

# Recent advances in ancient DNA research and their implications for archaeobotany

Terence A. Brown · Enrico Cappellini ·  
Logan Kistler · Diane L. Lister · Hugo R. Oliveira ·  
Nathan Wales · Angela Schlumbaum

Received: 6 March 2014 / Accepted: 29 September 2014 / Published online: 7 October 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** The scope and ambition of biomolecular archaeology is undergoing rapid change due to the development of new ‘next generation’ sequencing (NGS) methods for analysis of ancient DNA in archaeological specimens. These methods have not yet been applied extensively to archaeobotanical material but their utility has been demonstrated with desiccated, waterlogged and charred remains. The future use of NGS is likely to open up new areas of investigation that have been difficult or impossible with the traditional approach to aDNA sequencing. Species identification should become more routine with archaeobotanical explants, not just with

charred grain but with most if not all species likely to be encountered in an archaeobotanical setting. Distinctions between different subspecies groups such as cereal landraces will also be possible in the near future. Phenotypic characterization, in which aDNA sequencing is used to infer the biological characteristics of an archaeological specimen, will become possible, improving our understanding of traits such as flowering behaviour of cereals, and when combined with studies of preserved RNA and protein will enable complex phenotypes such as environmental tolerance and nutritional quality to be assessed. The sequencing of entire ancient plant genomes is also likely to have significant impact. As with past studies of ancient plant DNA, realization of the new potential provided by NGS will require productive collaboration between archaeologists and geneticists within the archaeobotanical research community.

---

Communicated by S. M. Valamoti.

---

T. A. Brown  
Faculty of Life Sciences, Manchester Institute of Biotechnology,  
University of Manchester, Manchester M1 7DN, UK

E. Cappellini · N. Wales  
Centre for GeoGenetics, University of Copenhagen,  
1350 Copenhagen K, Denmark

L. Kistler  
Department of Anthropology, Pennsylvania State University,  
University Park, PA 16802, USA

D. L. Lister  
McDonald Institute for Archaeological Research, University of  
Cambridge, Cambridge CB2 3ER, UK

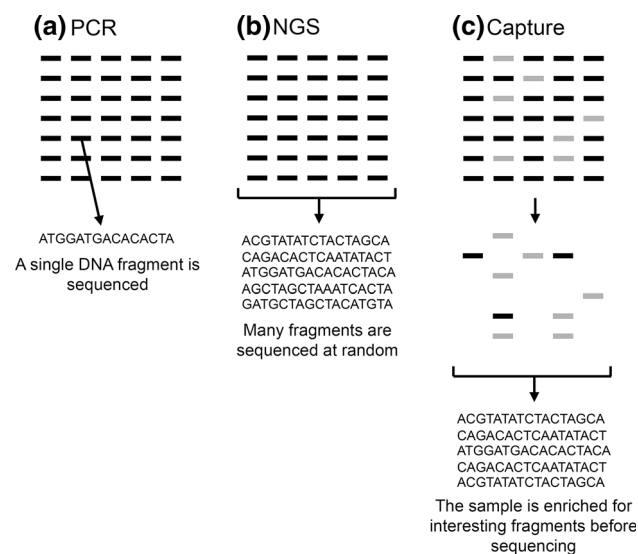
H. R. Oliveira  
Research Centre in Biodiversity and Genetic Resources, Campus  
Agrário de Vairão, 4485–661 Vairão, Portugal

A. Schlumbaum (✉)  
Integrative Prähistorische und Naturwissenschaftliche  
Archäologie, Spalenring 145, 4055 Basel, Switzerland  
e-mail: angela.schlumbaum@unibas.ch

**Keywords** Ancient DNA · Charred plant remains ·  
Desiccated plant remains · Genomes · Next generation  
DNA sequencing · Phenotype characterization · Species  
identification · Waterlogged plant remains

## Introduction

Technical advances in the ways in which nucleic acids are studied have led to dramatic developments in many areas of biomolecular archaeology. Most notable has been the application of ‘next generation’ sequencing (NGS) methods to ancient DNA (aDNA) from human remains, which has enabled complete genome sequences to be obtained for two types of extinct hominin and various archaic and prehistoric examples of *Homo sapiens* (Shapiro and Hofreiter 2014). These methods are also being applied to



**Fig. 1** The differences between sequencing aDNA by **a** PCR, **b** NGS, and **c** NGS after capture of interesting fragments

preserved plant macrofossils, and promise to have an equally dramatic impact in archaeobotany. The purpose of this article is to explore the likely extent of this impact.

