New Phytologist

Review

Research review

Ancient plant DNA: review and prospects

Author for correspondence:
Felix Gugerli
Tel: +41 1 7392590
Fax: +41 1 7392215
Email: gugerli@wsl.ch

Received: *25 October 2004*
Accepted: *14 December 2004*

Felix Gugerli[1], Laura Parducci[2] and Rémy J. Petit[3]

[1]Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH–8903 Birmensdorf, Switzerland; [2]Palaeobiology programme, Department of Earth Sciences, Department of Evolutionary Biology, Uppsala University, Norbyvägen 18/D, SE–752 36 Uppsala, Sweden; [3]Institut National de la Recherche Agronomique, UMR Biodiversité, Gènes et Ecosystèmes, 69, Route d'Arcachon, F–33612 Cestas Cedex, France

Key words: ancient DNA, fossil, intraspecific genetic variation, molecular markers, palaeogenetics, population history.

Summary

Ancient DNA has received much attention since the mid-1980s, when the first sequence of an extinct animal species was recovered from a museum specimen. Since then, the majority of ancient DNA studies have focused predominantly on animal species, while studies in plant palaeogenetics have been rather limited, with the notable exception of cultivated species found in archaeological sites. Here, we outline the recent developments in the analysis of plant ancient DNA. We emphasize the trend from species identification to population-level investigation and highlight the potential and the difficulties in this field, related to DNA preservation and to risks of contamination. Further efforts towards the analysis of ancient DNA from the abundant store of fossil plant remains should provide new research opportunities in palaeoecology and phylogeography. In particular, intraspecific variation should be considered not only in cultivated plants but also in wild taxa if palaeogenetics is to become a fully emancipated field of plant research.

New Phytologist (2005) **166**: 409–418

© *New Phytologist* (2005) **doi**: 10.1111/j.1469-8137.2005.01360.x

Introduction

Population genetic studies based on molecular markers indicate how Quaternary climate oscillations have influenced the extant genetic structure of plants (Lascoux *et al.*, 2004). Similarly, the establishment of large-scale pollen and macrofossil databanks and the comparative analyses of genetic and fossil data (Brewer *et al.*, 2002; Taberlet & Cheddadi, 2002; Petit *et al.*, 2003) have considerably improved our knowledge of the distribution of plants since the last glaciation and has led to a better understanding of their ecological setting (Bennett, 1997; Elenga *et al.*, 2000; Litt *et al.*, 2003). A more direct and largely complementary approach to study past population processes would be provided by palaeogenetics (i.e. the study of ancient DNA (aDNA)). Ancient DNA research deals with DNA molecules that have been subjected to post mortem degradation primarily through hydrolysis and oxidation. In plants, these DNA molecules can be found in tissues of extant organisms, such as wood or sclerenchymatic tissue of fruits or seeds, in herbarium specimens, in various types of fossilized plant remains, in feces as secondary components or even adsorbed to sediments after leakage from their original cellular source.

In principle, insights into the evolutionary history of plant populations can be obtained by (1) relying on direct evidence from the fossil record, (2) studying the current genetic structure of living organisms, (3) genetically analysing germinated seeds or spores buried in the soil, or (4) directly identifying species and genotypes through the analysis of aDNA (palaeogenetics). Plant palaeontological research is often limited by taxonomic resolution because fossil remains, such as wood or pollen, can

rarely be identified beyond the species (or even genus) level, often impeding inferences on infraspecific processes. On the other hand, extinct lineages or genotypes cannot be detected when relying exclusively on samples taken from living specimens, precluding a direct spatiotemporal analysis. An elegant solution to this problem has been provided by analysing soil diaspore banks, which represent living records from past populations. Successful studies based on these 'time capsules' include the circumarctic cotton grass *Eriophorum vaginatum* (McGraw, 1993), the local endemic ground plum *Astragalus bibullatus* (Morris *et al.*, 2002) or the temperate fern *Athyrium filix-femina* (Schneller, 1998). Even though one viable seed recovered from a permafrost deposit was claimed to be at least 10 000 yr old (Porsild *et al.*, 1967), a *c.* 1300-yr-old lotus fruit from China seems to be the oldest germinated and directly ^{14}C-dated fruit known (Shen-Miller, 2002). Unfortunately, however, such living archives are limited to the few taxa that produce persistent diaspores and are generally not older than a few centuries. Therefore, palaeogenetics represents the most direct approach to unravel the evolutionary history of plant populations.

