in particular the Statistical Assignment Package (SAP), were developed to address the latter limitation. SAP uses an automated method based on Bayesian phylogenetics for clustering sequences within supported taxonomic clades (Munch et al. 2008a, 2008b). Other classification methods based on similarities in base composition or k-mer frequencies have been applied to modern data sets, (see e.g., Sharpston 2014; Hoffman and Matsen 2015) but their potential for analyzing ancient metagenomic data sets is currently unknown.

Ancient metagenomic analyses have been carried out using computational pipelines, such as OBITools (<https://git.metabarcoding.org/obitools/obitools/wikis/home>) and QIIME for microbes (Caporaso et al. 2010), which perform all analytical steps from post-sequencing DNA read processing, sequence comparison to reference databases and clustering, to taxonomic assignment, visualization, and statistical analyses. In addition, the program SourceTracker, which is part of the QIIME package, allows prediction of the source of microbial taxa representing distinct microbial communities within a microbiome (Knights et al. 2011). It uses 16S rRNA data from different environments tested as sources, and directly estimates the proportions of contaminants from each source in the sink microbiome using Bayesian modeling. Applied to ancient coprolites from three sites (Tito et al. 2012), SourceTracker revealed variable preservation levels for the original gut microbiomes. While one sample was found to be representative of gut microbiomes typical of rural individuals, extensive microbial contamination from compost-like or unknown sources was identified in the other two samples.

#### *Limitations of Ancient Metagenomics*

Although ancient metabarcoding can help describe past biodiversity in a range of samples and organisms, and across wide temporal scales, this approach suffers from several limitations. First, the regions selected as metabarcodes should enable an even representation of all taxa and minimize false negatives/positives. False negatives can result from limited template availability, which is common in ancient DNA, or from the preferential amplification of particular DNA

templates. The latter has been recently demonstrated when classical 16S barcodes was applied to ancient DNA extracts from dental plaque, and resulted in skewed bacterial taxonomic profiles (Zieseemer et al. 2015). This advocates for the use of shotgun sequencing to retrieve more accurate microbial profiles. False positives can result from contamination, nucleotide mis-incorporations at damaged nucleotides, and more generally from sequence errors, but can be detected through multiple replications and controls (Ficetola et al. 2015). Second, the evolutionary relationships among metabarcodes do not necessarily reflect the true evolutionary relationships among species, for instance due to gene flow, incomplete lineage sorting, or horizontal gene transfer. Third, as so far restricted to taxonomic assignment, ancient metabarcoding analyses do not provide any information about the functional content of an environmental sample. Finally, the accuracy of taxonomic assignment requires high-quality comparative databases, and databases skewed toward specific taxonomic groups can mislead species-level assignment. Importantly, a large part of the past and present biodiversity still remains undescribed at the molecular level and taxonomic assignment methods should take the incompleteness of the databases into account, for example, correcting assignment scores for the possibility of unobserved species (Munch et al. 2008a).

#### *Efficient and Accurate Microbial Profiling from Ancient Samples*

Considering the limitations of ancient metabarcoding (Zieseemer et al. 2015), it is expected that microbial profiling of ancient specimens will mainly rely on HTS in the near future. However, HTS data sets are rapidly increasing in size, causing growing computational challenges. Moreover, the assignment accuracy can be limited by the short read length of ancient DNA templates. These problems are particularly acute for characterizing ancient microbial diversity from shotgun HTS data using BLAST-based approaches (Warinner et al. 2014), but can be overcome with the program MetaPhlAn (Segata et al. 2012; Truong et al. 2015). The MetaPhlAn database consists of a collection of specific taxonomically informative markers identified from pairwise genome comparisons of characterized microbial genomes. It is, however, not immune to representation bias, as the vast majority of environmental microbes is still largely unknown, although likely to improve in the shorter term due to decreasing sequencing costs and the multiplication of microbial genome sequencing projects.

MetaPhlAn was first applied to ancient shotgun HTS data to characterize the microbial diversity present in the DNA extracts from 7 horse remains dated back between 13 kyrs and 200 years ago (Der Sarkissian et al. 2014), representing a recent colonisation from the depositional soil. MetaPhlAn and the series of statistical analyses used

to compare microbial profiles in Der Sarkissian et al. (2014) have been implemented in two pipelines: as a module of the PALEOMIX package, dedicated to whole-genome resequencing analyses (Schubert et al. 2014a), and in metaBIT, which is entirely dedicated to microbial profiling (Louvel et al. in press). The approach has so far been applied to a limited number of archaeological, herbarium, and museum specimens (Schubert et al. 2014a, 2014b; Der Sarkissian et al. 2015b; Librado et al. 2015; Seguin-Orlando et al. 2015) but routine microbial profiling using PALEOMIX or metaBIT on a diversity of ancient DNA extracts will, in the future, enhance our understanding of depositional environments and their relationship with taphonomic processes, past diets and health statuses, and help identify contamination sources.

#### CONCLUSIONS

The recent development of new molecular methods, most remarkably HTS technologies, has driven significant advances in ancient DNA research. This review highlights how these advances have led to the implementation of dedicated models and novel tools in statistical palaeogenomics, some of which have even become standard for the analysis of present-day genetic variability. The recent findings described here illustrate how these methods exploit the information provided by ancient DNA sequences to circumvent some of the known limitations in modern genetic analyses, and help reconstruct complex population histories and adaptive processes. With the ever-increasing success in retrieving DNA from ever-larger numbers of ancient samples, the future for ancient DNA research has perhaps never been as promising.

#### SUPPLEMENTARY MATERIAL

Supplementary material can be found in the Dryad data repository at <http://dx.doi.org/10.5061/dryad.tt78r>.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### FUNDINGS

This work was supported by the Danish Council for Independent Research, Natural Sciences (Grant 4002-00152B); the Danish National Research Foundation (Grant DNRF94); Initiative d'Excellence Chaires d'attractivité, Université de Toulouse (OURASI); the International Research Group Program (Grant IRG14-08), Deanship of Scientific Research, King Saud University; a Marie-Curie Individual Fellowship (MSCA-IF-657852); a Marie-Curie Intra-European fellowship (FP7-IEF-328024); and a Villum Fonden Blokstipendier grant.

#### REFERENCES

- Achilli A., Olivieri A., Soares P., Lancioni H., Kashani B.H., Perego U.A., Nergadze S.G., Carossa V., Santagostino M., Capomaccio S., Felicetti M., Al-Achkar W., Penedo M.C.T., Verini-Supplizi A., Houshmand M., Woodward S.R., Semino O., Silvestrelli M., Giulotto E., Pereira L., Bandelt H.-J., Torroni A. 2012. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc. Natl. Acad. Sci. USA* 109:2449–2454.
- Adler C.J., Dobney K., Weyrich L.S., Kaidonis J., Walker A.W., Haak W., Bradshaw C.J.A., Townsend G., Sołtysiak A., Alt K.W., Parkhill J., Cooper A., et al. 2013. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* 45:450–455.
- Adler C.J., Haak W., Donlon D., Cooper A. 2011. Survival and recovery of DNA from ancient teeth and bones. *J. Archaeol. Sci.* 38:956–964.
- Alexander D.H., Novembre J., Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664.
- Allentoft M.E.E., Sikora M., Sjögren K.-G., Rasmussen S., Rasmussen M., Stenderup J., Damgaard P.B., Schroeder H., Ahlström T., Vinner L., Malaspina A.-S., Margaryan A., Higham T.F.G., Chivall D., Lynnerup N., Harvig L., Baron J., Casa P. Della, Dąbrowski P., Duffy P.R., Ebel A. V., Epimakhov A., Frei K., Furmanek M., Gralak T., Gromov A., Gronkiewicz S., Grupe G., Hajdu T., Jarysz R., Khartanovich V., Khokhlov A., Kiss V., Kolář J., Kriiska A., Lasak I., Longhi C., McGlynn G., Merkevicius A., Merkyte I., Metspalu M., Mkrtychyan R., Moiseyev V., Paja L., Pálfi G., Pokutta D., Pospieszny L., Price T.D., Saag L., Sablin M., Shishlina N., Smrčka V., Soenov V.I., Szeverényi V., Tóth G., Trifanova S. V., Varul L., Vicze M., Yepiskoposyan L., Zhitenev V., Orlando L., Sichert-Pontén T., Brunak S., Nielsen R., Kristiansen K., Willerslev E. 2015. Population genomics of Bronze Age Eurasia. *Nature.* 522:167–172.
- Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
- Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Ammerman A.J., Cavalli-Sforza L.L. 1984. *The Neolithic transition and the genetics of populations in Europe.* Princeton University Press.
- Anderson-Carpenter L.L., McLachlan J.S., Jackson S.T., Kuch M., Lumibao C.Y., Poinar H.N. 2011. Ancient DNA from lake sediments: Bridging the gap between paleoecology and genetics. *BMC Evol. Biol.* 11:30.
- Anderson C.N.K., Ramakrishnan U., Chan Y.L., Hadly E. a. 2005. Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics.* 21:1733–1734.
- Avise J.C., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A., Saunders N.C. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* 18:489–522.
- Axelsson E., Willerslev E., Gilbert M.T.P., Nielsen R. 2008. The effect of ancient DNA damage on inferences of demographic histories. *Mol. Biol. Evol.* 25:2181–2187.
- Beaumont M.A. 2010. Approximate Bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Syst.* 41:379–406.
- Bollback J.P., York T.L., Nielsen R. 2008. Estimation of 2Nes from temporal allele frequency data. *Genetics.* 179:497–502.
- Bos K.I., Harkins K.M., Herbig A., Coscolla M., Weber N., Comas I., Forrest S.A., Bryant J.M., Harris S.R., Schuenemann V.J., Campbell T.J., Majander K.K., Wilbur A.K., Guichon R.A., Wolfe Steadman D.L., Cook D.C., Niemann S., Behr M.A., Zumarraga M., Bastida R., Huson D., Nieselt K., Young D., Parkhill J., Buikstra J.E., Gagneux S., Stone A.C., Krause J. 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature.* 514:494–497.
- Bos K.I., Herbig A., Sahl J., Waglechner N., Fourment M., Forrest S.A., Klunk J., Schuenemann V.J., Poinar D., Kuch M., Golding G.B., Dutour O., Keim P., Wagner D.M., Holmes E.C., Krause J., Poinar H.N. 2016. Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *Elife.* 5: e12994.