#### Next generation sequencing of aDNA

Until to the mid-2000s, virtually all studies of aDNA were dependent on the polymerase chain reaction (PCR), which was used to amplify short regions of DNA, containing genes or other sequences of interest, from the small amounts of fragmentary aDNA present in an archaeological specimen (Brown and Brown 2011). This approach had three important limitations. First, it was necessary to design a different PCR for each sequence being studied (Fig. 1a), with each PCR using up a significant amount of the available aDNA, restricting the information that could be obtained from a sample. Second, to design a PCR at least part of the sequence of the target DNA region must already be known, meaning that aDNA studies could only be directed at genes that had previously been studied in living organisms. Third, fragments of aDNA shorter than about 80 bp are difficult to study by PCR, whereas most aDNA fragments, especially in charred plant remains, are less than 60 bp in length (Brown 1999). The new NGS approach sidesteps these limitations by providing sequences for a random sample of all the DNA molecules in an extract (Fig. 1b), regardless of their length and regardless of how much is known about those sequences in living plants.

The output of an NGS experiment is millions of individual sequence ‘reads’, the lengths of these reads giving insights about the lengths of the aDNA fragments in the sample being studied. Bioinformatics is then used to identify overlaps between these reads, to build up the

contiguous sequence of the original DNA molecule. This is a massive challenge when the genome of a species is being sequenced for the first time, but is less daunting if the genome sequence is already known and can be used as a reference to direct assembly of the reads. This is the strategy that enables a personal human genome, which comprises 3.2 billion nucleotides, to be sequenced for as little as \$5,000, and it can equally be applied to aDNA to sequence the genome of an archaeological specimen (Shapiro and Hofreiter 2014). But for many archaeological applications such extensive sequencing is overkill as the information being sought about the specimen can be obtained by studying just a few genes rather than the entire genome. In these cases, a pretreatment can be carried out to ‘capture’ those aDNA fragments that map to the regions of interest, so that the sequences of these fragments are enriched in the collection of reads provided by NGS (Fig. 1c) (Shapiro and Hofreiter 2010).

Although NGS methods are technically more difficult than the old-fashioned PCR approach to aDNA sequencing, they are becoming routine in many labs, and the cost of this type of work is decreasing annually. The bioinformatics skills needed to handle and analyse the vast number of sequence reads are also becoming routine thanks to the development of online and commercial computer packages that perform most of the standard tasks. The primary challenge with these new methods is the same as with the PCR approach—ensuring that the sequences that are obtained derive from aDNA and not modern contaminating DNA. A particular danger is that the techniques used to capture aDNA fragments spanning DNA regions of interest will preferentially capture undamaged modern DNA as opposed to partially degraded aDNA fragments. Various methods have been proposed for distinguishing genuine aDNA sequence reads from modern contaminating ones, based on the patterns of DNA damage displayed by aDNA (Skoglund et al. 2014), but such methods may never be entirely reliable (Green et al. 2009). For this reason, the standard precautions against contamination that have been *de rigueur* for aDNA research for the last decade, centering on the use of specialized laboratories in which samples are handled under ultraclean conditions (Brown and Brown 2011), must still be followed when an NGS approach is being taken.

#### The potential of NGS with archaeobotanical material

Any discussion of biomolecular archaeobotany must make a distinction between the different modes of preservation for plant material in the archaeological record, as these different preservation types have enormous implications for the nature of the biomolecular studies that are feasible. The ancient biomolecules in desiccated plant remains often

display exceptional preservation, specimens of this type being among the most amenable of all archaeological material for aDNA sequencing, and good preservation has also been reported for anoxic material, generally from waterlogged contexts (Schlumbaum and Edwards 2013). Charred plant remains, conversely, are among the worst types of material for this work, but there have been successful extractions of ancient DNA from various types of charred material and there is clear potential for future studies using NGS. Mineralised remains have not received a great deal of attention from biomolecular archaeobotanists, the assumption being that this type of preservation is unlikely to be compatible with survival of DNA.

#### *Desiccated and waterlogged remains*

The exceptional preservation of ancient biomolecules in desiccated plant remains was first demonstrated in the mid-1990s with a combined lipid and nucleic acid study of 3,300 year-old *Raphanus sativus* (radish) seeds from Qasr Ibrim, Upper Egypt (O'Donoghue et al. 1996). This work established not only that DNA and fatty acids from the original seeds were preserved in the desiccated remains, but that RNA was also present (O'Donoghue et al. 1994). This confirmed earlier reports of RNA in 3,400 year-old *Lepidium* (cress) seeds (Rollo 1985), but was still surprising because RNA is thought to be more labile than DNA and hence much less likely to survive in any type of archaeological material.

