Genomic approaches and their contributions to understanding the European Neolithisation

Cristina Gamba
Centre for GeoGenetics, Natural History Museum of Denmark, Copenhagen, DK
cristinagamba@gmail.com

ABSTRACT – *The contribution of ancient DNA to the understanding of past events has been increasing exponentially in recent years. This is mainly due to the synergy of technical advances, such as the molecular technique of high-throughput DNA sequencing, which has allowed for the reconstruction of complete genomes as old as 750 000 years. Another step toward the cost-effective characterisation of ancient genomes is the sampling of petrous bone, which has allowed sequencing of the first ancient African genome. Here I review the significant contribution of ancient genomics to our understanding of the European Neolithisation process.*

IZVLEČEK – *V zadnji letih se je eksponentno povečal prispevek stare DNK k razumevanju preteklosti, kar je predvsem posledica sinergije tehničnih napredkov, npr. molekularne diagnostike visoke zahtevnosti DNK sekvence, s katero lahko rekonstruiramo celoten genom tudi do 750 000 let starih vzorcev. Drug korak k stroškovno učinkovitejši karakterizaciji starega genoma je vzorčenje kosti skalnice v lobanji, s katero smo lahko dobili sekvenco prvega starega afriškega genoma. V članku nudim pregled nad glavnimi prispevki raziskave starega genoma k razumevanju evropskega procesa neolitizacije.*

KEY WORDS – *ancient DNA; palaeogenomics; Neolithic*

Introduction

The field of ancient DNA was 30 years old in 2014, and the anniversary was celebrated with an international conference at the Royal Society of London entitled "*Ancient DNA: the first three decades*" (*Hagelberg* et al. *2015*). This conference was dominated by the most recent results in the field, including an extensive genome-wide proof of Neanderthal interbreeding with modern humans (*Sankararaman* et al. *2014*). However, ancient DNA has entered the 'genomic era' only very recently, with the first ancient human genome published in 2010 (*Rasmussen* et al. *2010*), the first scattered genomic data from a Neolithic sample in 2012 (*Skoglund* et al. *2012*), and the first Near Eastern Neolithic genomes only very recently (*Mathieson* et al. *2015*).

Before the 'genomic era', the field advanced thanks to the complicated and low-success recovery-rate of small DNA fragments, mainly from mitochondrial DNA (mtDNA). The mitochondria are cell organelles that govern the energy supply, in several copies per cell. They contain one or more small circular molecules of DNA; therefore, 1000 to 10 000 copies of mtDNA can be counted in a cell, as compared to only two copies of maternally and paternally inherited nuclear DNA (nuDNA) (*Robin, Wong 1988*). This has made mtDNA the preferred target for ancient DNA studies, given >1000 higher probability of recovering DNA after the environmental degradation that affects cells and molecules after death.

The first mtDNA sequence from ancient material was isolated in 1984 from the dried muscle of an early 1900s quagga, an extinct equid (*Higuchi* et al. *1984*). Since then, a wide range of other ancient samples and tissues have been analysed, especially ancient bones and teeth, the most frequently preserved tissues in ancient samples, with the seminal work of

DOI: 10.4312/dp.43.12

Hagelberg and colleagues (*Hagelberg* et al. *1989*). Other tissues such as hair (*Gilbert* et al. *2004*), coprolites (*Poinar* et al. *1998*), and arctic sediments (*Willerslev* et al. *2003*) were also successfully investigated in the first two decades of the aDNA field.

In the specific research on the Neolithic transition in Europe, several studies based on short mtDNA sequences were published, with the pioneering work of Wolfgang Haak and colleagues (*Haak* et al. *2005*) on Neolithic LBK (samples from Central Europe). This study was the first population survey of an ancient Neolithic group to help us understand the Neolithisation process. The authors showed an ancient Neolithic genetic composition very different from the current European population, including variants virtually absent nowadays (*i.e.* haplogroup N1a). Several other mtDNA works followed, including a comparison of contemporaneous farmers and hunter-gatherers in Northern Europe ion the 3rd millennium BC (*Malmström* et al. *2009*), revealing a genetic break. For the Iberian Peninsula, Cristina Gamba and colleagues (*Gamba* et al. *2011*) also showed an important component of new incomers at the beginning of the Neolithic.

The jump to the second-generation or the 'genomic era' has allowed a quantum leap in retrieving genetic information from ancient specimens. After the first ancient human genome to be sequenced, a 4000 years old specimen from Greenland called 'Saqqaq' after his culture (*Rasmussen* et al. *2010*), many other ancient genomes followed. Most of the studies included one or very few samples, due to the high cost of high-throughput DNA sequencing (HTS) of whole genomes (3 billion base pairs for the human genome – around 200 000 times longer than the whole mtDNA). This is especially true for degraded samples with endogenous DNA below 5%, the other 95% being mostly co-purified microbial DNA from the environment and, unfortunately, as costly as the endogenous DNA fraction. Basically, to sequence the genome of an ancient sample costs at least 20 times more than a modern one, restricting the whole-genome sequencing of ancient sample to a few highly funded laboratories and to very well-preserved specimens.

