Mitigating destructive sampling to study population history

By Marc Vander Linden
Department of Archaeology, University of Cambridge

Nearly a decade has passed since the first sequencing of a human ancient genome and it feels like ancient DNA has been continuously in the spotlight of high-ranked journals. Its practitioners can boast of impressive scientific results, most broadly bringing to light the key yet complex role of admixture in human societies and evolution. Admixture is especially evident in regional time transects, as published in this week's issue by Olalde and colleagues for the Iberian Peninsula (1). Their new sampling and analysis retrace the many events that shaped regional long-term genomic variation including a more complex population structure of the last foragers than previously thought; and a possible total replacement of the local male population by people with Steppe ancestry during the Bronze Age, a pattern comparable to the situation for Bronze Age Britain (2). Olalde and colleagues also document for the first time episodes of gene flow from North Africa and the eastern Mediterranean during both Classical and Medieval periods, thus echoing known historical processes. Other identified events are perhaps less expected, as the presence of a northern African genomic component during the Chalcolithic period, independently identified by another team (3).

The archaeological community has had a mixed reaction to aDNA research. Some are delighted by its revolutionary prospects as aDNA allows to identify directly processes otherwise only inferred, especially migrations (4). Others, more cautious, have pointed out that biological relatedness cannot be conflated with social or cultural identities (5). Many archaeologists have questioned limited aDNA sample numbers, unevenly distributed across time and space. Whilst geneticists stress that each sample provides a wealth of information about entire past lineages, recent research on the European Bell Beaker Phenomenon demonstrates how extensive sampling is required to encapsulate in genomic terms the complexity of the past evident in archaeological data (2).

Yet, it is naive to assume that more data is always necessary. aDNA involves the destructive sampling of a finite resource and it is therefore imperative to evaluate the cost to irreplaceable resources and the benefits of knowledge gained. Given the self-acknowledged backlog of samples already processed in certain laboratories, cynics might be tempted to call for a temporary embargo on destructive analyses until we have learned how or whether the samples in the queue have changed the state of knowledge, whether or not new questions have arisen and what new data might be helpful.

As with any destructive technique, and given the cost resources involved in aDNA sequencing, simultaneous sampling for additional information should be systematically undertaken.

Radiocarbon dating is often done, though unfortunate exceptions occur (1). Beyond the savings of the resource, there are clear analytical advantages of such iterative sampling. By combining aDNA with stable isotopes and material culture studies, Knipper and colleagues showed, for the Early Bronze Age of the Lech valley (southern Germany), that the majority of women were non local, contrary to men who were mostly local, as well as a diversification of maternal lineages over time. These multiple patterns point to patrilocality and female exogamy rules over several generations (6). In a similar vein, Amorim and colleagues show how two early Medieval cemeteries from Hungary and northern Italy were organized around biological kinship, genetic differences being echoed in the deposition of grave goods, diet, and mobility patterns, thus suggesting complex family and social systems (7).

It is worth reminding that the range of sequencing techniques (targeted SNP, whole genome capture and shotgun) differ in terms of required endogamous DNA preversation, costs, and scientific outputs (8). Likewise, different human tissues provide different DNA yields: sampling of the petrous bone has widened the geographical range of aDNA studies, but is of limited use for pathogens (9). These are not mere technical details of interest to geneticists only, but key information that must be explained to museum curators, archeologists and local stakeholders prior to sample submission. It is the ethical responsibility of all parties involved to assure that full potential of all samples is tapped, so that best practice becomes the normal practice.

Multidisciplinary collaboration also offers stimuli to push the theoretical and methodological limits of each partner. Kinship is of long-standing interest to archaeologists, and proves methodologically stimulating for geneticists (10). Such mutual methodological opportunity is perhaps less obvious for population history. A pragmatic approach, advocated by practitioners of both disciplines (4, 11), is to let the geneticists sequence and analyse, and then the archaeologists to provide the in-depth explanations. Whilst apparently playing the strengths of each community, such post-hoc collaboration is at best statisfactory, at worst an opportunity lost. An alternative road lies in explicit hypothesis-building. In a recent paper, Mondal and colleagues first built several demographic computational simulations, from which they inferred the associated genetic signatures. Using a combination of statistical framework and deep learning algorithms, they then searched for matching signals in existing genomes, eventually suggesting the presence of a yet unidentified Pleistocene human species (12). Such work provides a model for future collaborative venture whereby archaeological information provides essential prior knowledge to be tested with genomics. For instance, admixture events sometimes seem dissociated from any demographic context, whilst a growing body of archaeological methods and data actually points to complex patterns of fluctuations in population size and distribution (Figure 1). It has been suggested that this variation

echoes aDNA signals for the spread of the Neolithic in Europe (13), though this hypothesis requires further modelling and testing.