In view of this early demonstration of the suitability of desiccated plant remains for biomolecular study, it is perhaps surprising that during the intervening years there have been relatively few projects attempting to exploit this potential (Palmer et al. 2012b). This is a reflection of the scarcity of desiccated remains in the archaeological record. Most of the work has focussed on *Zea mays* (maize), making opportunistic use of desiccated maize cobs and kernels from caves and rockshelters in various parts in the New World (Freitas et al. 2003; Jaenicke-Després et al. 2003), but other species from these sites, such as chenopods, are now also beginning to attract attention (Kistler and Shapiro 2011). Other projects have made use of the remarkable range of desiccated material at Qasr Ibrim, early work on *Sorghum* being followed up more recently by studies of the Qasr Ibrim *Hordeum* (barley), which revealed unusual genetic features probably reflecting adaptation to the local environment (Palmer et al. 2009). As well as these genuinely archaeological specimens, important information on recent crop evolution has been inferred from studies of herbaria samples of *Hordeum* and *Pisum* (pea) (e.g. Leino et al. 2013; Lister et al. 2013) and from *Triticum* (wheat) chaff recovered from the floors of historic buildings (Blatter et al. 2002).

The studies mentioned so far all used the traditional PCR approach to aDNA sequencing. NGS has been applied to desiccated maize, but to date this work has focussed on technical issues (Ávila-Arcos et al. 2011). In contrast, a project with *Gossypium* (cotton) has used NGS to make a genuine advance in our understanding of the relationship between Old and New World varieties, as well as shedding new light on the evolutionary processes underlying domestication of this species (Palmer et al. 2012a). A combination of capture and NGS has also been used to sequence a contiguous 86 kb region of the chloroplast genome from eight desiccated and one waterlogged *Lagenaria siceraria* (bottle gourd) specimens, from New World sites up to 10,000 years in age, comparisons with extant African gourds suggesting that the New World varieties arrived during the late Pleistocene by floating across the Atlantic (Kistler et al. 2014). Other projects have focussed on waterlogged grape pips and wood, to make comparisons between archaeological grape varieties and modern ones (Manen et al. 2003; Wales et al. 2014; Schlumbaum et al. unpublished). As well as plant genomes, some important pathogens have been sequenced from historic and archaeological plant remains, including the fungus *Phytophthora infestans*, causative agent of the Irish potato famine, obtained from potato leaves in herbarium collections (Martin et al. 2013; Yoshida et al. 2013), and a 750-year-old barley stripe mosaic virus from desiccated barley grain (Smith et al. 2014).

NGS has also been used to obtain sequences of components of the preserved transcriptomes of archaeological barley (Smith et al. 2014) and maize (Fordyce et al. 2013), the transcriptome being the collection of RNA molecules in a sample, which in turn identifies those genes that are active in a particular tissue at a particular time. There has also been preliminary success in directly investigating the protein content of *Vitis* (grape) seeds (Cappellini et al. 2010). We return to these approaches later as they point the way to one of the future developments of ancient biomolecular research.

#### *Charred archaeobotanical remains*

Most archaeobotanists would agree that aDNA sequences from charred remains represent the 'holy grail' for biomolecular research. Extensive work during the 1990s (reviewed by Palmer et al. 2012b) showed that aDNA is preserved in at least some assemblages of charred wheat, establishing that this type of preservation should not be referred to as 'carbonized', the latter implying that the burning process has resulted in complete conversion of the remains to inorganic carbon. Even so, the preservation of aDNA within a positive assemblage is sporadic, and the aDNA fragments have very short lengths. More recently,

the focus has moved away from cereal grains to larger and more robust types of charred material, within which aDNA might be more likely to survive. Charred maize cobs were used in one of the earliest plant aDNA projects (Goloubinoff and Pääbo 1993) and later in a comparison of landraces from Andean sites in southern South America (Lia et al. 2007), and charred grape seeds have yielded aDNA sequences that have helped unravel the evolutionary history of this crop (Manen et al. 2003).

NGS studies of charred remains are much less well advanced than those with desiccated samples, but the initial results are promising. A mixture of charred grain from the Bronze Age site of Assiros Toumba gave sequence reads that could be assigned to *Triticum*, *Hordeum* and, to a lesser extent, *Panicum miliaceum* (broomcorn millet), all of which were present in the assemblage (Bunning et al. 2012). A second project, with 2,000-year-old *Hordeum* from Kawa in the Sudan, used capture to direct sequencing at chosen parts of the *Hordeum* genome, the results indicating that it might be possible to obtain aDNA sequences spanning entire genes from charred grain (Allaby et al. 2014).