Ancient DNA challenges and validation

In order to understand aDNA data, it is essential to understand the challenges involved in performing a molecular analysis based on degraded material. Firstly, the DNA starts to degrade immediately after the organism dies due to the action of microorganisms in the depositional environment, enzymes released from the cells themselves and chemical reactions due to the presence of oxygen and water (*Hofreiter* et al. *2001*). The main outcomes are DNA fragmentation and chemical changes. The first implies that only short sequences can be retrieved from ancient samples, and a sequence is informative only above 25–30 base pairs, because below this length the probability of a precise match between the DNA fragment in different locations of a genome or even in different genomes significantly increases, making the genetic data uninformative.

The main chemical DNA modification known to date is the deamination at cytosines, one of the four nucleotide bases composing the DNA, which implies their transformation into uraciles (usually only found in RNA and not in the DNA of living organisms), and ultimately into thymines through the amplification process, which is necessary for sequencing DNA (*Hofreiter* et al. *2001*). This means that we detect a thymine, which was originally a cytosine, which renders misleading the genetic information retrieved. This happens mainly around DNA breaks, thus at the beginning or end of DNA fragments. Consequently, the deamination might lead to confusion concerning the correct retrieval of genetic information, unless many fragments from the same genomic region are analysed and compared (high coverage), given that deamination should happen randomly.

On the other hand, detecting deamination provides a unique signature of DNA degradation and has been largely used as a criterion from ancient DNA (*Ginolhac* et al. *2011*). Interestingly, there is an experimental way to eliminate these miscoding lesions from the DNA strands, by using an enzyme (UDG, Uracil-DNA-Glycosylase) that can remove the majority of the uracils from ancient DNA before amplification, making downstream analysis less noisy (see *e.g.*, *Rohland* et al. *2015*).

Another significant weakness of aDNA, especially before the HTS era, is its susceptibility to contamination from fresh genetic material, either from the depositional environment or introduced after an excavation during manipulation of the sample. This undermined many of the first aDNA studies, which claimed to recover sequences from material millions of years old, such as a magnolia leaf claimed to be 17–20 million years old (*Golenberg* et al. *1990*), amber-preserved bees, termites and plants (*Cano* et al. *1993; DeSalle* et al. *1992; Poinar* et al. *1993*), and

even fossil bacteria (*Cano* et al. *1994*). These studies were later shown to be fallacious, as only contaminant DNA of modern origin was actually retrieved (*Poinar* et al. *1998; Yousten, Rippere 1997*).

All these million-year studies revealed the high risk of combining a highly sensitive technique, such as PCR, with the study of scarce and degraded ancient molecules. This led to outsiders taking a very sceptical view of the field, especially archaeologist and palaeontologists unwilling to destroying precious samples in order to obtain unreliable results.

However, this was an important turning point in the aDNA field, highlighting the importance of authenticating and validating results. The so-called 'authenticity criteria' were formalised for the first time by Matthias Krings and colleagues through the validation of the DNA results obtained from the first Neanderthal specimen analysed (*Krings* et al. *1997*), followed by several comprehensive reviews of all the criteria needed to validate results (*Cooper, Poinar 2000; Pääbo* et al. *2004*). Despite its controversial beginnings, ancient DNA can now be considered a well-established scientific discipline, with greatly improved standards, especially since the development of HTS.

A highly valuable piece: the petrous bone

Recently, Gamba *et alii* (*2014*) demonstrated that the densest area of the petrous bone, part of the temporal bone, could provide very high levels of endogenous DNA, up to 180-times more than other bone pieces from the same individual, preserving up to more than 80% of the ancient endogenous DNA of the specimen. This implies a sequencing cost only slightly higher than modern genomes, which are currently systematically generated by the hundreds. This is linked to the high density of this bone in mammals (*Lam* et al. *1999*), which might be responsible for the low penetration of bacteria during the decomposition processes of the body. After this seminal study, different parts of the petrous bone were shown to provide lower amounts of DNA (*Pinhasi* et al. *2015*) than the densest part chosen in the first case (*Gamba* et al. *2014*).

In the latter study, the closest samples to the petrous bone in terms of DNA preservation were the molar roots, only 3-times less efficient (*Gamba* et al. *2014*). In another study, Peter B. Damgaard and colleagues demonstrated that the cementum of teeth is where the best-preserved genetic material is harboured (*Damgaard* et al. *2015*). However, although especially well-preserved teeth can also have very high levels of endogenous DNA, on average they do not perform nearly as well as petrous bone (Fig. 1).

Thanks to the specific sampling of the petrous bone, it became possible to retrieve and sequence the genome of a sample from an especially degrading environment such as Africa for the first time (*Gallego-Llorente* et al. *2015*), and more recently, even a handful of the first farmers from the Near East (*Broushaki* et al. *2016; Gallego-Llorente* et al. *in press; Lazaridis* et al. *2016*)

Palaeogenomics and the Neolithic transition in Europe

Ancient genomes from prehistoric Europeans have provided highly valuable insights into the understanding of the population dynamics of the past. This