- Bos K.I., Schuenemann V.J., Golding G.B., Burbano H.A., Waglechner N., Coombes B.K., McPhee J.B., DeWitte S.N., Meyer M., Schmedes S., Wood J., Earn D.J.D., Herring D.A., Bauer P., Poinar H.N., Krause J. 2011. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature*. 478:506–510.
- Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537.
- Bramanti B., Thomas M.G., Haak W., Unterlaender M., Jores P., Tambets K., Antanaitis-Jacobs L., Haidle M.N., Jankauskas R., Kind C.-J., Lueth F., Terberger T., Hiller J., Matsumura S., Forster P., Burger J. 2009. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science*. 326:137–140.
- Briggs A.W., Stenzel U., Johnson P.L.F., Green R.E., Kelso J., Prüfer K., Meyer M., Krause J., Ronan M.T., Lachmann M., Pääbo S., Prüfer K., Meyer M., Krause J., Ronan M.T., Lachmann M., Paabo S. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA* 104:14616–14621.
- Briggs A.W., Stenzel U., Meyer M., Krause J., Kircher M., Pääbo S. 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* 38:e87.
- Burger J., Kirchner M., Bramanti B., Haak W., Thomas M.G. 2007. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc. Natl. Acad. Sci. USA* 104:3736–3741.
- Burokas A., Moloney R.D., Dinan T.G., Cryan J.F. 2015. Microbiota regulation of the mammalian gut-brain axis. *Adv. Appl. Microbiol.* 91:1–62.
- Burri R., Nater A., Kawakami T., Mugal C.F., Olason P.I., Smeds L., Suh A., Dutoit L., Bureš S., Garamszegi L.Z., Hogner S., Moreno J., Qvarnström A., Ruzić M., Sæther S.-A., Sætre G.-P., Török J., Ellegren H. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula flycatchers*. *Genome Res.*:gr.196485.115.
- Campos P.F., Willerslev E., Sher A., Orlando L., Axelsson E., Tikhonov A., Aaris-Sørensen K., Greenwood A.D., Kahlke R.-D., Kosintsev P.A., Krakhmalnaya T., Kuznetsova T., Lemey P., MacPhee R., Norris C.A., Shepherd K., Suchard M.A., Zazula G.D., Shapiro B., Gilbert M.T.P. 2010. Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proc. Natl. Acad. Sci. USA* 107:5675–80.
- Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Peña A.G., Goodrich J.K., Gordon J.L., Huttley G.A., Kelley S.T., Knights D., Koenig J.E., Ley R.E., Lozupone C.A., McDonald D., Muegge B.D., Pirrung M., Reeder J., Sevinsky J.R., Turnbaugh P.J., Walters W.A., Widmann J., Yatsunenko T., Zaneveld J., Knight R., et al. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*. 7:335–336.
- Carpenter M.L., Buenrostro J.D., Valdiosera C., Schroeder H., Allentoft M.E.E., Sikora M., Rasmussen M., Gravel S., Guilleán S., Nekhrizov G., Leshtakov K., Dimitrova D., Theodossiev N., Pettener D., Luiselli D., Sandoval K., Moreno-Estrada A., Li Y., Wang J., Gilbert M.T.P., Willerslev E., Greenleaf W.J., Bustamante C.D. 2013. Pulling out the 1%: Whole-Genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am. J. Hum. Genet.* 93: 852–864.
- Cassidy L.M., Martiniano R., Murphy E.M., Teasdale M.D., Mallory J., Hartwell B., Bradley D.G. 2016. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proc. Natl. Acad. Sci. USA* 113:368–73.
- Chan J.Z.-M., Sergeant M.J., Lee O.Y.-C., Minnikin D.E., Besra G.S., Pap I., Spigelman M., Donoghue H.D., Pallen M.J. 2013. Metagenomic Analysis of Tuberculosis in a Mummy. *N. Engl. J. Med.* 369:289–290.
- Chan Y.L., Anderson C.N.K., Hadly E. a. 2006. Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. *PLoS Genet.* 2:e59.
- Cui Y., Yu C., Yan Y., Li D., Li Y., Jombart T., Weinert L.A., Wang Z., Guo Z., Xu L., Zhang Y., Zheng H., Qin N., Xiao X., Wu M., Wang X., Zhou D., Qi Z., Du Z., Wu H., Yang X., Cao H., Wang H., Wang J., Yao S., Rakin A., Li Y., Falush D., Balloux F., Achtman M., Song Y., Wang J., Yang R. 2013. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc. Natl. Acad. Sci. USA* 110:577–582.
- Curat M., Excoffier L., Maddison W., Otto S.P., Ray N., Whitlock M.C., Yeaman S. 2006. Comment on “Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*” and “Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans”. *Science*. 313:172; author reply 172.
- D’Costa V.M., King C.E., Kalan L., Morar M., Sung W.W.L., Schwarz C., Froese D., Zazula G., Calmels F., Debruyne R., Golding G.B., Poinar H.N., Wright G.D. 2011. Antibiotic resistance is ancient. *Nature*. 477:457–461.
- Dabney J., Knapp M., Glocke I., Gansauge M.-T.M.-T., Weihmann A., Nickel B., Valdiosera C., García N., Pääbo S., Arsuaga J.-L., Meyer M., García N., Pääbo S., Arsuaga J.-L., Meyer M. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* 110:15758–15763.
- Damgaard P.B., Margaryan A., Schroeder H., Orlando L., Willerslev E., Allentoft M.E.E. 2015. Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* 5:11184.
- Depaulis F., Orlando L., Hänni C. 2009. Using classical population genetics tools with heterochronous data: time matters! *PLoS One*. 4:e5541.
- Drummond A.J., Nicholls G.K., Rodrigo A.G., Solomon W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*. 161:1307–1320.
- Drummond A.J., Rambaut A., Shapiro B., Pybus O.G. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22:1185–1192.
- Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29:mss075.
- Durand E.Y., Patterson N., Reich D., Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28:2239–2252.
- Epp L.S., Boessenkool S., Bellemain E.P., Haile J., Esposito A., Riaz T., Erséus C., Gusarov V.I., Edwards M.E., Johnsen A., Stenøien H.K., Hassel K., Kauseud H., Yoccoz N.G., Bråthen K.A., Willerslev E., Taberlet P., Coissac E., Brochmann C., et al. 2012. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol. Ecol.* 21:1821–1833.
- Eriksson A., Manica A. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc. Natl. Acad. Sci. USA* 109:13956–13960.
- Excoffier L., Foll M. 2011. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*. 27:1332–1334.
- Fabre P.-H., Vilstrup J.T., Raghavan M., Der Sarkissian C., Willerslev E., Douzery E.J.P., Orlando L. 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biol. Lett.* 10:20140266–20140266.
- Ficetola G.F., Pansu J., Bonin A., Coissac E., Giguet-Covex C., De Barba M., Gielly L., Lopes C.M., Boyer F., Pompanon F., Rayé G., Taberlet P., et al. 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Mol. Ecol. Resour.* 15:543–556.
- Fletcher H.A., Donoghue H.D., Holton J., Pap I., Spigelman M. 2003. Widespread occurrence of *Mycobacterium tuberculosis* DNA from 18th–19th century Hungarians. *Am. J. Phys. Anthropol.* 120: 144–152.
- da Fonseca R.R., Smith B.D., Wales N., Cappellini E., Skoglund P., Fumagalli M., Samaniego J.A., Caroe C., Ávila-Arcos M.C., Hufnagel D.E., Korneliusen T.S., Vieira F.G., Jakobsson M., Arriaza B., Willerslev E., Nielsen R., Hufford M.B., Albrechtsen A., Ross-Ibarra J., Gilbert M.T.P. 2015. The origin and evolution of maize in the Southwestern United States. *Nat. Plants*. 1:14003.
- Fu Q., Hajdinjak M., Moldovan O.T., Constantin S., Mallick S., Skoglund P., Patterson N., Rohland N., Lazaridis I., Nickel B., Viola B., Prüfer K., Meyer M., Kelso J., Reich D., Pääbo S. 2015. An early modern human from Romania with a recent Neanderthal ancestor. *Nature*. 524:216–219.