#### Future directions for biomolecular archaeobotany

Now we explore the major research opportunities made possible by the application of NGS to archaeological plant remains, opportunities that we look on as the key future directions for biomolecular archaeobotany.

#### *Taxonomic identification of seeds and other explants*

Several attempts have been made, with varying success, to use the PCR approach to aDNA sequencing for species identification with charred wheat remains that lack sufficient morphological traits for categorization by conventional methods (Brown 1999). This is a particular problem with the naked wheats, whose chaff is only infrequently preserved, as criteria such as the grain dimensions are extremely imprecise. One challenge is to distinguish between *Triticum turgidum* ssp. *durum* respectively *T. durum* (durum wheat) and *T. aestivum* (bread wheat), which is possible because the latter possesses the D genome, which is absent from *T. durum* (Schlumbaum et al. 2008). A related question concerns the ‘new glume wheat’, thought to be a tetraploid species but distinct from *T. dicoccum* (emmer) and *T. durum*, and possibly more closely related to *T. timopheevi* (Jones et al. 2000). This hypothesis predicts that the new glume wheat would possess a G genome. PCR has been used since the 1990s to sequence short regions of the wheat genome displaying genome-specific variations, such as the promoter regions of the glutenin genes and the spacer components of the ribosomal

DNA loci, in order to distinguish charred examples of *T. durum* and *T. aestivum* (e.g. Allaby et al. 1997; Schlumbaum et al. 1998; Allaby et al. 1999). The same approach is routinely used today with desiccated remains (e.g. Li et al. 2011; Oliveira et al. 2012), but has not been extensively pursued with charred specimens due to anomalous results arising from artefacts synthesized during PCR by the process called template switching. These artefacts contain segments of different aDNA molecules linked head to tail, so that the sequence that is obtained combines polymorphisms from different genomes. The chimeric sequences can be unravelled to reveal the authentic genome sequences (Allaby et al. 1999), but the analysis is difficult and is rarely carried out.

The difficulties in distinguishing genomes in charred wheat are entirely due to the limitations of the PCR approach to aDNA sequencing, template switching being much less likely to occur during sequencing by a NGS method. The literature contains many examples of genes and other regions of DNA that contain sequence features specific to the different wheat genomes, and which provide a much more secure means of identifying the genomes that are present in a charred specimen. These sequences are too long, both individually and in combination, to be accessible by PCR, as this would require many more PCRs than would be possible with the limited amounts of aDNA obtainable from any preserved specimen. They would, however, make ideal targets for enrichment, as the reads obtained after NGS would be enriched for ones mapping to these regions, enabling the diagnostic polymorphisms to be typed.

Although likely to have immediate impact in identifying enigmatic charred wheat specimens, the use of NGS for taxonomic identification can be extended to any species, because comparison of the sequence reads obtained from an unknown specimen with the DNA databases will reveal at least the genus. A little further in the future, but by no means unfeasible, would be the use of NGS to distinguish between subspecies groups such as cereal landraces.

#### *Phenotypic characterization of archaeobotanical remains*

The second future direction for biomolecular archaeobotany is to use NGS to sequence key genes that specify important biological characteristics of the living plant, in order to infer the phenotypic features of an archaeological specimen. There are many different types of question that could be answered in this way, but we will use the flowering behaviour of barley as an example with which to illustrate the approach. Flowering behaviour is an adaptive trait whose study can help to identify the selective pressures, both human and environmental, that drove the evolution of crops as they were taken into regions outside their natural

environment. Most wild barleys flower only after a period of prolonged cold exposure—vernalization—and in response to increasing daylength. These phenotypes are referred to as winter growth habit and daylength responsiveness, respectively. They ensure that wild barley does not begin to flower until after winter, but completes its reproductive cycle before the midsummer heat and drought of its natural habitats in southwest Asia. Some, though not all, cultivated barleys have lost the vernalization requirement, enabling them to be sown in spring, and some have become nonresponsive to daylength, so they flower later in the season, permitting a longer period of vegetative growth and increased resource storage. In a crop with a spring growth habit, the ability to extend the growing period later into the year in a climate with moist summers is of obvious selective advantage, and may have been a critical factor in enabling cereals to spread into the more northerly regions of Europe and into high altitude locations such as the Alpine foreland (Jones et al. 2012). The genetic basis to the vernalization and daylength responses of barley are now reasonably well understood. The vernalization requirement is controlled by three genes, and daylength responsiveness by two others, and the various phenotypes can be read from the DNA sequences of these genes. All that is necessary in order to infer the flowering behaviour of an archaeological barley sample is to obtain aDNA sequences of these genes. With NGS, using capture to direct sequencing at the genes of interest, phenotypic characterization of these genes, and of others specifying interesting biological characteristics, should become increasingly routine.