Here, we summarize the recent development achieved in plant aDNA research, examine conditions necessary for the preservation of DNA in plant tissue and indicate promising fields of research. We deliberately refrain from thoroughly dealing with the issue of authenticity, as it has been extensively elaborated elsewhere (Austin *et al.*, 1997; Yang, 1997; Wayne *et al.*, 1999; Cooper & Poinar, 2000; Hummel, 2003; Pääbo *et al.*, 2004). One of our aims is to show that intraspecific aDNA studies in plants could provide a link between palaeoecology, population genetics and phylogeography, which should improve our understanding of past population processes in plants.

Publication record of aDNA

A literature search on ISI Web of Science carried out in October 2004, using 'ancient DNA' as search term, retrieved 537 entries. This is by no means exhaustive but it should be representative of the literature available in the field. Forty-six of these were excluded as being erroneously listed or just replies to other articles. Well-known publications on Miocene plant sequences, on the first archaeo-anthropological DNA study, and on DNA extraction from the extinct quagga were not retrieved using this key word combination and therefore were manually added, giving a total of 496 papers. There was a continuous, exponential increase in the number of publications over time. The diversity of organisms investigated was relatively low, with the bulk of the efforts focusing on humans (35%), other vertebrate taxa (28%) and microorganisms (9%; Fig. 1). Only 35 studies (7%) dealt with plant species, which is less than the 55 review or opinion articles. It