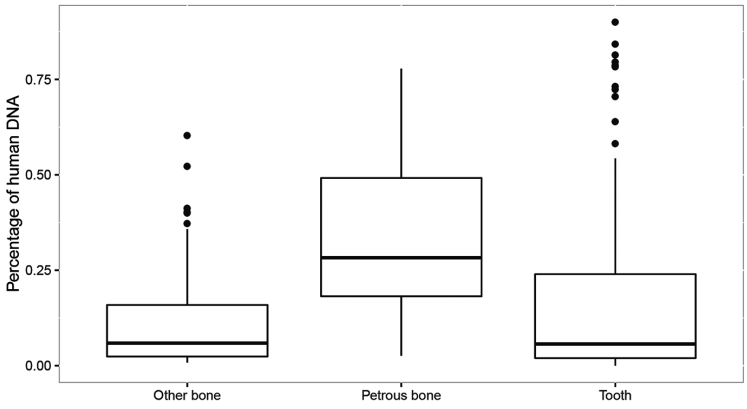

Fig. 1. Boxplot representing the percentage of endogenous human DNA retrieved from the petrous bone versus other skeletal elements retrieved from ancient specimens. The bottom and top of the box refer to the first and third quartiles; the band inside the box is the median, while the ends of the whiskers represent the 5th and 95th percentiles. Outliers are represented as points. Database from Mathieson and colleagues (*2015*), which also includes data (*Allentoft *et al.* 2015; Gamba *et al.* 2014; Haak *et al.* 2015; Keller *et al.* 2012; Lazaridis *et al.* 2014; Olalde *et al.* 2014) completed with these data (*Cassidy *et al.* 2016; Günther *et al.* 2015; Jones *et al.* 2015; Omrak *et al.* 2016; Skoglund *et al.* 2014; Broushaki *et al.* 2016; Gallego-Llorente *et al.* in press). Only data at a minimum coverage of 0.025x were included. Plot realised in R (*Core Team 2014*) using the library ggplot2 (*Wickham 2010*).

is especially true for the Neolithic period, for which many complete genomes or genome-wide data are now available (Fig. 2).

The Neolithisation of Europe has been thoroughly investigated by different disciplines (*i.e. Bocquet-Appel, Bar-Yosef 2008; Pinhasi, Stock 2011*) which showed the relative contribution of demographic movements and the transmission of ideas and new practices that accompanied the economic revolution (*Zvelebil 2001*).

Investigations of this process with modern DNA data has yielded contradictory results (see *Jobling* et al. *2013.Ch. 12 and references therein*) because of the specificity of the genetic markers studied and the important genetic reshaping of the European genetic pool after the Neolithic, recently directly detected through ancient genomics (see *Mathieson* et al. *2015* and references therein). However, it seems that now this long-lasting debate on the mechanisms of Neolithic expansion has been resolved thanks to ancient genetics and genomics. Ancient genomes support the first studies based on mtDNA (*Bramanti* et al. *2009; Haak* et al. *2010; Brandt* et al. *2013; Szécsényi-Nagy* et al. *2015*), pointing to a genetic break between the Mesolithic and the Neolithic periods in Europe, with the arrival of new incomers from the Near East.

The first clue to the genomic landscape of Neolithic Europeans was provided from the complete sequencing of a frozen mummy from the Copper Age (5300 BP) found in the Italian Alps (*Keller* et al. *2012*). The so-called Iceman surprisingly showed high genetic affinities with modern-day Sardinians, despite their geographic distance. These data suggested for the first time that (1) major genomic reshaping occurred after the Neolithic and (2) Sardinians might be relics of the original Neolithic population, because of the late peopling of the island and the implicit geographic isolation.

Understanding palaeogenomics

Molecular approaches

Palaeogenomic analyses start in the laboratory, where a sample, usually a piece of bone, is ground and the DNA extracted. The short DNA molecules retrieved are then built into so-called *libraries* by attaching to each side of the DNA strand two sequences, the *adapters*, which contain well-known DNA stretches compatible with further processing. The library is then amplified and (usually at this stage) short unique DNA, including unique *indexes*, are attached to the adapters to tag each sample differently and to allow the sequencing machine to distinguish between samples. The amplified libraries are then either sequenced or enriched. In the first case, the whole DNA molecules extracted from the sample are sequenced. This method is called *shotgun sequencing*, because there is no selection of the DNA to be sequenced, and then, apart from the endogenous DNA of interest, also DNA from the environment (contaminant DNA) is co-sequenced. In the second case, the amplified library is firstly combined with a set of DNA sequences, so-called *probes*, which *capture* regions of interest, while all the other molecules are washed away. In this case, principally these regions are selected, *enriched* and then sequenced, reducing sequencing costs while increasing the coverage of these positions of interest. The drawback of this second approach is that the genome to be sequenced needs to be well characterised in order to produce specific probes for DNA *enrichment*.

Analytical approaches

Given the massive amount of sequences retrievable from HTS machines, *bioinformatic* tools should be applied to the data analysis and interpretation. The first step is to align the sequence retrieved to a *reference genome*, whenever available. Sometimes, a reference genome is not available, *e.g.*, in the case of species not characterised or extinct; therefore, the genome should be *assembled* from scratch. For human samples, a very well characterised genome is available and constantly improved. So the first step consists of *aligning* the sequences to the human reference genome, followed by *quality filtering*, which takes into account how well a genome is aligned to the reference (*mapping quality*) and how confidently each base was identified during the sequencing (*base quality*). This is followed by *genotyping*, mainly focusing on the identification and analysis of punctual genomic variants, also called *SNPs* (Single Nucleotide Polymorphism), which are positions in the genome that vary among different individuals and populations. Further filters can be applied, such as the setting of a minimum *coverage* (how many times a position has been sequenced), which reflects

Further genomic data from northern European farmers and hunter-gatherers extended to much higher latitudes and much later periods (~5000 years old) showed the affinity of ancient farmers with modern-day southern Europeans (*Skoglund* et al. *2012*). On the other hand, northern hunter-gatherers showed genetic discontinuity with farmers, falling outside the genomic variability of modern Europeans (*Skoglund* et al. *2012*). Despite the very low amount of genomic data retrieved, covering only ~10% of the genome, this study allowed for the correlation of two different genomic backgrounds with two differentiated cultural groups: farmers and hunter-gatherers, suggesting the arrival of new incomers with the advent of the Neolithic.