The limitation to this approach is unlikely to be our future ability to obtain aDNA sequences from archaeobotanical specimens. The limitation is more likely to be the complexity of the link between DNA sequence and phenotype, which often makes it difficult or impossible to make a clear interpretation of phenotype simply by studying a DNA sequence. This difficulty has three underlying causes. First, even if a phenotype is specified by a single gene or small number of genes, as with the vernalization and photoperiod responses of barley, the initial identification of those genes, and of the variants associated with different versions of a phenotype, presents a complex problem that requires high-powered genetic analysis. In general, this is not area in which biomolecular archaeobotany is able to take the lead, and the list of phenotypes that can be studied by aDNA sequencing is always likely to be those that crop breeders and plant geneticists have deemed important enough to justify the effort and resources needed for their initial characterization. The second difficulty in interpreting the link between DNA sequence and phenotype arises from the complexity of the genetic basis to many phenotypes, which are influenced not just by two or three genes, but by many genes, possibly hundreds. There are many examples of such

complex phenotypes, including interesting ones such as tolerance to waterlogging in barley and drought resistance in maize. Characterizing phenotypes such as these from DNA sequences is currently impossible even with extant plants, and we therefore cannot expect to distinguish them by NGS of aDNA.

The third complication in the link between gene and phenotype lies with the possibility that different phenotypes arise not from sequence differences in the relevant genes, but from changes in the activities of those genes. Geneticists are becoming increasingly aware of the importance of epigenetic variation, which refers to a heritable variation that is not caused by differences in DNA sequence but instead results from chemical modification of the genome, an example being the addition of methyl groups to cytosine nucleotides, which can suppress the activity of adjacent genes. The methylation pattern is retained during DNA replication, and so can be transmitted to the next generation of plants. Epigenetic variation has been linked to the responses of some plants to environmental challenges (Henderson and Jacobsen 2007). Special methods of DNA sequencing that can detect the location of cytosine methylations, enabling epigenetic variation to be identified, have been applied to aDNA from Pleistocene bison bones (Llamas et al. 2012) and desiccated barley remains (Smith et al. 2014). The recent breakthroughs in sequencing RNA and characterizing proteins in archaeobotanical remains (Cappellini et al. 2010; Smith et al. 2010; Fordyce et al. 2013) are also important in this context. Many phenotypes, including the nutritional value of cereal grain and cold adaptation in barley, are affected not just by the gene variants present in the genome of an individual plant, but also by the rate at which each gene directs the synthesis of RNA and protein in a particular tissue at a particular time. Characterizing the transcriptome (RNA content) and proteome (protein content) of a tissue or structure such as a mature seed is therefore just as important as sequencing the genes responsible for a phenotype. Doing this with archaeobotanical material still presents a major challenge. It is not sufficient simply to identify the RNA or protein molecules in a transcriptome or proteome. Their identification tells us which genes are active, but does not provide the critical information that is needed on the relative degrees of activity of those genes. Information is therefore needed on the relative amounts of individual RNAs or proteins in order to interpret the phenotype from the composition of a transcriptome or proteome. Quantification of the individual components of transcriptomes and proteomes is routine when living plants are being studied, and is equally routine with transcriptomes and proteomes obtained from archaeological specimens. The problem is that the vagaries of diagenesis might mean that the composition of the transcriptome or

**Table 1** Examples of plants for which complete or near-complete genome sequences are available

Species <sup>a</sup>	Common name	Genome size (Mb) <sup>b</sup>
<i>Beta vulgaris</i>	Sugar beet	730
<i>Cicer arietinum</i>	Chickpea	740
<i>Cucumis melo</i>	Melon	425
<i>Glycine max</i>	Soybean	1,115
<i>Gossypium raimondii</i>	Cotton D genome	761
<i>Hordeum vulgare</i>	Barley	5,100
<i>Linum usitatissimum</i>	Flax	350
<i>Manihot esculenta</i>	Cassava	760
<i>Oryza sativa</i>	Rice	370
<i>Phaseolus vulgaris</i>	Common bean	520
<i>Ricinus communis</i>	Castor bean	320
<i>Solanum tuberosum</i>	Potato	856
<i>Sorghum bicolor</i>	Sorghum	730
<i>Triticum aestivum</i>	Bread wheat	15,000
<i>Triticum urartu</i>	Wheat A <sup>u</sup> genome	4,940
<i>Vitis vinifera</i>	Grape	505
<i>Zea mays</i>	Maize	2,300