- Fu Q., Li H., Moorjani P., Jay F., Slepchenko S.M., Bondarev A.A., Johnson P.L.F., Aximu-Petri A., Prüfer K., de Filippo C., Meyer M., Zwyns N., Salazar-García D.C., Kuzmin Y.V., Keates S.G., Kosintsev P.A., Razhev D.I., Richards M.P., Peristov N.V., Lachmann M., Douka K., Higham T.F.G., Slatkin M., Hublin J.-J., Reich D., Kelso J., Viola T.B., Pääbo S. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*. 514:445–449.
- Fu Q., Posth C., Hajdinjak M., Petr M., Mallick S., Fernandes D., Furtwängler A., Haak W., Meyer M., Mittnik A., Nickel B., Peltzer A., Rohland N., Slon V., Talamo S., Lazaridis I., Lipson M., Mathieson I., Schiffels S., Skoglund P., Derevianko A.P., Drozdov N., Slavinsky V., Tsybankov A., Cremonesi R.G., Mallegni F., Gély B., Vacca E., Morales M.R.G., Straus L.G., Neugebauer-Maresch C., Teschler-Nicola M., Constantin S., Moldovan O.T., Benazzi S., Peresani M., Coppola D., Lari M., Ricci S., Ronchitelli A., Valentin F., Thevenet C., Wehrberger K., Grigorescu D., Rougier H., Crevecoeur I., Flas D., Semal P., Mannino M.A., Cupillard C., Bocherens H., Conard N.J., Harvati K., Moiseyev V., Drucker D.G., Svoboda J., Richards M.P., Caramelli D., Pinhasi R., Kelso J., Patterson N., Krause J., Pääbo S., Reich D. 2016. The genetic history of Ice Age Europe. *Nature*.
- Fumagalli M., Vieira F.G., Korneliussen T.S., Linderoth T., Huerta-Sánchez E., Albrechtsen A., Nielsen R. 2013. Quantifying population genetic differentiation from next-generation sequencing data. *Genetics*. 195:979–992.
- Gamba C., Fernández E., Tirado M., Deguilloux M.-F., Pemonge M.H., Utrilla P., Edo M., Molist M., Rasteiro R., Chikhi L., Arroyo-Pardo E. 2012. Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. *Mol. Ecol.* 21:45–56.
- Gamba C., Hanghøj K., Gauntz C., Alfarhan A.H., Alquraishi S.A., Al-Rasheid K.A.S., Bradley D.G., Orlando L. 2016. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol. Ecol. Resour.* 16:459–469.
- Gamba C., Jones E.R., Teasdale M.D., McLaughlin R.L., Gonzalez-Fortes G., Mattiangeli V., Domboróczki L., Kóvári I., Pap I., Anders A., Whittle A., Dani J., Raczky P., Higham T.F.G., Hofreiter M., Bradley D.G., Pinhasi R. 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* 5:5257.
- Gansauge M.-T., Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* 8:737–748.
- Gansauge M.-T., Meyer M. 2014. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.* 24:1543–1549.
- Gavryushkina A., Welch D., Stadler T., Drummond A.J. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput. Biol.* 10:e1003919.
- Giguet-Covex C., Pansu J., Arnaud F., Rey P.-J., Griggo C., Gjelly L., Domaizon I., Coissac E., David F., Choler P., Poulenard J., Taberlet P. 2014. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat. Commun.* 5:3211.
- Gilbert M.T.P., Wilson A.S., Bunce M., Hansen A.J., Willerslev E., Shapiro B., Higham T.F.G., Richards M.P., O'Connell T.C., Tobin D.J., Janaway R.C., Cooper A. 2004. Ancient mitochondrial DNA from hair. *Curr. Biol.* 14:R463–R464.
- Ginolhac A., Vilstrup J., Stenderup J., Rasmussen M., Stiller M., Shapiro B., Zazula G., Froese D., Steinmann K.E., Thompson J.F., Al-Rasheid K.A.S., Gilbert M.T.P., Willerslev E., Orlando L. 2012. Improving the performance of true single molecule sequencing for ancient DNA. *BMC Genomics*. 13:177.
- Gokhman D., Lavi E., Prüfer K., Fraga M.F., Riancho J.A., Kelso J., Pääbo S., Meshorer E., Carmel L. 2014. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science*. 344:523–527.
- Goldstein D.B., Chikhi L. 2002. Human migrations and population structure: what we know and why it matters. *Annu. Rev. Genomics Hum. Genet.* 3:129–152.
- Gould B.A., Leon B., Buffen A.M., Thompson L.G. 2010. Evidence of a high-Andean, mid-Holocene plant community: An ancient DNA analysis of glacially preserved remains. *Am. J. Bot.* 97:1579–1584.
- Grealy A.C., McDowell M.C., Scofield P., Murray D.C., Fusco D.A., Haile J., Prideaux G.J., Bunce M. 2015. A critical evaluation of how ancient DNA bulk bone metabarcoding complements traditional morphological analysis of fossil assemblages. *Quat. Sci. Rev.* 128:37–47.
- Green R.E., Briggs A.W., Krause J., Prüfer K., Burbano H.A., Siebauer M., Lachmann M., Pääbo S. 2009. The Neandertal genome and ancient DNA authenticity. *EMBO J.* 28:2494–502.
- Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H.-Y.Y., Hansen N.F., Durand E.Y., Malaspinas A.-S.S., Jensen J.D., Marques-Bonet T., Alkan C., Prüfer K., Meyer M., Burbano H.A., Good J.M., Schultz R., Aximu-Petri A., Butthof A., Höber B., Höffner B., Siegemund M., Weihmann A., Nusbaum C., Lander E.S., Russ C., Novod N., Affourtit J.P., Egholm M., Verna C., Rudan P., Brajkovic D., Kucan Ž., Gušić I., Doronichev V.B., Golovanova L. V., Lalueza-Fox C., de la Rasilla M., Fortea J., Rosas A., Schmitz R.W., Johnson P.L.F., Eichler E.E., Falush D., Birney E., Mullikin J.C., Slatkin M., Nielsen R., Kelso J., Lachmann M., Reich D., Pääbo S., Kucan Z., Gusic I., Rasilla M. de la, Prüfer K., Hober B., Hoffner B., Paabo S. 2010. A draft sequence of the Neandertal genome. *Science*. 328:710–722.
- Günther T., Valdiosera C., Malmström H., Ureña I., Rodriguez-Varela R., Sverrisdóttir Ó.O., Daskalaki E.A., Skoglund P., Naidoo T., Svensson E.M., Bermúdez de Castro J.M., Carbonell E., Dunn M., Stora J., Iriarte E., Arsuaga J.-L., Carretero J.-M., Götherström A., Jakobsson M. 2015. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc. Natl. Acad. Sci. USA* 112:11917–11922.
- Haak W., Balanovsky O., Sanchez J.J., Koshel S., Zaporozhchenko V., Adler C.J., Der Sarkissian C., Brandt G., Schwarz C., Nicklisch N., Dresely V., Fritsch B., Balanovska E., Villemers R., Meller H., Alt K.W., Cooper A., Consortium M. of the G. 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biol.* 8:e1000536.
- Haak W., Lazaridis I., Patterson N., Rohland N., Mallick S., Llamas B., Brandt G., Nordenfelt S., Harney E., Stewardson K., Fu Q., Mittnik A., Bánffy E., Economou C., Francken M., Friederich S., Pena R.G., Hallgren F., Khartanovich V., Khokhlov A., Kunst M., Kuznetsov P., Meller H., Mochalov O., Moiseyev V., Nicklisch N., Pichler S.L., Risch R., Rojo Guerra M.A., Roth C., Szécsényi-Nagy A., Wahl J., Meyer M., Krause J., Brown D., Anthony D., Cooper A., Alt K.W., Reich D. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 522:207–211.
- Hagen O., Hartmann K., Steel M., Stadler T. 2015. Age-dependent speciation can explain the shape of empirical phylogenies. *Syst. Biol.* 64:432–440.
- Haile J., Froese D.G., MacPhee R.D.E., Roberts R.G., Arnold L.J., Reyes A. V., Rasmussen M., Nielsen R., Brook B.W., Robinson S., Demuro M., Gilbert M.T.P., Munch K., Austin J.J., Cooper A., Barnes I., Möller P., Willerslev E. 2009. Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proc. Natl. Acad. Sci. USA* 106:22352–22357.
- Haile J., Holdaway R., Oliver K., Bunce M., Gilbert M.T.P., Nielsen R., Munch K., Ho S.Y.W., Shapiro B., Willerslev E. 2007. Ancient DNA Chronology within Sediment Deposits: Are Paleobiological Reconstructions Possible and Is DNA Leaching a Factor? *Mol. Biol. Evol.* 24:982–989.
- Haouchar D., Haile J., McDowell M.C., Murray D.C., White N.E., Allcock R.J.N., Phillips M.J., Prideaux G.J., Bunce M. 2014. Thorough assessment of DNA preservation from fossil bone and sediments excavated from a late Pleistocene–Holocene cave deposit on Kangaroo Island, South Australia. *Quat. Sci. Rev.* 84:56–64.
- Harris K., Nielsen R. 2013. Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genet.* 9:e1003521.
- Heath T.A., Huelsenbeck J.P., Stadler T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proc. Natl. Acad. Sci. USA* 111:E2957–E2966.
- Heintzman P.D., Soares A.E.R., Chang D., Shapiro B. 2015. Paleogenom. *Rev. Cell Biol. Mol. Med.* 1:243–267.
- Heled J., Drummond A.J. 2008. Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* 8:289.
- Heller R., Chikhi L.L., Siegismund H.R. 2013. The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS One.* 8:e62992.