Later studies showed the same pattern at lower latitudes, such as the characterisation of two high-quality genomes, one Neolithic farmer from Germany and one Mesolithic hunter-gatherer from Luxembourg, further supported by eight lower-quality genomes from Swedish hunter-gatherers (*Lazaridis* et al. *2014*).

The study of a time series from the Early Neolithic to the Iron Age in Eastern Hungary (*Gamba* et al. *2014*), provided a pool of nine new Neolithic genomes (one high-quality, NE1), including two early Neolithic individuals associated with the Körös culture. One of them, KO2, showed affinities with the Neolithic pool, close to modern-day Sardinians, with some Near-Eastern influence. Surprisingly, the other Körös sample, KO1, fell outside modern-day variability, clustering together with hunter-gatherers' genomes, despite the Neolithic cultural context. This finding points to possible admixture between hunter-gatherers and farmers at the beginning of the Neolithic. The following investigations mainly focused on genome-wide informative markers (*Haak* et al. *2015; Mathieson* et al. *2015*), allowing for the characterisation of more than one hundred Neolithic genomes to date.

how confidently the genotyping was assigned. *High-quality* or *high-coverage* genomes usually refer to genomes that have been sequenced with an average coverage of the whole genome of 20x and above. Specific filters that take into account post-mortem deamination reactions are frequently applied to ancient genomic data, *i.e.* considering only SNPs that involve transversions (from C or G to T or A and *vice versa*), which cannot derive from deamination events.

The genomic variability can be summarised and compared to other individuals with *Principal Component Analyses* (PCA), relying on those components that best explain the diversity between samples.

So-called *admixture plots* are also frequently used to visualise with stacked ancestry components in barplots. The number of these components is set *a priori* and can identify portions of ancestry related to geographic distributions, temporal periods or ethnicities shared among different individuals.

Relationships between individuals can also be drawn by adopting phylogenetic approaches. One of the most interesting approaches used in palaeogenomics relies on whole genome data and allows for the representation of a phylogenetic tree including arrows between branches, pointing to possible admixture events (*e.g.* interbreeding between species or inbreeding between unrelated populations).

Many methods have been specifically developed to identify admixture events among populations or species including ancient DNA. These rely on different statistical tests, such as so-called D-statistics and the f-statistics.

Adaptation to the environment and the selection of advantageous variants (positive selection) can also be tested at a population or species level, and rely on the comparison between patterns of genetic variability (such as FST, which provides a measure of population differentiation).

All the genomic analyses described above can also be implemented on either *genotype likelihood* or *imputed genotypes* (genotype likelihood incorporating information from available databases), which provide *genotype probabilities* instead of observed genotypes. This is especially interesting for ancient genomes, usually showing few data mined by molecular damage, which are not suitable for extensive genotyping.

Further reading

For a review of aDNA molecular methods see Shapiro, Hofreiter (*2012*), and for an updated and comprehensive review of aDNA analytical tools see Leonardi *et al.* (*2016*).

Within the Neolithic pool, there is a Southeast-Northwest cline, with a decreasing Near-Eastern affinity, pointing to a dilution of the original gene pool along with the expansion. The same trend can be detected in modern European populations, overlapping with an East-West gradient due to the influence of Bronze Age incomers from the Steppe (*Mathieson* et al. *2015*) at later stages.

It is worth noting that the Neolithic cluster also includes a Spanish genome associated with the Cardial Neolithic cultural expansion, the first Neolithic incomers into Mediterranean regions, suggesting a common origin with the LBK Neolithic culture, which spread in parallel into Central Europe (*Olalde* et al. *2015*). Moreover, it was possible to identify an eastern European hunter-gatherer component in this Cardial genome (*ibid.*), reinforcing the hypothesis of a certain admixture of hunter-gatherers and farmers at the beginning of the Neolithic expansion.

Similarly, Zuzana Hofmanová and colleagues (*2016*) detected a low-level admixture of migrating farmers and local hunter-gatherers in the earliest stages of the Neolithic, consistent with sporadic occurrences.

The admixture with local hunter-gatherers increased substantially at later stages (*Haak* et al. *2015; Hofmanová* et al. *2016*) at the transition to the Middle Neolithic across Europe, while Late Neolithic and Bronze Age periods were characterised by increasing input from steppe populations (*Haak* et al. *2015*).

The analysis of north-western Anatolian Neolithic samples from the Marmara region in Turkey (*Mathieson* et al. *2015*), which also clearly cluster within the ancient Europeans' Neolithic pool, confirmed that the source of the agricultural incomers reached Europe through northern Anatolia, and probably followed a route across Greece to Europe (*Hofmanová* et al. *2016*).