<sup>a</sup> For some species, especially those with larger genomes, sufficient sequence reads to cover the entire genome have been obtained, but these have not been completely assembled into contiguous DNA sequences, or only parts of the genome, such as the genes, have been assembled

<sup>b</sup> In comparison, the human genome is 3,500 Mb

proteome in a preserved seed is quite different to what it was when the seed reached ripeness several hundred or thousand years ago. If we assume that every RNA and every protein degrades at equal rate, then there is no problem, but we cannot make that assumption based on our current knowledge of the ways in which these biomolecules degrade over time, and hence we cannot as yet make secure interpretations of phenotype from RNA and protein extracted from archaeobotanical remains. Progress to meet this challenge does not therefore require further breakthroughs in the methodology used to study nucleic acids and proteins, but instead needs a greater refinement in our knowledge of the patterns of biomolecular breakdown during diagenesis of different types of preserved plant material.

#### *Entire genome sequences for archaeological varieties of plants*

In human biomolecular archaeology, the most spectacular use of the new aDNA sequencing methods has been in the construction of complete genome sequences from extinct hominins and archaic and prehistoric versions of *H. sapiens* (Shapiro and Hofreiter 2014). This work has revealed the evolutionary relationships and degree of cross-breeding

between *H. sapiens*, Neanderthals and Denisovans, as well as the physical and biochemical characteristics of prehistoric individuals. Can we hope to perform equivalent feats of genome reconstruction with archaeobotanical material?

The answer is a qualified yes. The main limitation in obtaining a complete genome sequence from aDNA is the need to have a reference sequence—essentially the complete genome sequence obtained from modern material—on which to assemble the millions of short sequence reads obtained by NGS. The genomes of several plant species have been sequenced, and the number is increasing rapidly (Table 1). However, some domesticates, including wheat and barley, have very large genomes, making sequence assembly from aDNA very challenging. Others have smaller genomes, but with some species, such as maize, the organization of the genome is complex, with a large portion made up of mixtures of repeated sequences, whose presence again complicates assembly of aDNA sequence reads.

Because of their huge importance in human archaeology, one might imagine that complete genome sequences of preserved plant specimens will have an equally major impact in archaeobotany. In reality, this is unlikely to be the case. The problem lies with the nature of the archaeobotanical questions that could be addressed from a complete sequence. Of course, a complete sequence contains all the information needed to distinguish individual genome types and to interpret phenotypes from individual genes or groups of genes. Hence, the challenges described above could be addressed by genome sequencing, but to do so would be immense overkill because the work and resources needed to obtain a complete genome are one or more orders of magnitude greater than those needed to sequence just the aDNA molecules covering those genes or other regions that are needed to identify a species and/or characterize its interesting phenotypes. Partial genome sequences from archaeobotanical specimens, obtained by capture methods directed at genes and other sequences of interest, are therefore equally as useful as complete genomes for addressing many research questions.

A second consideration is whether the sequence of an archaeological specimen would tell us much more than the sequence of a modern plant variety. With maize, for example, the morphological features of many archaeological specimens enable these to be associated with some confidence with extant landraces, and in these cases the limited genetic features of the archaeological variety, as determined by PCR, have proved to be identical with those of the modern version, except for polymorphisms that can be put down to intravarietal differences (Brown et al. unpublished). Even when the remains lack a distinctive phenotype, relationships with modern varieties can still be assessed by PCRs directed at short sequences, as has been demonstrated

by a comparison between desiccated wheat grains from pre-Hispanic sites on Gran Canaria with modern wheat varieties from the same island and from north Africa and Spain (Oliveira et al. 2012). Similar considerations apply to fruit trees and vines, which are vegetatively propagated, which means there may be only a few generations between a historic sample and the descendent modern variety.

The complete hominin genome sequences that have been obtained by aDNA sequencing, as well as the complete genomes of pathogenic bacteria such as the plague bacillus *Yersinia pestis* and *Mycobacterium leprae*, the causative agent of leprosy, have been of greatest value in studying broad questions in evolution and population affinity (Shapiro and Hofreiter 2014). Complete or near-complete genome sequences of archaeobotanical specimens are likely to be equally valuable in this context, as has already been shown by the pioneering work that has been carried out with desiccated cotton (Palmer et al. 2012a). This type of study is not limited to those species whose modern genome sequences are already known. An initial experiment in which NGS is used with DNA from an extant specimen will provide enough reads to assemble short parts of the modern genome, which might only cover a few percent of the entire sequence, but will be sufficient to design a capture system so NGS of the ancient specimens can be directed at the assembled regions. Comparative sequence data from ancient and modern varieties can therefore be obtained, enabling the evolutionary relationships to be assessed. We should bear in mind, however, that in archaeobotany the most interesting evolutionary questions tend to relate to the rather short time scale represented by the period since domestication, which for many crops is no more than a few thousand generations. For this relatively short period, evolutionary studies based on extant landraces can often provide fairly complete information on how a crop has changed (e.g. Jones et al. 2012, 2013). The questions that genome sequencing are expected to answer therefore need to be phrased carefully to ensure that they are not ones that could be answered more easily simply by sequencing genomes from living plants.