Like any revolution, ancient DNA's legacy will not only been measured in light of technological developments, but by its ability to generate meaningful results beyond the repeated demonstration of admixture, which can only be achieved by forging effective collaborations. There have been tensions, but the opportunity is there for a transdisciplinary approach to population history and a better understanding of how gene flow and variation is shaped by human behaviour at multiple, congruent or not, scales.

Figure caption: Comparing archaeological and aDNA records through space and time

The geographical distributions of aDNA samples and 14C dates are roughly matching for later prehistoric Iberian peninsula (8000-2000 cal BC), pointing to an improved aDNA record. Summed probability distributions of the 14C dates, used as a demographic proxy, point to the complexity of population history for this period (red and blue bands indicating positive and negative deviations respectively from the overall Iberian signal). As aDNA samples unevenly overlap with these fluctuations, it is difficult to assess to what extent the latter shaped, or not, gene flow. Sources for aDNA data: (1); 14C data (14)

References

(1) I. Olalde *et al.*, *Science* (2019)

(2) I. Olalde *et al.*, *Nature* **555,** 190 (2018)

(3) G. González-Fortes *et al*., *Proc.R.Soc.B* **286**, 20182288 (2018).

(4) K. Kristiansen, *Current Swedish Archaeology* **22** (2014)

(5) N. Johanssen *et al*., *Science* **356**, 1119 (2017)

(6) C. Knipper *et al*., *Proc.Natl.Acad.Sci*. (2017) doi: 10.1073/pnas.1706355114

(7) C. Amorim *et al*., *Nat.Comms* **9**, 3547 (2018)

(8) Cappellini *et al*., *Annu. Rev. Biochem*. **87**, 36.1 (2018)

(9) A. Margaryan *et al*., *Ecol. Evol*. **8**, 3534 (2018)

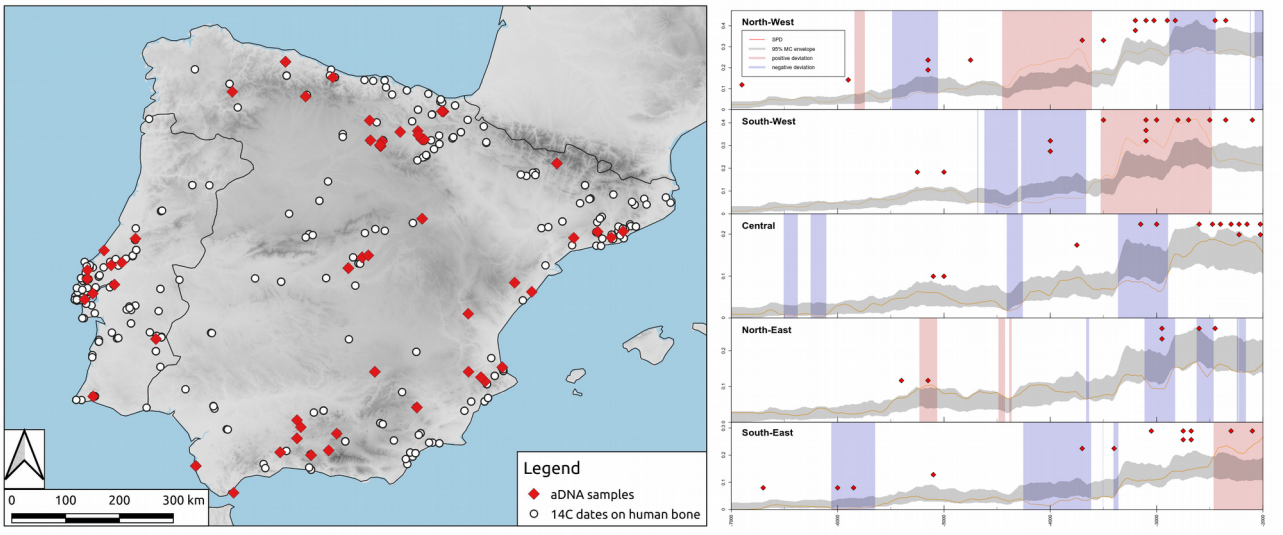
(10) J. Monroy Kuhn *et al*., *PlosOne* **13**, e195491 (2018)

(11) D. Reich, Who we are and how we got here (Pantheon, New-York) (2018)

(12) M. Mondal *et al*., *Nat.Comms.* **10**, 246 (2019)

(13) F. Silva, M. Vander Linden, *Sci.Rep*. **7**,11985 (2017)

(14) V. Balsera *et al*., *Quat.Int*. **386**, 208 (2015)


0
100
200
300 km
Legend
aDNA samples
14C dates on human bone
North-West
SPD
95% MC envelope
positive deviation
negative deviation
South-West
Central
North-East
South-East