- Higuchi R., Bowman B., Freiburger M., Ryder O.A., Wilson A.C. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature*. 312:282–284.
- Hinch A.G., Tandon A., Patterson N., Song Y., Rohland N., Palmer C.D., Chen G.K., Wang K., Buxbaum S.G., Akyzbekova E.L., Aldrich M.C., Ambrosone C.B., Amos C., Bandera E. V, Berndt S.I., Bernstein L., Blot W.J., Bock C.H., Boerwinkle E., Cai Q., Caporaso N., Casey G., Cupples L.A., Deming S.L., Diver W.R., Divers J., Fornage M., Gillanders E.M., Glessner J., Harris C.C., Hu J.J., Ingles S.A., Isaacs W., John E.M., Kao W.H.L., Keating B., Kittles R.A., Kolonel L.N., Larkin E., Le Marchand L., McNeill L.H., Millikan R.C., Murphy A., Musani S., Neslund-Dudas C., Nyante S., Papanicolaou G.J., Press M.F., Psaty B.M., Reiner A.P., Rich S.S., Rodriguez-Gil J.L., Rotter J.L., Rybicki B.A., Schwartz A.G., Signorello L.B., Spitz M., Strom S.S., Thun M.J., Tucker M.A., Wang Z., Wiencke J.K., Witte J.S., Wrensch M., Wu X., Yamamura Y., Zanetti K.A., Zheng W., Ziegler R.G., Zhu X., Redline S., Hirschhorn J.N., Henderson B.E., Taylor H.A., Price A.L., Hakonarson H., Chanock S.J., Haiman C.A., Wilson J.G., Reich D., Myers S.R. 2011. The landscape of recombination in African Americans. *Nature*. 476:170–175.
- Ho S.Y.W., Duchêne S. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. *Mol. Ecol.* 23:5947–65.
- Ho S.Y.W., Heupink T.H., Rambaut A., Shapiro B. 2007. Bayesian estimation of sequence damage in ancient DNA. *Mol. Biol. Evol.* 24:1416–22.
- Ho S.Y.W., Lanfear R., Bromham L., Phillips M.J., Soubrier J., Rodrigo A.G., Cooper A. 2011. Time-dependent rates of molecular evolution. *Mol. Ecol.* 20:3087–3101.
- Ho S.Y.W., Shapiro B. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol. Ecol. Resour.* 11:423–434.
- Ho S.Y.W., Tong K.J., Foster C.S.P., Ritchie A.M., Lo N., Crisp M.D. 2015. Biogeographic calibrations for the molecular clock. *Biol. Lett.* 11:20150194.
- Hoffman N.G., Matsen F.A. 2015. Computational Tools for Taxonomic Assignment. *Encycl. Metagenomics Environ. Metagenomics*:108–112.
- Hofreiter M., Jaenicke V., Serre D., von Haeseler A., Pääbo S. 2001a. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* 29:4793–4799.
- Hofreiter M., Mead J.I., Martin P., Poinar H.N. 2003. Molecular caving. *Curr. Biol.* 13:R693–R695.
- Hofreiter M., Serre D., Poinar H.N., Kuch M., Pääbo S. 2001b. Ancient DNA. *Nat. Rev. Genet.* 2:353–359.
- Hou W., Dong H., Li G., Yang J., Coolen M.J.L., Liu X., Wang S., Jiang H., Wu X., Xiao H., Lian B., Wan Y., et al. 2014. Identification of photosynthetic plankton communities using sedimentary ancient DNA and their response to late-Holocene climate change on the Tibetan Plateau. *Sci. Rep.* 4:6648.
- Hudson R.R. 1983. Properties of a neutral allele model with intragenic recombination. *Theor. Popul. Biol.* 23:183–201.
- Hudson R.R. 1990. Gene genealogies and the coalescent process. *Oxford Surv. Evol. Biol.* 7:44.
- Huson D.H., Auch A.F., Qi J., Schuster S.C. 2007. MEGAN analysis of metagenomic data. *Genome Res.* 17:377–386.
- Itan Y., Powell A., Beaumont M.A., Burger J., Thomas M.G. 2009. The origins of lactase persistence in Europe. *PLoS Comput. Biol.* 5:e1000491.
- Jones E.R., Gonzalez-Fortes G., Connell S., Siska V., Eriksson A., Martiniano R., McLaughlin R.L., Gallego Llorente M., Cassidy L.M., Gamba C., Meshveliani T., Bar-Yosef O., Müller W., Belfer-Cohen A., Matskevich Z., Jakeli N., Higham T.F.G., Currat M., Lordkipanidze D., Hofreiter M., Manica A., Pinhasi R., Bradley D.G., McLaughlin R.L., Llorente M.G., Cassidy L.M., Gamba C., Meshveliani T., Bar-Yosef O., Müller W., Belfer-Cohen A., Matskevich Z., Jakeli N., Higham T.F.G., Currat M., Lordkipanidze D., Hofreiter M., Manica A., Pinhasi R., Bradley D.G. 2015. Upper palaeolithic genomes reveal deep roots of modern eurasians. *Nat. Comm.* 6:8912.
- Jónsson H., Schubert M., Seguin-Orlando A., Ginolhac A., Petersen L., Fumagalli M., Albrechtsen A., Petersen B., Korneliusen T.S., Vilstrup J.T., Lear T., Myka J.L., Lundquist J., Miller D.C., Alfarhan A.H., Alquraishi S. a., Al-Rasheid K. a. S., Stagegaard J., Strauss G., Bertelsen M.F., Sicheritz-Ponten T., Antczak D.F., Bailey E., Nielsen R., Willerslev E., Orlando L. 2014. Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc. Natl. Acad. Sci. USA* 111(52):18655–18660.
- Jørgensen T., Haile J., Möller P., Andreev A., Boessenkool S., Rasmussen M., Kienast F., Coissac E., Taberlet P., Brochmann C., Bigelow N.H., Andersen K., Orlando L., Gilbert M.T.P., Willerslev E., et al. 2012. A comparative study of ancient sedimentary DNA, pollen and microfossils from permafrost sediments of northern Siberia reveals long-term vegetational stability: comparative study of ancient sedimentary DNA, pollen and microfossils. *Mol. Ecol.* 21:1989–2003.
- Kay G.L., Sergeant M.J., Zhou Z., Chan J.Z.-M., Millard A., Quick J., Szikossy I., Pap I., Spigelman M., Loman N.J., al et, Achtman M., Donoghue H.D., Pallen M.J. 2015. Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nat. Commun.* 6:6717.
- Keller A., Graefen A., Ball M., Matzas M., Boisguerin V., Maixner F., Leidinger P., Backes C., Khairat R., Forster M., Stade B., Franke A., Mayer J., Spangler J., McLaughlin S., Shah M., Lee C., Harkins T.T., Sartori A., Moreno-Estrada A., Henn B., Sikora M., Semino O., Chiaroni J., Rootsi S., Myres N.M., Cabrera V.M., Underhill P.A., Bustamante C.D., Vigl E.E., Samadelli M., Cipollini G., Haas J., Katus H., O'Connor B.D., Carlson M.R.J., Meder B., Blin N., Meese E., Pusch C.M., Zink A. 2012. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* 3:698.
- Kerpedjiev P., Frellsen J., Lindgreen S., Krogh A. 2014. Adaptable probabilistic mapping of short reads using position specific scoring matrices. *BMC Bioinformatics.* 15:100.
- Kim E.-S., Rothschild M.F. 2014. Genomic adaptation of admixed dairy cattle in East Africa. *Front. Genet.* 5:443.
- Kingman J.F.C. 1982. The coalescent. *Stoch. Process Appl.* 13:235–248.
- Knights D., Kuczynski J., Charlson E.S., Zaneveld J., Mozer M.C., Collman R.G., Bushman F.D., Knight R., Kelley S.T. 2011. Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods.* 8:761–763.
- Korlević P., Gerber T., Gansauge M.-T., Hajdinjak M., Nagel S., Aximu-Petri A., Meyer M. 2015. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques.* 59:87–93.
- Korneliusen T.S., Albrechtsen A., Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics.* 15:356.
- Krause-Kyora B., Makarewicz C., Evin A., Flink L.G., Dobney K., Larson G., Hartz S., Schreiber S., von Carnap-Bornheim C., von Wurmb-Schwark N., Nebel A. 2013. Use of domesticated pigs by Mesolithic hunter-gatherers in northwestern Europe. *Nat. Commun.* 4:2348.
- Kronforst M.R., Hansen M.E.B., Crawford N.G., Gallant J.R., Zhang W., Kulathinal R.J., Kapan D.D., Mullen S.P. 2013. Hybridization reveals the evolving genomic architecture of speciation. *Cell Rep.* 5:666–77.
- Kuhlwilm M., Gronau I., Hubisz M.J., de Filippo C., Prado-Martinez J., Kircher M., Fu Q., Burbano H.A., Lalueza-Fox C., de la Rasilla M., Rosas A., Rudan P., Brajkovic D., Kucan Ž., Gušić I., Marques-Bonet T., Andrés A.M., Viola B., Pääbo S., Meyer M., Siepel A., Castellano S. 2016. Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature.* 530:429–433.
- Lacan M., Keyser C., Ricaut F.-X., Brucato N., Tarrús J., Bosch A., Guilaine J., Crubézy E., Ludes B., Duranthon F., Guilaine J., Crubézy E., Ludes B. 2011a. Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc. Natl. Acad. Sci. USA* 108:9788–9791.
- Lacan M., Keyser C., Ricaut F.-X., Brucato N., Tarrús J., Bosch A., Guilaine J., Crubézy E., Ludes B. 2011b. Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc. Natl. Acad. Sci. USA* 108:18255–18259.
- Larson G., Albarella U., Dobney K., Rowley-Conwy P., Schibler J., Tresselt A., Vigne J.-D., Edwards C.J., Schlumbaum A., Dinu A., Balaşescu A., Dolman G., Tagliacozzo A., Manaseryan N., Miracle P., Van Wijngaarden-Bakker L., Masseti M., Bradley D.G., Cooper A. 2007. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc. Natl. Acad. Sci. USA* 104:15276–15281.

- Larson G., Dobney K., Albarella U., Fang M., Matisoo-Smith E., Robins J., Lowden S., Finlayson H., Brand T., Willerslev E., Rowley-Conwy P., Andersson L., Cooper A. 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*. 307:1618–1621.
- Lazaridis I., Patterson N., Mittnik A., Renaud G., Mallick S., Kirsanov K., Sudmant P.H., Schraiber J.G., Castellano S., Lipson M., Berger B., Economou C., Bollongino R., Fu Q., Bos K.I., Nordenfelt S., Li H., de Filippo C., Prüfer K., Sawyer S., Posth C., Haak W., Hallgren F., Fornander E., Rohland N., Delsate D., Francken M., Guinet J.-M., Wahl J., Ayodo G., Babiker H.A., Bailliet G., Balanovska E., Balanovsky O., Barrantes R., Bedoya G., Ben-Ami H., Bene J., Berrada F., Bravi C.M., Brisighelli F., Busby G.B.J., Cali F., Churnosov M., Cole D.E.C., Corach D., Damba L., van Driem G., Dryomov S., Dugoujon J.-M., Fedorova S.A., Gallego Romero I., Gubina M., Hammer M., Henn B.M., Hervig T., Hodoglugil U., Jha A.R., Karachanak-Yankova S., Khusainova R., Khusnutdinova E., Kittles R., Kivisild T., Klitz W., Kučinskas V., Kushniarevich A., Laredj L., Litvinov S., Loukidis T., Mahley R.W., Melegh B., Metspalu E., Molina J., Mountain J., Näkkäläjärvi K., Nesheva D., Nyambo T., Osipova L., Parik J., Platonov F., Posukh O., Romano V., Rothhammer F., Rudan I., Ruizbakiev R., Sahakyan H., Sajantila A., Salas A., Starikovskaya E.B., Tarekuev T., Toncheva D., Turdikulova S., Uktveryte I., Utevska O., Vasquez R., Villena M., Voevoda M., Winkler C.A., Yepiskoposyan L., Zalloua P., Zemunik T., Cooper A., Capelli C., Thomas M.G., Ruiz-Linares A., Tishkoff S.A., Singh L., Thangaraj K., Villemers R., Comas D., Sukernik R., Metspalu M., Meyer M., Eichler E.E., Burger J., Slatkin M., Pääbo S., Kelso J., Reich D., Krause J. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. 513: 409–413.
- Leonardi M., Gerbault P., Thomas M.G., Burger J. 2012. The evolution of lactase persistence in Europe. A synthesis of archaeological and genetic evidence. *Int. Dairy J.* 22:88–97.
- Li H., Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*. 475:493–496.
- Li J.Z., Absher D.M., Tang H., Southwick A.M., Casto A.M., Ramachandran S., Cann H.M., Barsh G.S., Feldman M.W., Cavalli-Sforza L.L., Myers R.M. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*. 319:1100–1104.
- Librado P., Der Sarkissian C., Ermini L., Schubert M., Jónsson H., Albrechtsen A., Fumagalli M., Yang M.M.A., Gamba C., Seguin-Orlando A., Mortensen C.C.D., Petersen B., Hoover C.A.C., Lorente-Galdos B., Nedoluzhko A., Boulygina E., Tsygankova S., Neuditschko M., Jagannathan V., Thèves C., Alfarhan A.H., Alquraishi S.A., Al-Rasheed K.A.S., Sicheritz-Ponten T., Popov R., Grigoriev S., Alekseev A.N., Rubin E.M.E., McCue M., Rieder S., Leeb T., Tikhonov A., Crubézy E., Slatkin M., Marques-Bonet T., Nielsen R., Willerslev E., Kantanen J., Prokhorchouk E., Orlando L., Albrechtsen A., Neuditschko M., Jagannathan V. 2015. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to arctic environments. *Proc. Natl. Acad. Sci. USA* 112:E6889–E6897.
- Lippold S., Knapp M., Kuznetsova T., Leonard J.A., Benecke N., Ludwig A., Rasmussen M., Cooper A., Weinstock J., Willerslev E., Shapiro B., Hofreiter M. 2011. Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nat. Commun.* 2:450.
- Lipson M., Loh P.-R., Levin A., Reich D., Patterson N., Berger B. 2013. Efficient moment-based inference of admixture parameters and sources of gene flow. *Mol. Biol. Evol.* 30:1788–1802.
- Lipson M., Loh P.-R., Patterson N., Moorjani P., Ko Y.-C., Stoneking M., Berger B., Reich D. 2014. Reconstructing Austronesian population history in Island Southeast Asia. *Nat. Commun.* 5: 4689.
- Llorente M.G., Jones E.R., Eriksson A., Siska V., Arthur K.W., Arthur J.W., Curtis M.C., Stock J.T., Coltorti M., Pieruccini P., Stretton S., Brock F., Higham T.F.G., Park Y., Hofreiter M., Bradley D.G., Bhak J., Pinhasi R., Manica A. 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science*. 350:820–822.
- Loh P.-R., Lipson M., Patterson N., Moorjani P., Pickrell J.K., Reich D., Berger B. 2013. Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. *Genetics*. 193:1233–1254.
- Lorenzen E.D., Nogués-Bravo D., Orlando L., Weinstock J., Binladen J., Marske K. a., Ugan A., Borregaard M.K., Gilbert M.T.P., Nielsen R., Ho S.Y.W., Goebel T., Graf K.E., Byers D., Stenderup J.T., Rasmussen M., Campos P.F., Leonard J. a., Koepfli K.-P., Froese D., Zazula G., Stafford T.W., Aaris-Sørensen K., Batra P., Haywood A.M., Singarayer J.S., Valdes P.J., Boeskorov G., Burns J.A., Davydov S.P., Haile J., Jenkins D.L., Kosintsev P.A., Kuznetsova T., Lai X., Martin L.D., McDonald H.G., Mol D., Meldgaard M., Munch K., Stephan E., Sablin M., Sommer R.S., Sipko T., Scott E., Suchard M. a., Tikhonov A., Willerslev R., Wayne R.K., Cooper A., Hofreiter M., Sher A., Shapiro B., Rahbek C., Willerslev E., Stafford Jr T.W. 2011. Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*. 479:359–364.
- Louvel G., Der Sarkissian C., Hanghøj K., Orlando L.A.A. metaBIT, an integrative and automated metagenomic pipeline for analysing microbial profiles from high-throughput sequencing shotgun data. *Mol. Ecol. Resour.* doi:10.1111/1755-0998.12546.
- Ludwig A., Pruvost M., Reissmann M., Benecke N., Brockmann G.A., Castanos P., Cieslak M., Lippold S., Llorente L., Malaspinas A.S., Slatkin M., Hofreiter M. 2009. Coat color variation at the beginning of horse domestication. *Science*. 324:485.
- Ludwig A., Reissmann M., Benecke N., Bellone R.R., Sandoval-Castellanos E., Cieslak M., Fortes G.G., Morales-Muñiz A., Hofreiter M., Pruvost M., Morales-Muniz A., Hofreiter M., Pruvost M. 2014. Twenty-five thousand years of fluctuating selection on leopard complex spotting and congenital night blindness in horses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370:20130386.
- Lynch V.J., Bedoya-Reina O.C., Ratan A., Sulak M., Drautz-Moses D.I., Perry G.H., Miller W., Schuster S.C. 2015. Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. *Cell Rep.* 12:217–228.
- Mailund T., Dutheil J.Y., Hobolth A., Lunter G., Schierup M.H. 2011. Estimating divergence time and ancestral effective population size of Bornean and Sumatran orangutan subspecies using a coalescent hidden Markov model. *PLoS Genet.* 7:e1001319.
- Maixner F., Krause-Kyora B., Turaev D., Herbig A., Hoopmann M.R., Hallows J.L., Kusebauch U., Vigil E.E., Malfertheiner P., Megraud F., OSullivan N., Cipollini G., Coia V., Samadelli M., Engstrand L., Linz B., Moritz R.L., Grimm R., Krause J., Nebel A., Moodley Y., Rattei T., Zink A. 2016. The 5300-year-old *Helicobacter pylori* genome of the Iceman. *Science*. 351:162–165.
- Malaspinas A.-S., Lao O., Schroeder H., Rasmussen M., Raghavan M., Moltke I., Campos P.F., Sagredo F.S., Rasmussen S., Gonçalves V.F., Albrechtsen A., Allentoft M.E.E., Johnson P.L.F., Li M., Reis S., Bernardo D. V., DeGiorgio M., Duggan A.T., Bastos M., Wang Y., Stenderup J., Moreno-Mayar J.V., Brunak S., Sicheritz-Ponten T., Hodges E., Hannon G.J., Orlando L., Price T.D., Jensen J.D., Nielsen R., Heinemeier J., Olsen J. V., Rodrigues-Carvalho C., Lahr M.M., Neves W.A., Kayser M., Higham T.F.G., Stoneking M., Pena S.D.J., Willerslev E. 2014. Two ancient human genomes reveal Polynesian ancestry among the indigenous Botocudos of Brazil. *Curr. Biol.* 24:R1035–1037.
- Malaspinas A.-S., Malaspinas O., Evans S.N., Slatkin M. 2012. Estimating allele age and selection coefficient from time-serial data. *Genetics*. 192:599–607.
- Malaspinas A.-S. 2016. Methods to characterize selective sweeps using time serial samples: an ancient DNA perspective. *Mol. Ecol.* 25:24–41.
- Malmström H., Gilbert M.T.P., Thomas M.G., Brandström M., Storå J., Molnar P., Andersen P.K., Bendixen C., Holmlund G., Götherström A., Willerslev E. 2009. Ancient DNA reveals lack of continuity between Neolithic hunter-gatherers and contemporary Scandinavians. *Curr. Biol.* 19:1758–1762.
- Malmström H., Linderholm A., Liden K., Stora J., Molnar P., Holmlund G., Jakobsson M., Götherström A., Malmström H., Linderholm A., Liden K., Storå J., Molnar P., Holmlund G., Jakobsson M., Götherström A. 2010. High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. *BMC Evol. Biol.* 10:89.
- Marjoram P., Wall J.D. 2006. Fast “coalescent” simulation. *BMC Genet.* 7:16.