## Conclusions

NGS methods for aDNA sequencing are revolutionizing biomolecular archaeology and are likely to have a similar important impact in archaeobotany, as indicated by the groundbreaking projects that have already been carried out. Care must be taken to ensure that large scale sequencing projects address questions that have genuine importance in archaeobotany, but our discipline has an excellent past record of productive collaboration between archaeology

and genetics, and there is every reason to believe that this will continue to be the case in the future.

**Acknowledgments** This paper is dedicated to Franco Rollo, who passed away on 3rd September 2014. Franco was the first to identify preserved nucleic acids in plant remains. His work in the 1980s and early 1990s ensured that biomolecular archaeobotany developed into the rigorous and respected discipline that it has now become.

## References

- Allaby RG, O'Donoghue K, Sallares R, Jones MK, Brown TA (1997) Evidence for the survival of ancient DNA in charred wheat seeds from European archaeological sites. *Anc Biomol* 1:119–129
- Allaby RG, Banerjee M, Brown TA (1999) Evolution of the high-molecular-weight glutenin loci of the A, B, D and G genomes of wheat. *Genome* 42:296–307
- Allaby RG, Gutaker R, Clarke A et al (2014) Using archaeogenomic and computational approaches to unravel the history of local adaptation in crops. *Phil Trans R Soc B* 279:4,727–4,733
- Ávila-Arcos MC, Cappellini E, Romero-Navarro JA et al (2011) Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci Rep* 1:74
- Blatter R, Jacomet S, Schlumbaum A (2002) Spelt-specific alleles in HMW glutenin genes from modern and historical European spelt (*Triticum spelta* L.). *Theor Appl Genet* 104:329–337
- Brown TA (1999) How ancient DNA may help in understanding the origin and spread of agriculture. *Phil Trans R Soc B* 354:89–98
- Brown TA, Brown KA (2011) Biomolecular archaeology: an introduction. Wiley-Blackwell, New York
- Bunning SL, Jones G, Brown TA (2012) Next generation sequencing of DNA in 3300-year-old charred cereal grains. *J Archaeol Sci* 39:2,780–2,784
- Cappellini E, Gilbert MTP, Geuna F et al (2010) A multidisciplinary study of archaeological grape seeds. *Naturwissenschaften* 97:205–217
- Fordyce L, Ávila-Arcos MC, Rasmussen M et al (2013) Deep sequencing of RNA from ancient maize kernels. *PLoS ONE* 8(1):e50961
- Freitas FO, Bandel G, Allaby RG, Brown TA (2003) DNA from primitive maize landraces and archaeological remains: implications for the domestication of maize and its expansion into South America. *J Archaeol Sci* 30:901–908
- Goloubinoff P, Pääbo S (1993) Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proc Natl Acad Sci USA* 90:1,997–2,001
- Green RE, Briggs AW, Krause J et al (2009) The Neandertal genome and ancient DNA authenticity. *EMBO J* 28:2,494–2,502
- Henderson IR, Jacobsen SE (2007) Epigenetic inheritance in plants. *Nature* 447:418–424
- Jaenicke-Després V, Buckler ES, Smith BD et al (2003) Early allelic selection in maize as revealed by ancient DNA. *Science* 302:1,206–1,208
- Jones G, Valamoti S, Charles M (2000) Early crop diversity: a “new” glume wheat from northern Greece. *Veget Hist Archaeobot* 9:133–146
- Jones G, Jones H, Charles MP et al (2012) Phylogeographic analysis of barley DNA as evidence for the spread of Neolithic agriculture through Europe. *J Archaeol Sci* 39:3,230–3,238
- Jones G, Charles MP, Jones MK et al (2013) DNA evidence for multiple introductions of barley into Europe following dispersed domestications in Western Asia. *Antiquity* 87:701–713