However, until very recently, there were no genomic data directly linked to the first appearance of Neolithic culture in the Fertile Crescent. Only short sequences from mitochondrial DNA were available from Pre-Pottery Neolithic samples (*Fernández* et al. *2014*). However, recently, whole genomes and genome-wide data from the Fertile Crescent region have become available, giving a first glimpse of the Near-Eastern genetic pool through time and space (*Broushaki* et al. *2016; Lazaridis* et al. *2016; Gallego-Llorente* et al. *in press*). The analysis of samples from such warm areas has now become possible thanks to the recovery of relatively high endogenous content from the petrous bone, used as the main DNA source in all three studies (see section above).

Four Early Neolithic (EN) genomes from Zagros in Iran show a distinct genetic signature from both European hunter-gatherers and farmers, close to modern Pakistanis and Afghans (*Broushaki* et al. *2016*). In this study, the authors suggested that the affinities of Zagros Neolithic individuals to modern pop-

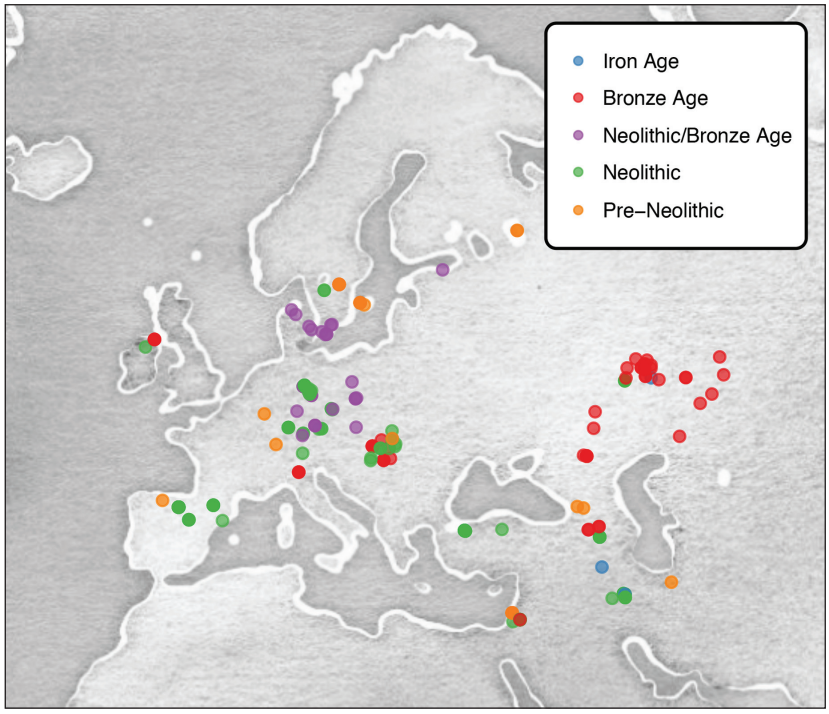

Fig. 2. Location of ancient human genomes sequenced at a minimum coverage of 0.025x (database* Mathieson *et al.* 2015 *which also includes data from* Allentoft *et al.* 2015; Gamba *et al.* 2014; Haak *et al.* 2015; Keller *et al.* 2012; Lazaridis *et al.* 2014; Olalde *et al.* 2014*; this database is completed with data from* Cassidy *et al.* 2016; Günther *et al.* 2015; Jones *et al.* 2015; Omrak *et al.* 2016; Skoglund *et al.* 2014; Broushaki *et al.* 2016; Gallego-Llorente *et al.* in press; Lazaridis *et al.* 2016*). Plot realised in R (*Core Team 2014*) using the libraries ggmap (*Kahle, Wickham 2013*) and ggplot2 (*Wickham 2010*).

ulations of southern Asia can be related to the spread of Indo-Iranian languages or Dravidian languages, along with the demographic expansion of farming into the region (*Broushaki* et al. *2016*). This study also pointed out that the European Neolithic migration probably had a different genetic source than the eastern Fertile Crescent.

Iosif Lazaridis and colleagues (*2016*) tried to answer to this question by analysing serial samples from different regions of the Near East, mainly from two areas, the Levant (Israel and Jordan) and Iran, for a total of 44 samples (also including some from Armenia in the Caucasus, and one from Turkey) spanning from the Epi-Paleolithic and Natufian (pre-Neolithic) to the Chalcolithic periods. The results retrieved from the Iranian samples supported the study by Farnaz Broushaki and colleagues (*2016*) and highlighted the high genetic differentiation of those samples not only with respect to the European Neolithic, but also to the Levantine pool. Interestingly, they also detected genetic continuity in both regions from pre-Neolithic to Neolithic periods, suggesting a major cultural spread of the Neolithic throughout the Near East (*Lazaridis* et al. *2016*).

The populations of the Levant in the Neolithic are genetically closer to the European and Anatolian Neolithic pool than the Iranian Neolithic, but nevertheless cluster separately (*Lazaridis* et al. *2016*). The authors identify the Levantine population as a good proxy for East African ancestry, pointing to the fact that the source population of the Neolithic expansion into Europe still remains to be identified (*Lazaridis* et al. *2016*). Further analysis from such crucial areas and periods will improve our understanding of the dynamics that influenced the initial development of the Neolithic period and its subsequent multidirectional expansion.