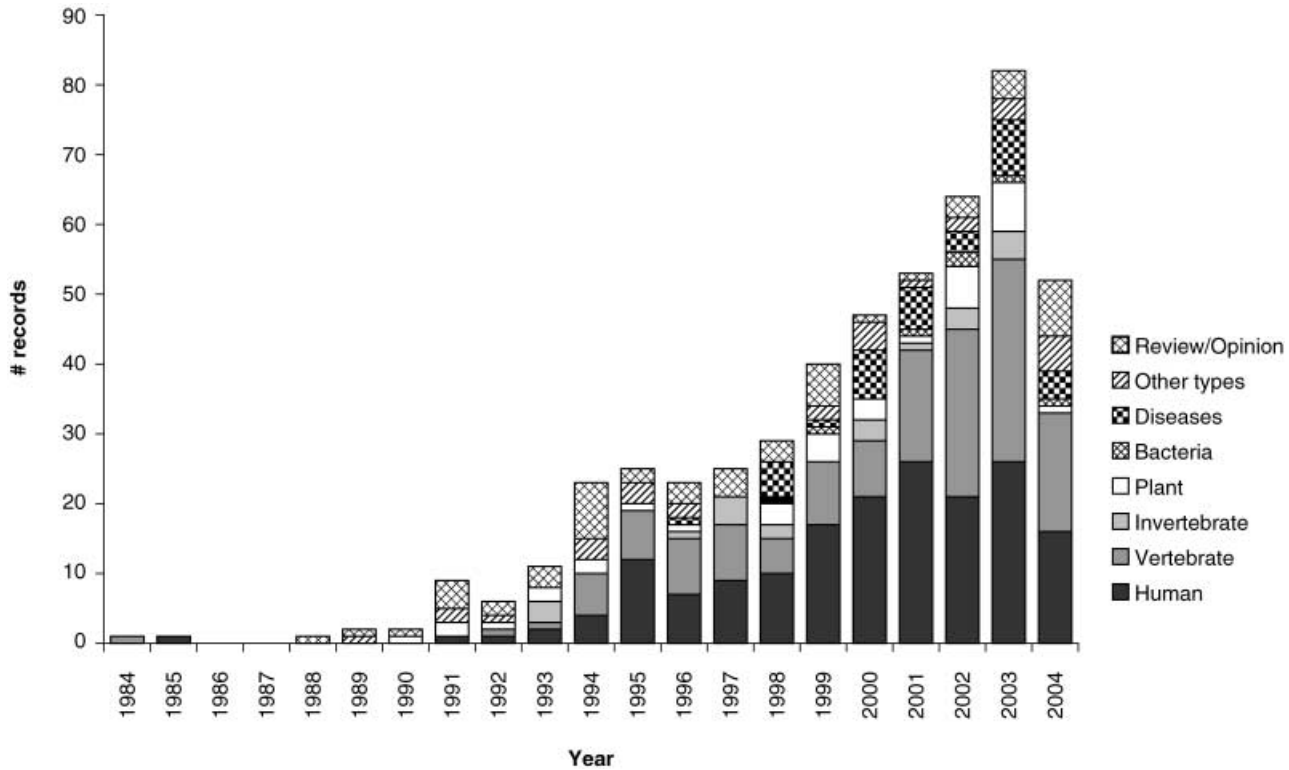

Fig. 1 Number of publications retrieved from a public reference data bank in October 2004, using 'ancient DNA' as search term, categorized according to the study organisms.