- Martin M.D., Cappellini E., Samaniego J.A., Zepeda M.L., Campos P.F., Seguin-Orlando A., Wales N., Orlando L., Ho S.Y.W., Dietrich F.S., Mieczkowski P.A., Heitman J., Willerslev E., Krogh A., Ristaino J.B., Gilbert M.T.P. 2013. Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nat. Commun.* 4:2172.
- Martin S.H., Davey J.W., Jiggins C.D. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32:244–257.
- Martiniano R., Caffell A.C., Holst M., Hunter-Mann K., Montgomery J., Müldner G., McLaughlin R.L., Teasdale M.D., van Rheeën W., Veldink J.H., van den Berg L.H., Hardiman O., Carroll M., Roskams S., Oxley J., Morgan C., Thomas M.G., Barnes I., McDonnell C., Collins M.J., Bradley D.G. 2016. Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nat. Commun.* 7:10326.
- Mathieson I., Lazaridis I., Rohland N., Mallick S., Patterson N., Roodenker S.A., Harney E., Stewardson K., Fernandes D., Novak M., Sirak K., Gamba C., Jones E.R., Llamas B., Dryomov S., Pickrell J., Arsuaga J.L., de Castro J.M.B., Carbonell E., Gerritsen F., Khokhlov A., Kuznetsov P., Lozano M., Meller H., Mochalov O., Moiseyev V., Guerra M.A.R., Roodenberg J., Vergès J.M., Krause J., Cooper A., Alt K.W., Brown D., Anthony D., Lalueza-Fox C., Haak W., Pinhasi R., Reich D. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature.* 528:499–503.
- Matisoo-Smith E., Roberts K., Welikala N., Tannock G., Chester P., Feek D., Flenley J. 2008. Recovery of DNA and pollen from New Zealand lake sediments. *Quat. Int.* 184:139–149.
- McVean G. 2009. A genealogical interpretation of principal components analysis. *PLoS Genet.* 5:e1000686.
- McVean G.A.T., Cardin N.J. 2005. Approximating the coalescent with recombination. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360:1387–1393.
- Mendum T.A., Schuenemann V.J., Roffey S., Taylor G.M., Wu H., Singh P., Tucker K., Hinds J., Cole S.T., Kierzek A.M., Niesel K., Krause J., Stewart G.R. 2014. Mycobacterium leprae genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. *BMC Genomics.* 15:270.
- Menozzi P., Piazza A., Cavalli-Sforza L.L. 1978. Synthetic maps of human gene frequencies in Europeans. *Science.* 201:786–792.
- Metcalf J.L., Prost S., Nogués-Bravo D., DeChaine E.G., Anderson C., Batra P., Araújo M.B., Cooper A., Guralnick R.P. 2014. Integrating multiple lines of evidence into historical biogeography hypothesis testing: a Bison bison case study. *Proc. Biol. Sci. USA* 281:20132782.
- Meyer M., Arsuaga J.-L., de Filippo C., Nagel S., Aximu-Petri A., Nickel B., Martínez I., Gracia A., de Castro J.M.B., Carbonell E., Viola B., Kelso J., Prüfer K., Pääbo S. 2016. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature. advance on.*
- 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., Pääbo S. 2014. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature.* 505:403–406.
- Meyer M., Kircher M., Gansauge M.-T., Li H., Racimo F., Mallick S., Schraiber J.G., Jay F., Prüfer K., de Filippo C., Sudmant P.H., Alkan C., Fu Q., Do R., Rohland N., Tandon A., Siebauer M., Green R.E., Bryc K., Briggs A.W., Stenzel U., Dabney J., Shendure J., Kitzman J., Hammer M.F., Shunkov M. V., Dereviyanko A.P., Patterson N., Andrés A.M., Eichler E.E., Slatkin M., Reich D., Kelso J., Pääbo S. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science.* 338:222–226.
- Mikheyev A.S., Tin M.M.Y., Arora J., Seeley T.D. 2015. Museum samples reveal rapid evolution by wild honey bees exposed to a novel parasite. *Nat. Commun.* 6:7991.
- Miller W., Drautz D.I., Ratan A., Pusey B., Qi J., Lesk A.M., Tomsho L.P., Packard M.D., Zhao F., Sher A., Tikhonov A., Raney B., Patterson N., Lindblad-Toh K., Lander E.S., Knight J.R., Irzyk G.P., Fredrikson K.M., Harkins T.T., Sheridan S., Pringle T., Schuster S.C. 2008. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature.* 456:387–390.
- Mitchell K.J., Llamas B., Soubrier J., Rawlence N.J., Worthy T.H., Wood J., Lee M.S.Y., Cooper A. 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science.* 344:898–900.
- Molak M., Suchard M. a., Ho S.Y.W., Beilman D.W., Shapiro B. 2015. Empirical calibrated radiocarbon sampler: a tool for incorporating radiocarbon-date and calibration error into Bayesian phylogenetic analyses of ancient DNA. *Mol. Ecol. Resour.* 15:81–86.
- Moorjani P., Patterson N., Hirschhorn J.N., Keinan A., Hao L., Atzmon G., Burns E., Ostrer H., Price A.L., Reich D. 2011. The History of African Gene Flow into Southern Europeans, Levantines, and Jews. *PLoS Genet.* 7:e1001373.
- Moorjani P., Sankararaman S., Fu Q., Przeworski M., Patterson N., Reich D. 2016. A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proc. Natl. Acad. Sci. USA.* 1514696113.
- Mourier T., Ho S.Y.W., Gilbert M.T.P., Willerslev E., Orlando L. 2012. Statistical guidelines for detecting past population shifts using ancient DNA. *Mol. Biol. Evol.* 29:2241–2251.
- Munch K., Boomsma W., Huelsenbeck J.P., Willerslev E., Nielsen R. 2008a. Statistical assignment of DNA sequences using Bayesian phylogenetics. *Syst. Biol.* 57:750–757.
- Munch K., Boomsma W., Willerslev E., Nielsen R. 2008b. Fast phylogenetic DNA barcoding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363:3997–4002.
- Murray D.C., Haile J., Dortch J., White N.E., Haouchar D., Bellgard M.I., Alcock R.J.N., Prideaux G.J., Bunce M. 2013. Scrapheap Challenge: a novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. *Sci. Rep.* 3:3371.
- Nelson M.R., Bryc K., King K.S., Indap A., Boyko A.R., Novembre J., Briley L.P., Maruyama Y., Waterworth D.M., Waeber G., Vollenweider P., Oksenberg J.R., Hauser S.L., Stirnadel H.A., Kooner J.S., Chambers J.C., Jones B., Mooser V., Bustamante C.D., Roses A.D., Burns D.K., Ehm M.G., Lai E.H. 2008. The population reference sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am. J. Hum. Genet.* 83:347–358.
- Nielsen R., Beaumont M.A. 2009. Statistical inferences in phylogeography. *Mol. Ecol.* 18:1034–1047.
- Nielsen R., Korneliussen T.S., Albrechtsen A., Li Y., Wang J. 2012. SNP calling, genotype calling, and sample allele frequency estimation from New-Generation Sequencing data. *PLoS One.* 7:e37558.
- Nordborg M. 1998. On the probability of Neanderthal ancestry. *Am. J. Hum. Genet.* 63:1237–1240.
- Nordborg M. 2008. Coalescent theory. In: Balding D.J., Bishop M., Cannings C., editors. *Handbook of statistical genetics.* West Sussex, UK: John Wiley & Sons. p. 843–872.
- Nosil P., Schluter D. 2011. The genes underlying the process of speciation. *Trends Ecol. Evol.* 26:160–167.
- Novembre J., Johnson T., Bryc K., Kutalik Z., Boyko A.R., Auton A., Indap A., King K.S., Bergmann S., Nelson M.R., Stephens M., Bustamante C.D. 2008. Genes mirror geography within Europe. *Nature.* 456:98–101.
- Olalde I., Allentoft M.E.E., Sánchez-Quinto F., Santpere G., Chiang C.W.K., DeGiorgio M., Prado-Martinez J., Rodríguez J.A., Rasmussen S., Quilez J., Ramírez O., Marigorta U.M., Fernández-Callejo M., Prada M.E., Encinas J.M.V., Nielsen R., Netea M.G., Novembre J., Sturm R.A., Sabeti P., Marquès-Bonet T., Navarro A., Willerslev E., Lalueza-Fox C. 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature.* 507:225–228.
- Olalde I., Schroeder H., Sandoval-Velasco M., Vinner L., Lobón I., Ramirez O., Civit S., García Borja P., Salazar-García D.C., Talamo S., María Fullola J., Xavier Oms F., Pedro M., Martínez P., Sanz M., Daura J., Zilhão J., Marquès-Bonet T., Gilbert M.T.P., Lalueza-Fox C. 2015. A common genetic origin for early farmers from Mediterranean Cardial and Central European LBK Cultures. *Mol. Biol. Evol.* msv181.
- Orlando L., Gilbert M.T.P., Willerslev E. 2015. Reconstructing ancient genomes and epigenomes. *Nat. Rev. Genet.* 16:395–408.
- Orlando L., Ginolhac A., Zhang G., Froese D., Albrechtsen A., Stiller M., Schubert M., Cappellini E., Petersen B., Moltke I., Johnson P.L.F., Fumagalli M., Vilstrup J.T., Raghavan M., Korneliussen T.S., Malaspina A.-S., Vogt J., Szklarczyk D., Kelstrup C.D., Vinther J., Dolocan A., Stenderup J., Velazquez A.M.V., Cahill J.A., Rasmussen M., Wang X., Min J., Zazula G.D., Seguin-Orlando A., Mortensen C., Magnusson K., Thompson J.F., Weinstock J., Gregersen K., Røed K.H., Eisenmann V., Rubin C.-J., Miller D.C., Antczak D.F., Bertelsen