Although palaeogenomics data have significantly contributed to deciphering the mechanisms involved in European Neolithisation, this field would benefit from deeper interaction and integration with related disciplines, such as archaeology and anthropology.

Positive selection and phenotypes during the Neolithic

Genomic research is sufficiently advanced to allow the identification of the genes involved in several phenotypic attributes – including complex traits such as eye and hair colour – controlled by several genes. European hunter-gatherers had quite dark skin, dark hair and, interestingly, light-coloured eyes, while the incoming farmers typically had lighter skin and dark eyes (*Gamba* et al. *2014; Lazaridis* et al. *2014; Mathieson* et al. *2015; Olalde* et al. *2014*).

A recent study specifically focused on detecting the positive selection of phenotypes through time (*Mathieson* et al. *2015*) identified several candidates. The authors investigated the temporal progression of allele frequency of those genes for spotting the selection timing and consequences.

A strong signal of selection was detected for the light skin pigmentation variant of the gene SLC45A2 (SNP rs16891982), now almost fixed in Europeans, significantly increasing in frequency through time. Another gene, SLC24A5, also associated with light skin pigmentation, was not identified as under selection in the analysis, but the selected allele frequency increased at the beginning of the Neolithic in Europe, probably due to the migration pattern from the Near East.

The primary determinant of light eye colour is linked to the gene HERC2/OCA2 (SNP rs12913832). This has been found in all European hunter-gatherers, while at later stages up to the present, the allele associated with light eye colour increases with high latitudes, suggesting selection due to the environment.

Interestingly, the derived allele of the SNP rs3827760 in the gene EDAR, almost absent in present-day Europeans, was detected with high frequency in Scandinavian hunter-gatherers from Motala in Sweden (5898–5531 cal BC). This gene, which affects tooth morphology and hair thickness, is highly frequent in East Asian and Native Americans and has previously been suggested to have originated in East Asia (*Kamberov* et al. *2013*), but a different scenario emerge from these ancient data.

However, the strongest signal of selection was retrieved from the LCT gene (SNP rs4988235), associated with lactase persistence, which allows to adults to digest milk. The authors confirmed previous results (*Gamba* et al. *2014; Burger* et al. *2007; Allentoft* et al. *2015*) pointing out the late occurrence of this allele, which appeared for the first time only around 4000 years ago, much later than the advent of the Neolithic. The earliest date for lactase persistence is from a central European Bell Beaker sample dated to 2457–2142 cal BC (*Mathieson* et al.

2015). Other variants in genes associated with diet have been identified, suggesting components of adaptation to a variety of diets (fatty acid), different food sources and environments (vitamin D) and others possibly linked to coeliac disease (*Mathieson* et al. *2015*).

In this study, Iain Mathieson and colleagues (*2015*) were also able to investigate an even more complex trait which depends on hundreds of variants: height. A North-South cline in Europe is evident, with height decreasing in modern Europeans, probably linked to selection processes that occurred in the past reflecting better adaptation to the environment. Based on 169 genomic variants, the authors explained this gradient by detecting a significant signal of selection of reduced height in Iberian Neolithic and Chalcolithic samples, as well as increased height in steppe populations relative to the central European Neolithic. This suggests that the height gradient detected now is mainly due to the increased steppe ancestry of northern Europeans and selection for lower height in Southern Europe.

New perspectives in ancient genomics

Despite the short history of the aDNA field, the genomic boom has resulted in an explosion of ancient genomes from many species, including humans, archaic hominids, animals, and plants. Also, other satellite approaches are attracting more and more attention, providing new ways to study ancient genomics, such as ancient metagenomic and ancient epigenomic analysis. The first refers to the analysis of the whole sequencing output, the exogenous DNA, that vast amount of unused sequencing obtained through shotgun approaches, especially vast in those samples with very low amounts of endogenous DNA. Why should we be interested in such DNA portion? Because it might contain other interesting organisms, which might come from the depositional environment, from manipulation, or from other organisms, such as pathogens, that were inhabiting the specimen. Recent work has demonstrated that it is possible to collaterally detect and fully characterise Yersinia pestis, the agent of the plague, in ancient samples (*Rasmussen* et al. *2015*), pushing back the presence of this pathogen at least 3000 years before any historical record.

A very interesting substrate for metagenomic analysis is the dental calculus (*Adler* et al. *2013; Warinner* et al. *2014*), allowing for the investigation of ancient oral microbiomes, clearly detecting shifts correlating with dietary changes during the Neolithic and the Industrial revolutions (*Adler* et al. *2013*).

On the other hand, epigenomics studies DNA modifications that do not imply changes in the sequence, but only reversible modifications to the genetic material that mainly influence gene expression (*i.e.* how the genetic information coded in the genes is differently expressed in cells). One of the most common epigenetic modifications is DNA methylation, which is the addition of a methyl group to a nitrogenous base in specific genomic contexts. Recent studies have demonstrated that it is possible to retrieve this information from ancient genomes, by using either direct techniques (bisulphite sequencing, which is difficult to apply to ancient specimens, as it requires a large amount of DNA) (*Llamas* et al. *2012*), or by analysing patterns of molecular damage directly linked to methylation rates (*Gokhman* et al. *2014; Pedersen* et al. *2014*).