- Kistler L, Shapiro B (2011) Ancient DNA confirms a local origin of domesticated chenopod in eastern North America. *J Archaeol Sci* 38:3,549–3,554
- Kistler L, Montenegro A, Smith BD et al (2014) Transoceanic drift and the domestication of African bottle gourds in the Americas. *Proc Natl Acad Sci USA* 118(8):2,937–2,941
- Leino M, Boström E, Hagenblad J (2013) Twentieth-century changes in the genetic composition of Swedish field pea metapopulations. *Heredity* 110:338–346
- Li CX, Lister DL, Li HJ et al (2011) Ancient DNA analysis of desiccated wheat grains excavated from a Bronze Age cemetery in Xinjiang. *J Archaeol Sci* 38:115–119
- Lia VV, Confalonieri VA, Ratto N et al (2007) Microsatellite markers provide insights into the genetic constitution of ancient maize in southern South America. *Proc R Soc Lond B* 274:545–554
- Lister DL, Jones H, Jones MK, O'Sullivan DM, Cockram J (2013) Analysis of DNA polymorphism in ancient barley herbarium material: validation of the KASP SNP genotyping platform. *Taxon* 62:779–789
- Llamas B, Holland ML, Chen K et al (2012) High-resolution analysis of cytosine methylation in ancient DNA. *PLoS ONE* 7(1):e30226
- Manen JF, Bouby L, Dalnoki O et al (2003) Microsatellites from archaeological *Vitis vinifera* seeds allow a tentative assignment of the geographical origin of ancient cultivars. *J Archaeol Sci* 30:721–729
- Martin MD, Cappellini E, Samaniego JA (2013) Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nat Comm* 4(2):172
- O'Donoghue K, Brown TA, Carter JF, Evershed RP (1994) Detection of nucleotide bases in ancient seeds using gas chromatography/mass spectrometry and gas chromatography/mass spectrometry/mass spectrometry. *Rapid Comm Mass Spectrom* 8:503–508
- O'Donoghue K, Clapham A, Evershed RP, Brown TA (1996) Remarkable preservation of biomolecules in ancient radish seeds. *Proc R Soc Lond B* 263:541–547
- Oliveira HR, Civán P, Morales J et al (2012) Ancient DNA in archaeological wheat grains: preservation conditions and the study of pre-Hispanic agriculture on the island of Gran Canaria (Spain). *J Archaeol Sci* 39:828–835
- Palmer SA, Moore JD, Clapham AJ, Rose P, Allaby RG (2009) Archaeogenetic evidence of ancient Nubian barley evolution from six to two-row indicates local adaptation. *PLoS ONE* 4(7):e6301
- Palmer SA, Clapham AJ, Rose P et al (2012a) Archaeogenomic evidence of punctuated genome evolution in *Gossypium*. *Mol Biol Evol* 29:2,031–2,038
- Palmer SA, Smith O, Allaby RG (2012b) The blossoming of plant archaeogenetics. *Ann Anat* 194:146–156
- Rollo F (1985) Characterisation by molecular hybridization of RNA fragments isolated from ancient (1400 BC) seeds. *Theor Appl Genet* 71:330–333
- Schlumbaum A, Edwards CJ (2013) Ancient DNA research on wetland archaeological evidence. In: Menotti F, O'Sullivan A (eds) *The Oxford handbook of wetland archaeology*. Oxford University Press, Oxford, pp 569–583
- Schlumbaum A, Neuhaus J-M, Jacomet S (1998) Coexistence of tetraploid and hexaploid naked wheat in a Neolithic lake dwelling of central Europe: evidence from morphology and ancient DNA. *J Archaeol Sci* 25:1,111–1,118
- Schlumbaum A, Tensen M, Jaenicke-Després V (2008) Ancient plant DNA in archaeobotany. *Veget Hist Archaeobot* 17:233–244
- Shapiro B, Hofreiter M (2010) Analysis of ancient human genomes. *BioEssays* 32:388–391
- Shapiro B, Hofreiter M (2014) A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. *Science* 343:1,236,573
- Skoglund P, Northoft BH, Shunkov MV et al (2014) Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci USA*. doi:10.1073/pnas.1414542111
- Smith O, Manning K, Clapham AJ et al. (2010) Detection of epigenesis in archaeogenetic systems. In: Abstracts, fourth international symposium on biomolecular archaeology, p 104
- Smith O, Clapham A, Rose P et al (2014) A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological barley stripe mosaic virus. *Sci Rep* 4:5,559
- Wales N, Anderson K, Cappellini E et al (2014) Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PLoS ONE* 9(1):e86827
- Yoshida K, Schuenemann VJ, Cano LM et al. (2013) The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLIFE* 2:e00731