- M.F., Brunak S., Al-Rasheid K. a. S., Ryder O., Andersson L., Mundy J., Krogh A., Gilbert M.T.P., Kjær K., Sicheritz-Ponten T., Jensen L.J., Olsen J. V., Hofreiter M., Nielsen R., Shapiro B., Wang J., Willerslev E. 2013. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*. 499:74–78.
- Orlando L., Willerslev E. 2014. An epigenetic window into the past? *Science*. 511:10–12.
- Pääbo S. 2014. The human condition—a molecular approach. *Cell*. 157:216–226.
- Pääbo S. 2015. The diverse origins of the human gene pool. *Nat. Rev. Genet.* 16:313–314.
- Pickrell J.K., Pritchard J.K. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8(11):e1002967.
- Padhukasahasram B. 2014. Inferring ancestry from population genomic data and its applications. *Front. Genet.* 5:204.
- Palkopoulou E., Mallick S., Skoglund P., Enk J., Rohland N., Li H., Omrak A., Vartanyan S., Poinar H., Götherström A., Reich D., Dalén L. 2015. Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth. *Curr. Biol.* 25: 1395–1400.
- Pansu J., Giguët-Covex C., Ficotola G.F., Gielly L., Boyer F., Zinger L., Arnaud F., Poulencard J., Taberlet P., Choler P. 2015. Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA. *Mol. Ecol.* 24: 1485–1498.
- Pakendorf B., Stoneking M. 2005. Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* 6:165–83.
- Parducci L., Jørgensen T., Tollefsrud M.M., Elverland E., Alm T., Fontana S.L., Bennett K.D., Haile J., Matetovici I., Suyama Y., Edwards M.E., Andersen K., Rasmussen M., Boessenkool S., Coissac E., Brochmann C., Taberlet P., Houmark-Nielsen M., Larsen N.K., Orlando L., Gilbert M.T.P., Kjær K.H., Alsos I.G., Willerslev E., et al. 2012. Glacial survival of boreal trees in Northern Scandinavia. *Science*. 335:1083–1086.
- Parducci L., Matetovici I., Fontana S.L., Bennett K.D., Suyama Y., Haile J., Kjær K.H., Larsen N.K., Drouzas A.D., Willerslev E. 2013. Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Mol. Ecol.* 22:3511–3524.
- Park S.D.E., Magee D.A., McGettigan P.A., Teasdale M.D., Edwards C.J., Lohan A.J., Murphy A., Braud M., Donoghue M.T., Liu Y., Chamberlain A.T., Rue-Albrecht K., Schroeder S., Spillane C., Tai S., Bradley D.G., Sonstegard T.S., Loftus B.J., MacHugh D.E. 2015. Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. *Genome Biol.* 16:234.
- Patterson N., Moorjani P., Luo Y., Mallick S., Rohland N., Zhan Y., Genschoreck T., Webster T., Reich D. 2012. Ancient admixture in human history. *Genetics*. 192:1065–1093.
- Pawłowska J., Lejzerowicz F., Esling P., Szczuciński W., Zaączkowski M., Pawłowski J. 2014. Ancient DNA sheds new light on the Svalbard foraminiferal fossil record of the last millennium. *Geobiology*. 12:277–288.
- Pease J.B., Hahn M.W. 2015. Detection and polarization of introgression in a five-taxon phylogeny. *Syst. Biol.* 64:651–662.
- Pedersen J.S., Valen E., Velazquez A.M.V., Parker B.J., Rasmussen M., Lindgreen S., Lilje B., Tobin D.J., Kelly T.K., Vang S., Andersson R., Jones P.A., Hoover C.A., Tikhonov A., Prokhortchouk E., Rubin E.M., Sandelin A., Gilbert M.T.P., Krogh A., Willerslev E., Orlando L. 2014. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* 24:454–466.
- Pedersen M.W., Ginolhac A., Orlando L., Olsen J., Andersen K., Holm J., Funder S., Willerslev E., Kjær K.H. 2013. A comparative study of ancient environmental DNA to pollen and microfossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quat. Sci. Rev.* 75:161–168.
- Pedersen M.W., Overballe-Petersen S., Ermini L., Der Sarkissian C., Haile J., Hellstrom M., Spens J., Thomsen P.F., Bohmann K., Cappellini E., Schnell I.B., Wales N.A., Carøe C., Campos P.F., Schmidt A.M.Z., Gilbert M.T.P., Hansen A.J., Orlando L., Willerslev E., Thomsen F., Børholm I., Astrid M., Schmidt Z., B.P.T.R.S. 2015. Ancient and modern environmental DNA. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370:20130383.
- Plantinga T.S., Alonso S., Izagirre N., Hervella M., Fregel R., van der Meer J.W., Netea M.G., de la Rúa C. 2012. Low prevalence of lactase persistence in Neolithic South-West Europe. *Eur. J. Hum. Genet.* 20:778–782.
- Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155:945–959.
- Prüfer K., Racimo F., Patterson N., Jay F., Sankararaman S., Sawyer S., Heinze A., Renaud G., Sudmant P.H., de Filippo C., Li H., Mallick S., Dannemann M., Fu Q., Kircher M., Kuhlwillm M., Lachmann M., Meyer M., Ongyerth M., Siebauer M., Theunert C., Tandon A., Moorjani P., Pickrell J., Mullikin J.C., Vohr S.H., Green R.E., Hellmann I., Johnson P.L.F., Blanche H., Cann H., Kitzman J.O., Shendure J., Eichler E.E., Lein E.S., Bakken T.E., Golovanova L. V., Doronichev V.B., Shunkov M. V., Derevianko A.P., Viola B., Slatkin M., Reich D., Kelso J., Pääbo S. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. 505:43–49.
- Pybus O.G., Rambaut A., Harvey P.H. 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*. 155:1429–1437.
- R Core Team. 2015. R: a language and environment for statistical computing.
- Raghavan M., DeGiorgio M., Albrechtsen A., Moltke I., Skoglund P., Korneliussen T.S., Grønnow B., Appelt M., Gulløv H.C., Friesen T.M., Fitzhugh W., Malmström H., Rasmussen S., Olsen J., Melchior L., Fuller B.T., Fahrni S.M., Stafford T.W., Grimes V., Renouf M.A.P., Cybulski J., Lynnerup N., Lahr M.M., Britton K., Knecht R., Arneborg J., Metspalu M., Cornejo O.E., Malaspina A.-S., Wang Y., Rasmussen M., Raghavan V., Hansen T.V.O., Khusnutdinova E., Pierre T., Dneprovsky K., Andreasen C., Lange H., Hayes M.G., Coltrain J., Spitsyn V.A., Götherström A., Orlando L., Kivisild T., Villemis R., Crawford M.H., Nielsen F.C., Dissing J., Heinemeier J., Meldgaard M., Bustamante C.D., O'Rourke D.H., Jakobsson M., Gilbert M.T.P., Nielsen R., Willerslev E. 2014a. The genetic prehistory of the New World Arctic. *Science*. 345:1255832.
- Raghavan M., Skoglund P., Graf K.E., Metspalu M., Albrechtsen A., Moltke I., Rasmussen S., Stafford T.W., Orlando L., Metspalu E., Karmin M., Tambets K., Rootsi S., Mägi R., Campos P.F., Balanovska E., Balanovsky O., Khusnutdinova E., Litvinov S., Osipova L.P., Fedorova S.A., Voevoda M.I., DeGiorgio M., Sicheritz-Ponten T., Brunak S., Demeshchenko S., Kivisild T., Villemis R., Nielsen R., Jakobsson M., Willerslev E., Stafford Jr T.W. 2014b. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*. 505:87–91.
- Raghavan M., Steinrücken M., Harris K., Schiffels S., Rasmussen S., Degiorgio M., Albrechtsen A., Valdiosera C., Avila-Arcos M.C., Malaspina A.-S.A.-S., Eriksson A., Moltke I., Metspalu M., Homburger J.R., Wall J., Cornejo O.E., Moreno-Mayar J.V., Korneliussen T.S., Pierre T.L., Rasmussen M., Campos P.F., Damgaard P. d. B., Allentoft M.E.E., Lindo J., Metspalu E., Rodriguez-Varela R., Mansilla J., Henriksen C., Seguin-Orlando A., Malmstrom H., Stafford T.W., Shringarpure S.S., Moreno-Estrada A., Karmin M., Tambets K., Bergstrom A., Xue Y., Warmuth V.M., Friend A.D., Singarayer J., Valdes P., Balloux F., LeBoreiro I., Vera J.L., Rangel-Villalobos H., Pettener D., Luiselli D., Davis L.G., Heyer E., Zolliker C.P.E., Ponce de Leon M.S., Smith C.I., Grimes V., Pike K.-A.K., Deal M., Fuller B.T., Arriaza B., Standen V., Luz M.F., Ricaut F.-X., Guidon N., Osipova L.P., Voevoda M.I., Posukh O.L., Balanovsky O., Lavryashina M., Bogunov Y., Khusnutdinova E., Gubina M., Balanovska E., Fedorova S. a, Litvinov S., Malyarchuk B. a, Derenko M. V, Mosher M.J., Archer D., Cybulski J., Petzelt B., Mitchell J.C., Worl R., Norman P.J., Parham P., Kemp B.M., Kivisild T., Tyler-Smith C., Sandhu M.S., Crawford M.H., Villemis R., Smith D.G., Waters M.R., Goebel T., Johnson J.R., Malhi R.S., Jakobsson M., Meltzer D.J., Manica A., Durbin R., Bustamante C.D., Song Y.S., Nielsen R., Willerslev E. 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*. 349:aab3884.
- Ramakrishnan U., Hadly E. a., Mountain J.L. 2005. Detecting past population bottlenecks using temporal genetic data. *Mol. Ecol.* 14:2915–2922.
- Ramakrishnan U., Hadly E. a. 2009. Using phylogenetics to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Mol. Ecol.* 18:1310–1330.