The applications of both epigenomics and metagenomics to ancient substrates pinpoint the progressive incorporation of new approaches to palaeogenomics, the result of a continuously updated multidisciplinary field.

ACKNOWLEDGEMENTS

I would like to thank the reviewers for their help in improving this manuscript and all my colleagues for their support. My contract was funded by the Marie-Curie Intra-European Fellowship program (FP7-IEF-328024).

References

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. and Cooper A. 2013. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nature Genetics 45: 450–455, 455e1*.

Allentoft M. E. and 65 co-authors. 2015. Population genomics of Bronze Age Eurasia. *Nature 522: 167–172*.

Bocquet-Appel J.-P., Bar-Yosef O. 2008. *The Neolithic Demographic Transition and its Consequences*. Springer Netherlands. Dordrecht.

Bramanti B. and 15 co-authors. 2009. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science 326: 137–140*.

Brandt G. and 17 co-authors. 2013. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science 342: 257–261*.

Broushaki F. and 31 co-authors. 2016. Early Neolithic genomes from the eastern Fertile Crescent. *Science: aaf7943*.

Burger J., Kirchner M., Bramanti B., Haak W. and Thomas M. G. 2007. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proceedings of the National Academy of Sciences USA 104: 3736–3741*.

Cano R. J., Borucki M. K., Higby-Schweitzer M., Poinar H. N., Poinar G. O. Jr. and Pollard K. J. 1994. Bacillus DNA in fossil bees: an ancient symbiosis? *Applied and Environmental Microbiology 60: 2164–2167*.

Cano R. J., Poinar H. N., Pieniazek N. J., Acra A. and Poinar G. O. Jr. 1993. Amplification and sequencing of DNA from a 120–135-million-year-old weevil. *Nature 363: 536–538*.

Cassidy L. M., Martiniano R., Murphy E. M., Teasdale M. D., Mallory J., Hartwell B. and Bradley D. G. 2016. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proceedings of the National Academy of Sciences USA 113: 368–373*.

Cooper A., Poinar H. N. 2000. Ancient DNA: Do It Right or Not at All. *Science 289: 1139*.

Core Team R. 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Damgaard P. B., Margaryan A., Schroeder H., Orlando L., Willerslev E. and Allentoft M. E. 2015. Improving access to endogenous DNA in ancient bones and teeth. *Scientific Reports 5: 11184*.

DeSalle R., Gatesy J., Wheeler W. and Grimaldi D. 1992. DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science 257: 1933–1936*.

Fernández E., Pérez-Pérez A., Gamba C., Prats E., Cuesta P., Anfruns J,. Molist M,. Arroyo-Pardo E. and Turbón D. 2014. Ancient DNA analysis of 8000 B.C. near eastern farmers supports an early neolithic pioneer maritime colonization of Mainland Europe through Cyprus and the Aegean Islands. *PLoS Genetics 10: e1004401*.

Gallego Llorente M. and 18 co-authors. 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science 350: 820–822*.

Gallego-Llorente M. G. and 15 co-authors. 2016. The genetics of an early Neolithic pastoralist from the Zagros, Iran. *Scientific Reports 6: 31326*.

Gamba C., Fernández E., Tirado M., Deguilloux M. F., Pemonge M. H., Utrilla P., Edo M., Molist M., Rasteiro R., Chikhi L. and Arroyo-Pardo E. 2012. Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. *Molecular Ecology 21: 45–56*.

Gamba C. and 16 co-authors. 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications 5: 5257*.

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. and Cooper A. 2004. Ancient mitochondrial DNA from hair. *Current Biology 14: R463–R464*.

Ginolhac A., Rasmussen M., Gilbert M. T. P., Willerslev E. and Orlando L. 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics 27: 2153–2155*.

Gokhman D., Lavi E., Prüfer K., Fraga M. F., Riancho J. A., Kelso J., Pääbo S., Meshorer E. and Carmel L. 2014. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science 344: 523–527*.

Golenberg E. M., Giannasi D. E., Clegg M. T., Smiley C. J., Durbin M., Henderson D. and Zurawski G. 1990. Chloroplast DNA sequence from a miocene Magnolia species. *Nature 344: 656–658*.

Günther T. and 18 co-authors. 2015. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proceedings of the National Academy of Sciences USA 112: 11917–11922*.

Haak W. and 17 co-authors. 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biology 8: e1000536*.

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

Haak W. and 38 co-authors. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature 522: 207–211*.

Hagelberg E., Hofreiter M. and Keyser C. 2015. Introduction. Ancient DNA: the first three decades. *Philosophical Transactions of the Royal Society B: Biological Sciences 370: 20130371.*

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

Higuchi R., Bowman B., Freiberger M., Ryder O. A. and Wilson A. C. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature 312: 282–284*.

Hofmanová Z. and 38 co-authors. 2016. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences USA 113: 6886–6891*.

Hofreiter M., Serre D., Poinar H. N., Kuch M. and Pääbo S. 2001. Ancient DNA. *Nature Review Genetics 2: 353–359*.

Jobling M., Tyler-Smith C., Hollox E., Hurles M. and Kivisild T. 2013. *Human Evolutionary Genetics*. Second Edition. Garland Science. New York and London.