- Rambaut A., Ho S.Y.W., Drummond A.J., Shapiro B. 2009. Accommodating the effect of ancient DNA damage on inferences of demographic histories. *Mol. Biol. Evol.* 26:245–248.
- Rambaut A., Suchard M., Xie D., Drummond A. 2014. Tracer v1.6.
- Rasmussen M., Anzick S.L., Waters M.R., Skoglund P., DeGiorgio M., Stafford T.W., Rasmussen S., Moltke I., Albrechtsen A., Doyle S.M., Poznik G.D., Gudmundsdottir V., Yadav R., Malaspinas A.-S., White S.S., Allentoft M.E.E., Cornejo O.E., Tambets K., Eriksson A., Heintzman P.D., Karmin M., Korneliusson T.S., Meltzer D.J., Pierre T.L., Stenderup J., Saag L., Warmuth V.M., Lopes M.C., Malhi R.S., Brunak S., Sicheritz-Ponten T., Barnes I., Collins M.J., Orlando L., Balloux F., Manica A., Gupta R., Metspalu M., Bustamante C.D., Jakobsson M., Nielsen R., Willerslev E. 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature*. 506:225–259.
- Rasmussen M., Guo X., Wang Y., Lohmueller K.E., Rasmussen S., Albrechtsen A., Skotte L., Lindgreen S., Metspalu M., Jombart T., Kivisild T., Zhai W., Eriksson A., Manica A., Orlando L., De La Vega F.M., Tridico S., Metspalu E., Nielsen K., Ávila-Arcos M.C., Moreno-Mayar J.V., Muller C., Dortch J., Gilbert M.T.P., Lund O., Wesolowska A., Karmin M., Weinert L.A., Wang B., Li J., Tai S., Xiao F., Hanihara T., van Driem G., Jha A.R., Ricaut F.-X., de Knijff P., Migliano A.B., Gallego Romero I., Kristiansen K., Lambert D.M., Brunak S., Forster P., Brinkmann B., Nehlich O., Bunce M., Richards M.P., Gupta R., Bustamante C.D., Krogh A., Foley R.A., Lahr M.M., Balloux F., Sicheritz-Pontén T., Villemers R., Nielsen R., Wang J., Willerslev E. 2011. An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science*. 334:94–98.
- Rasmussen M., Li Y., Lindgreen S., Pedersen J.S., Albrechtsen A., Moltke I., Metspalu M., Metspalu E., Kivisild T., Gupta R., Bertalan M., Nielsen K., Gilbert M.T.P., Wang Y., Raghavan M., Campos P.F., Kamp H.M., Wilson A.S., Gledhill A., Tridico S., Bunce M., Lorenzen E.D., Binladen J., Guo X., Zhao J., Zhang X., Zhang H., Li Z., Chen M., Orlando L., Kristiansen K., Bak M., Tommerup N., Bendixen C., Pierre T.L., Grønnow B., Meldgaard M., Andreassen C., Fedorova S. a, Osipova L.P., Higham T.F.G., Ramsey C.B., Hansen T.V.O., Nielsen F.C., Crawford M.H., Brunak S., Sicheritz-Pontén T., Villemers R., Nielsen R., Krogh A., Wang J., Willerslev E. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*. 463:757–762.
- Rasmussen M., Sikora M., Albrechtsen A., Korneliusson T.S., Moreno-Mayar J.V., Poznik G.D., Zollikofer C.P.E., Ponce de León M.S., Allentoft M.E.E., Moltke I., Jónsson H., Valdiosera C., Malhi R.S., Orlando L., Bustamante C.D., Stafford T.W., Meltzer D.J., Nielsen R., Willerslev E. 2015a. The ancestry and affiliations of Kennewick Man. *Nature*. 523:455–458.
- Rasmussen S., Allentoft M.E.E., Nielsen K., Orlando L., Sikora M., Sjögren K.-G., Pedersen A.G., Schubert M., Van Dam A., Kapel C.M.O., Nielsen H.B., Brunak S., Avetisyan P., Epimakhov A., Khalyapin M.V., Gnuni A., Kriiska A., Lasak I., Metspalu M., Moiseyev V., Gromov A., Pokutta D., Saag L., Varul L., Yepiskoposyan L., Sicheritz-Pontén T., Foley R.A., Lahr M.M., Nielsen R., Kristiansen K., Willerslev E. 2015b. Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell*. 163:571–582.
- 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., Maricic T., Good J.M., Marques-Bonet T., Alkan C., Fu Q., Mallick S., Li H., Meyer M., Eichler E.E., Stoneking M., Richards M.P., Talamo S., Shunkov M. V., Dereviako A.P., Hublin J.-J., Kelso J., Slatkin M., Pääbo S. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*. 468:1053–1060.
- Reich D., Thangaraj K., Patterson N., Price A.L., Singh L. 2009. Reconstructing Indian population history. *Nature*. 461:489–494.
- Rodrigo A.G., Felsenstein J. 1999. Coalescent approaches. *Evol. HIV*:233.
- Rosenbaum M., Knight R., Leibel R.L. 2015. The gut microbiota in human energy homeostasis and obesity. *Trends Endocrinol. Metab.* 26:493–501.
- Sabeti P.C., Schaffner S.F., Fry B., Lohmueller J., Varilly P., Shamovsky O., Palma A., Mikkelsen T.S., Altshuler D., Lander E.S. 2006. Positive natural selection in the human lineage. *Science*. 312:1614–1620.
- Sankararaman S., Patterson N., Li H., Pääbo S., Reich D. 2012. The date of interbreeding between Neandertals and modern humans. *PLoS Genet.* 8:e1002947.
- Der Sarkissian C., Allentoft M.E.E., Ávila-Arcos M.C., Barnett R., Campos P.F., Cappellini E., Ermini L., Fernández R., da Fonseca R.R., Ginolhac A., Hansen A.J., Jónsson H., Korneliusson T.S., Margaryan A., Martin M.D., Moreno-Mayar J.V., Raghavan M., Rasmussen M., Velasco M.S., Schroeder H., Schubert M., Seguin-Orlando A., Wales N., Gilbert M.T.P., Willerslev E., Orlando L. 2015a. Ancient genomics. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370:20130387.
- Der Sarkissian C., Ermini L., Jónsson H., Alekseev A.N., Crubézy E., Shapiro B., Orlando L. 2014. Shotgun microbial profiling of fossil remains. *Mol. Ecol.* 23:1780–1798.
- Der Sarkissian C., Ermini L., Schubert M., Yang M.A., Librado P., Fumagalli M., Jónsson H., Bar-Gal G.K., Albrechtsen A., Vieira F.G., Petersen B., Ginolhac A., Seguin-Orlando A., Magnussen K., Fages A., Gamba C., Lorente-Galdos B., Polani S., Steiner C., Neuditschko M., Jagannathan V., Feh C., Greenblatt C.L., Ludwig A., Abramson N.I., Zimmermann W., Schafberg R., Tikhonov A., Sicheritz-Ponten T., Willerslev E., Marques-Bonet T., Ryder O.A., McCue M., Rieder S., Leeb T., Slatkin M., Orlando L. et al. 2015b. Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr. Biol.* 25:2577–2583.
- Sawyer S., Krause J., Guschanski K., Savolainen V., Pääbo S. 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One*. 7:e34131.
- Sawyer S., Renaud G., Viola B., Hublin J.-J., Gansauge M.-T., Shunkov M. V., Dereviako A.P., Prüfer K., Kelso J., Pääbo S. 2015. Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc. Natl. Acad. Sci. USA* 112:15696–15700.
- Scally A., Durbin R. 2012. Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* 13:745–753.
- Schiffels S., Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46:919–925.
- Schiffels S., Haak W., Paajanen P., Llamas B., Popescu E., Loe L., Clarke R., Lyons A., Mortimer R., Sayer D., Tyler-Smith C., Cooper A., Durbin R. 2016. Iron age and Anglo-Saxon genomes from East England reveal British migration history. *Nat. Commun.* 7:10408.
- Schraiber J.G., Evans S.N., Slatkin M. 2016. Bayesian inference of natural selection from allele frequency time series. *Genetics*. 203:493–511.
- Schroeder H., Ávila-Arcos M.C., Malaspinas A.-S., Poznik G.D., Sandoval-Velasco M., Carpenter M.L., Moreno-Mayar J.V., Sikora M., Johnson P.L.F., Allentoft M.E., Samaniego J.A., Haviser J.B., Dee M.W., Stafford T.W., Salas A., Orlando L., Willerslev E., Bustamante C.D., Gilbert M.T.P. 2015. Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proc. Natl. Acad. Sci. USA* 112:3669–3673.
- Schubert M., Ermini L., Der Sarkissian C., Jónsson H., Ginolhac A., Schaefer R., Martin M.D., Fernández R., Kircher M., McCue M., Willerslev E., Orlando L., et al. 2014a. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* 9:1056–1082.
- Schubert M., Ginolhac A., Lindgreen S., Thompson J.F., Al-Rasheid K.A.S., Willerslev E., Krogh A., Orlando L. 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics*. 13:178.
- Schubert M., Jónsson H., Chang D., Der Sarkissian C., Ermini L., Ginolhac A., Albrechtsen A., Dupanloup I., Foucal A., Petersen B., Fumagalli M., Raghavan M., Seguin-Orlando A., Korneliusson T.S., Velazquez A.M. V., Stenderup J., Hoover C.A., Rubin C.-J., Alfarhan A.H., Alquraishi S.A., Al-Rasheid K.A.S., MacHugh D.E., Kalbfleisch T., MacLeod J.N., Rubin E.M., Sicheritz-Ponten T., Andersson L., Hofreiter M., Marques-Bonet T., Gilbert M.T.P., Nielsen R., Excoffier L., Willerslev E., Shapiro B., Orlando L. 2014b. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. USA* 111:201416991.
- Schuenemann V.J., Singh P., Mendum T.A., Krause-Kyora B., Jäger G., Bos K.I., Herbig A., Economou C., Benjak A., Busso P., Nebel A., Boldsen J.L., Kjellström A., Wu H.H.H.T., Stewart G.R., Taylor G.M., Bauer P., Lee O.Y.-C.O.Y.-C., Wu H.H.H.T., Minnikin D.E., Besra G.S., Tucker K., Roffey S., Sow S.O., Cole S.T., Nieselt K., Krause J. 2013. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science*. 341:179–183.

- Segata N., Waldron L., Ballarini A., Narasimhan V., Jousson O., Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods*. 9: 811–814.
- Seguin-Orlando A., Gamba C., Der Sarkissian C., Ermini L., Louvel G., Boulygina E., Sokolov A., Nedoluzhko A., Lorenzen E.D., Lopez P.M., McDonald H.G., Scott E., Tikhonov A., Stafford T.W., Alfarhan A.H., Alquraishi S. a., Al-Rasheid K. a. S., Shapiro B., Willerslev E., Prokhortchouk E., Orlando L. 2015. Pros and cons of methylation-based enrichment methods for ancient DNA. *Sci. Rep.* 5: 11826.
- Seguin-Orlando A., Korneliusen T.S., Sikora M., Malaspina A.-S., Manica A., Moltke I., Albrechtsen A., Ko A., Margaryan A., Moiseyev V., Goebel T., Westaway M., Lambert D.M., Khartanovich V., Wall J.D., Nigst P.R., Foley R.A., Lahr M.M., Nielsen R., Orlando L., Willerslev E. 2014. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science*. 346: 1113–1118.
- Ségurel L., Wyman M.J., Przeworski M. 2014. Determinants of mutation rate variation in the human germline. *Annu. Rev. Genomics Hum. Genet.* 15:47–70.
- Shapiro B., Drummond A.J., Rambaut A., Wilson M.C., Matheus P.E., Sher A. V., Pybus O.G., Gilbert M.T.P., Barnes I., Binladen J., Willerslev E., Hansen A.J., Baryshnikov G.F., Burns J.A., Davydov S., Driver J.C., Froese D.G., Harington C.R., Keddie G., Kosintsev P.A., Kunz M.L., Martin L.D., Stephenson R.O., Storer J., Tedford R., Zimov S., Cooper A. 2004. Rise and fall of the Beringian steppe bison. *Science*. 306:1561–1565.
- Shapiro B., Ho S.Y.W., Drummond A.J., Suchard M. a., Pybus O.G., Rambaut A. 2011. A Bayesian phylogenetic method to estimate unknown sequence ages. *Mol. Biol. Evol.* 28:879–887.
- Sharpton T.J. 2014. An introduction to the analysis of shotgun metagenomic data. *Front. Plant Sci.* 5:209.
- Sjodin P., Skoglund P., Jakobsson M. 2014. Assessing the maximum contribution from ancient populations. *Mol. Biol. Evol.* 31:1248–1260.
- Skoglund P., Ersmark E., Palkopoulou E., Dalén L. 2015. Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds. *Curr. Biol.* 25:1515–1519.
- Skoglund P., Malmström H., Omrak A., Raghavan M., Valdiosera C., Günther T., Hall P., Tambets K., Parik J., Sjögren K.-G., Apel J., Willerslev E., Storå J., Götherström A., Jakobsson M. 2014a. Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science*. 344:747–750.
- Skoglund P., Malmström H., Raghavan M., Storå J., Hall P., Willerslev E., Gilbert M.T.P., Götherström A., 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., Jakobsson M. 2014b. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. USA* 111:2229–2234.
- Skotte L., Korneliusen T.S., Albrechtsen A. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics*. 195:693–702.
- Smith J., Kronforst M.R. 2013. Do Heliconius butterfly species exchange mimicry alleles? *Biol. Lett.* 9:20130503.
- Soubrier J., Steel M., Lee M.S.Y., Der Sarkissian C., Guindon S., Ho S.Y.W., Cooper A. 2012. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol. Biol. Evol.* 29:3345–3358.
- Stadler T., Kühnert D., Bonhoeffer S., Drummond A.J. 2013. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci. USA* 110:228–233.
- Stadler T., Yang Z. 2013. Dating phylogenies with sequentially sampled tips. *Syst. Biol.* 62:674–688.
- Sverrisdóttir O.Ó., Timpson A., Toombs J., Lecocq C., Froguel P., Carretero J.-M., Ferreras J.L.A., Götherström A., Thomas M.G., Arsuaga Ferreras J.L. 2014. Direct estimates of natural selection in Iberia indicate calcium absorption was not the only driver of lactase persistence in Europe. *Mol. Biol. Evol.* 31:975–983.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 123:585–595.
- Taubenberger J.K., Reid A.H., Janczewski T.A., Fanning T.G. 2001. Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 356: 1829–1839.
- Tito R.Y., Knights D., Metcalf J., Obregon-Tito A.J., Cleland L., Najjar F., Roe B., Reinhard K., Sobolik K., Belknap S., Foster M., Spicer P., Knight R., Lewis Jr C.M., et al. 2012. Insights from Characterizing Extinct Human Gut Microbiomes. *PLoS One*. 7:e51146.
- Truong D.T., Franzosa E.A., Tickle T.L., Scholz M., Weingart G., Pasolli E., Tett A., Huttenhower C., Segata N. 2015. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat. Methods*. 12:902–903.
- Vernot B., Tucci S., Kelso J., Schraiber J.G., Wolf A.B., Gittelman R.M., Dannemann M., Grote S., McCoy R.C., Norton H., Scheinfeldt L.B., Merriwether D.A., Koki G., Friedlaender J.S., Wakefield J., Paabo S., Akey J.M. 2016. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science*. 352:235–239.
- Vieira F.G., Fumagalli M., Albrechtsen A., Nielsen R. 2013. Estimating inbreeding coefficients from NGS data: Impact on genotype calling and allele frequency estimation. *Genome Res.* 23: 1852–1861.
- Vila C., Leonard J.A., Götherström A., Marklund S., Sandberg K., Lidén K., Wayne R.K., Ellegren H. 2001. Widespread origins of domestic horse lineages. *Science*. 291:474–477.
- Warinner C., Rodrigues J.F.M., Vyas R., Trachsel C., Shved N., Grossmann J., Radini A., Hancock Y., Tito R.Y., Fiddyment S., Speller C., Hendy J., Charlton S., Luder H.U., Salazar-García D.C., Eppler E., Seiler R., Hansen L.H., Castruita J.A.S., Barkow-Oesterreicher S., Teoh K.Y., Kelstrup C.D., Olsen J. V., Nanni P., Kawai T., Willerslev E., von Mering C., Lewis Jr C.M., Collins M.J., Gilbert M.T.P., Rühli F., Cappellini E., et al. 2014. Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* 46:336–344.
- Watterson G.A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7:256–276.
- Wickham H. 2009. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.
- Wilde S., Timpson A., Kirsanow K., Kaiser E., Kayser M., Unterländer M., Hollfelder N., Potekhina I.D., Schier W., Thomas M.G., Burger J. 2014. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl. Acad. Sci. USA* 111:4832–4837.
- Willerslev E., Cappellini E., Boomsma W., Nielsen R., Hebsgaard M.B., Brand T.B., Hofreiter M., Bunce M., Poinar H.N., Dahl-Jensen D., Johnsen S., Steffensen J.P., Bennike O., Schwenninger J.-L., Nathan R., Armitage S., de Hoog C.-J., Alfimov V., Christl M., Beer J., Muscheler R., Barker J., Sharp M., Penkman K.E.H., Haile J., Taberlet P., Gilbert M.T.P., Casoli A., Campani E., Collins M.J. 2007. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science*. 317:111–114.
- Willerslev E., Davison J., Moora M., Zobel M., Coissac E., Edwards M.E., Lorenzen E.D., Vestergård M., Gussarova G., Haile J., Craine J., Gielly L., Boessenkool S., Epp L.S., Pearman P.B., Cheddadi R., Murray D., Bråthen K.A., Yoccoz N., Binney H., Cruaud C., Wincker P., Goslar T., Alsos I.G., Bellemain E.P., Brysting A.K., Elven R., Sønstebo J.H., Murton J., Sher A., Rasmussen M., Rønn R., Mourier T., Cooper A., Austin J.J., Möller P., Froese D., Zazula G., Pompanon F., Rioux D., Niderkorn V., Tikhonov A., Savvinov G., Roberts R.G., MacPhee R.D.E., Gilbert M.T.P., Kjær K.H., Orlando L., Brochmann C., Taberlet P. 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*. 506:47–51.
- Willerslev E., Hansen A.J., Christensen B., Steffensen J.P., Arctander P. 1999. Diversity of Holocene life forms in fossil glacier ice. *Proc. Natl. Acad. Sci. USA* 96:8017–8021.
- Willerslev E. 2003. Diverse Plant and animal genetic records from Holocene and Pleistocene sediments. *Science*. 300:791–795.
- Wiuf C., Hein J. 1999. Recombination as a point process along sequences. *Theor. Popul. Biol.* 55:248–259.
- Wu C.-I., Ting C.-T. 2004. Genes and speciation. *Nat. Rev. Genet.* 5:114–122.
- Yang M.A., Harris K., Slatkin M. 2014. The projection of a test genome onto a reference population and applications to humans and archaic hominins. *Genetics*. 198:1655–1670.

- Yang M.A., Slatkin M. 2015. Using ancient samples in projection analysis. G3 (Bethesda).
- Yoshida K., Schuenemann V.J., Cano L.M., Pais M., Mishra B., Sharma R., Lanz C., Martin F.N., Kamoun S., Krause J., Thines M., Weigel D., Burbano H.A. 2013. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *Elife*. 2: e00731.
- Ziesemer K.A., Mann A.E., Sankaranarayanan K., Schroeder H., Ozga A.T., Brandt B.W., Zaura E., Waters-Rist A., Hoogland M., Salazar-García D.C., Aldenderfer M., Speller C., Hendy J., Weston D.A., MacDonald S.J., Thomas G.H., Collins M.J., Lewis C.M., Hofman C., Warinner C. 2015. Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Sci. Rep.* 5: 16498.