Jones E. R. and 22 co-authors. 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature Communications 6: 8912*.

Kahle D., Wickham H. 2013. ggmap: Spatial Visualization with ggplot2. *The R Journal 5(1): 144–161*.

Kamberov Y. G. and 26 co-authors 2013. Modeling Recent Human Evolution in Mice by Expression of a Selected EDAR Variant. *Cell 152: 691–702*.

Keller A. and 40 co-authors. 2012. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature Communications 3: 698*.

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

Lazaridis I. and 51 co-authors. 2016. Genomic insights into the origin of farming in the ancient Near East. *Nature 536: 419–424.*

Lazaridis I. and 97 co-authors *et al.* 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature 513: 409–413*.

Leonardi M., Librado P., Der Sarkissian C., Schubert M., Alfarhan A. H., Alquraishi S. A., Al-Rasheid K. A. S., Gamba C., Willerslev E. and Orlando L. 2016. Evolutionary Patterns and Processes: Lessons from Ancient DNA. *Systematyc Biology doi: 10.1093/sysbio/syw059*

Llamas B., Holland M. L., Chen K., Cropley J. E., Cooper A. and Suter C. M. 2012. High-resolution analysis of cytosine methylation in ancient DNA. *PLoS One 7: e30226*.

Malmström H., Gilbert M. T. P., Thomas M. G., Brandström M., Storå J., Molnar P., Andersen P. K. , Bendixen C., Holmlund G., Götherström A. and Willerslev E. 2009. Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. *Current Biology 19: 1758–1762*.

Mathieson I. and 36 co-authors. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature 528: 499–503*.

Olalde I. and 23 co-authors. 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature 507: 225–228*.

Olalde I. and 19 co-authors. 2015. A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. *Molecular Biology and Evolution 32: 3132–3142*.

Omrak A., Günther T., Valdiosera C., Svensson E. M., Malmström H., Kiesewetter H., Aylward W., Storå J., Jakobsson M. and Götherström A. 2016. Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Current Biology 26: 270–275*.

Pääbo S., Poinar H., Serre D., Jaenicke-Despres V., Hebler J., Rohland N., Kuch M., Krause J., Vigilant L. and Hofreiter M. 2004. Genetic analyses from ancient DNA. *Annual Review of Genetics 38: 645–679*.

Pedersen J. S. and 20 co-authors. 2014. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Research 24: 454–466*.

Pinhasi R. and 18 co-authors. 2015. Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PLoS One 10: e0129102*.

Pinhasi R., Stock J. T. 2011. *Human Bioarchaeology of the Transition to Agriculture*. Wiley&Sons Ltd. Chichester.

Poinar H. N., Cano R. J. and Poinar G. O. 1993. DNA from an extinct plant. *Nature 363: 677–677*.

Poinar H. N., Hofreiter M., Spaulding W. G., Martin P. S., Stankiewicz B. A., Bland H., Evershed R. P., Possnert G. and Pääbo S. 1998. Molecular coproscopy: dung and diet of the extinct ground sloth Nothrotheriops shastensis. *Science 281: 402–406*.

Rasmussen M. and 51 co-authors. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature 463: 757–762*.

Rasmussen S. and 30 co-authors. 2015. Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago. *Cell 163: 571–582*.

Robin E. D., Wong R. 1988. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *Journal of Cellular Physiology 136: 507–513*.

Rohland N., Harney E., Mallick S., Nordenfelt S. and Reich D. 2015. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences 370: 20130624*.

Sankararaman S., Mallick S., Dannemann M., Prüfer K., Kelso J., Pääbo S., Patterson N. and Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature 507: 354–357*.

Shapiro B., Hofreiter M. 2012. *Ancient DNA: Methods and Protocols*. Methods in Molecular Biology 840. Humana Press. New York.

Skoglund P. and 14 co-authors. 2014. 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. and Jakobsson M. 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science 336: 466–469*.

Szécsényi-Nagy A. and 27 co-authors. 2015. Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. *Proceedings of the Royal Society B: Biological Sciences 282: 1805*.

Warinner C. and 31 co-authors. 2014. Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics 46: 336–344*.

Wickham H. 2010. *ggplot2: Elegant Graphics for Data Analysis.Use R!* Corr. 3rd printing 2010 edition. Springer-Verlag. New York.

Willerslev E., Hansen A. J., Binladen J., Brand T. B., Gilbert M. T. P., Shapiro B., Bunce M., Wiuf C., Gilichinsky D. A. and Cooper A. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science 300: 791–795*.

Woodward S. R., Weyand N. J. and Bunnell M. 1994. DNA sequence from Cretaceous period bone fragments. *Science 266: 1229–1232*.

Yousten A. A., Rippere K. E. 1997. DNA similarity analysis of a putative ancient bacterial isolate obtained from amber. *FEMS Microbiology Letters 152: 345–347*.

Zischler H., Höss M., Handt O., von Haeseler A., van der Kuyl A. C. and Goudsmit J. 1995. Detecting dinosaur DNA. *Science 268: 1192–1193*.

Zvelebil M. 2001. The agricultural transition and the origins of Neolithic society in Europe. *Documenta Praehistorica 28: 1–29*.