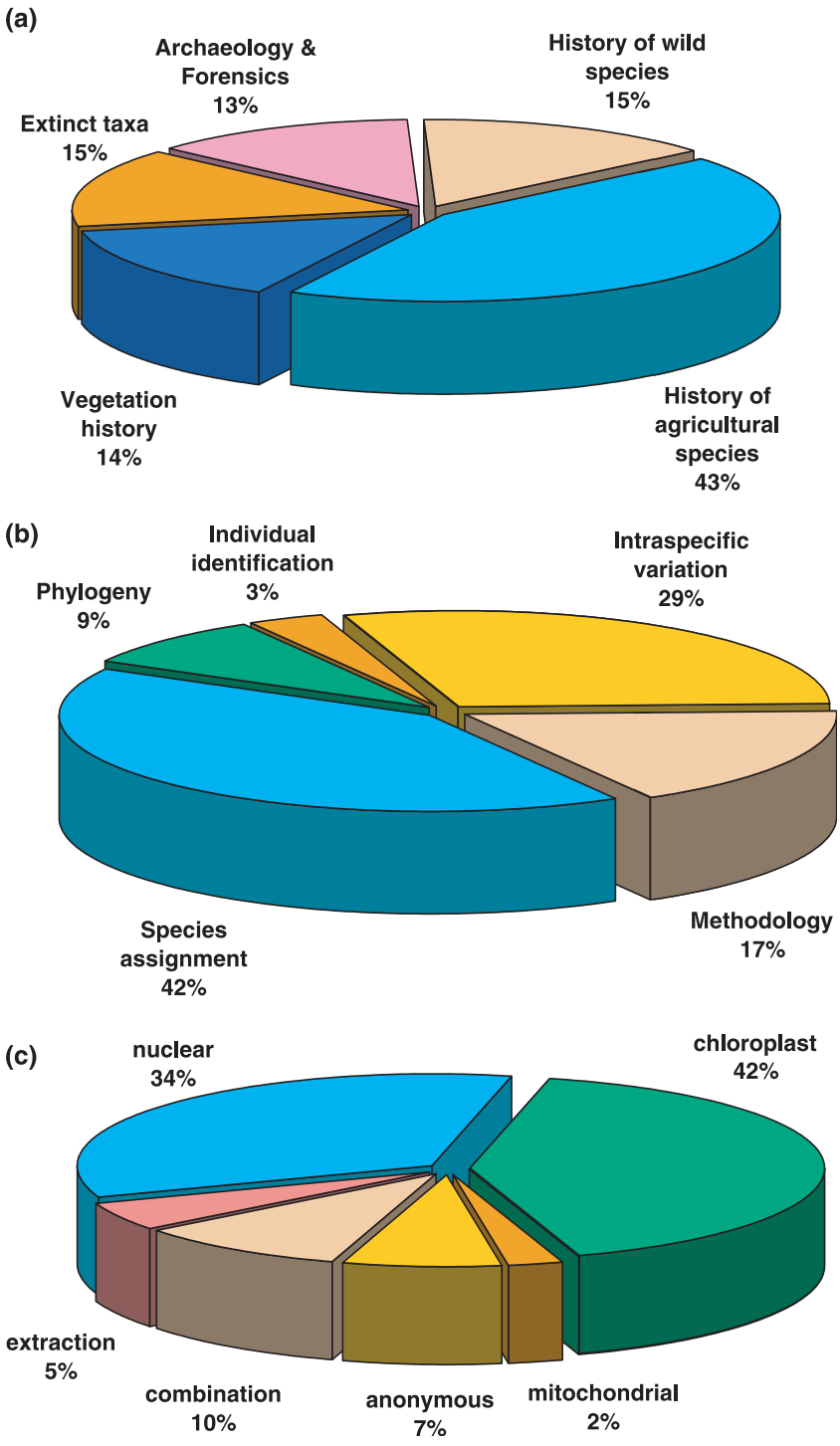

Fig. 2 Published papers on plant ancient DNA between 1983 and October 2004 (n = 57), retrieved from the Institute for Scientific Information (ISI) Web of Science. References are separated according to (a) the research area and types of study organisms (b) the study aims, and (c) the molecular approaches and type of markers used.

is notable that there was already a wealth of commentaries on the subject from the earliest studies, which illustrates the great controversy with regard to the prospects, the difficulties and the credibility of the results presented.

A more extensive literature search on aDNA studies in plants, excluding reviews and comments, yielded a total of 57 references. Almost half of the studies deal with cultivated taxa (Fig. 2a). Wild plant species have also been targeted, including a few extinct taxa. Finally, a suite of plant secondary products have been studied in the frame of forensics, archaeology or even for industrial purposes. Regardless of the type of species/material investigated, the main objectives have included species identification, phylogenetic assignment, and intraspecific diversity (Fig. 2b). It should be noted, however, that intraspecific

diversity is almost exclusively investigated in studies on allelic variation in wheat and maize, whereas only a few studies so far have dealt with the spatial distribution of genetic diversity over time (see later). Finally, the molecular markers used relate to the questions asked (Fig. 2c; see also later). For example, allelic diversity of cultivated plants was mainly resolved using well-characterized nuclear genes. By contrast, taxonomic identification and phylogenetic relationships relied on the analysis of more conserved sequences from the chloroplast (cp) genome, most often the large subunit of ribulose 1,5-biphosphate carboxylase/oxygenase (*rbc*L). Only two studies applied mitochondrial (mt) sequences, one of them in combination with nuclear and cpDNA markers (Deguilloux *et al.*, 2002).

Several reasons may account for the bias observed in the number of studies in plants compared with animals. First, a large fraction of aDNA studies are carried out in archaeology, in which the main focus is human history. Accordingly, the aDNA literature reflects the relative interests for various biological remains in this area (humans first, followed by animals and plants). Second, the bias could be due to differences in (1) sample preservation (2) DNA isolation, and (3) polymerase chain reaction (PCR) inhibition. We are not aware of any convincing evidence on fundamental differences between plant and animal DNA preservation, whereas there might be intrinsic difficulties in purifying aDNA from plant remains. Plant tissues are generally richer in primary or secondary metabolites, such as carbohydrates or phenolic compounds, than are those of animals (Ziegenhagen *et al.*, 1993). Moreover, the plant parts that are best represented among archaeological or palaeobotanical material, such as wood and charred seeds, are especially rich in PCR inhibitors. In fact, until just a few years ago, the only reference that would show up in a literature search using 'wood' and 'DNA' as keywords dealt with the inhibitory effects of wooden toothpicks, used to replicate bacterial colonies, on subsequent PCR amplification (Lee & Cooper, 1995). As had happened in the early 1990s for fresh plant tissues, technical adjustments of DNA isolation protocols are needed for plant fossil remains, such as post-purification procedures designed to remove inhibitors of the polymerase chain reaction (Banerjee & Brown, 2002). Another reason for the plant–animal study bias could be the low level of polymorphism present in cpDNA and mtDNA and the more limited number of appropriate markers that are readily available. More trivially, plant species, especially wild ones, might have been considered less attractive targets for aDNA investigations. Although human aDNA studies also present specific difficulties because of possible contamination through handling of ancient remains, the field has attracted considerable interest and experienced rapid progress.

Nevertheless, there are now signs that plant palaeogenetic research is attracting broader interest (Parducci & Petit, 2004). As primary producers, plants represent the bulk of both the biomass and the necromass (i.e. all living and dead organic material) present on Earth, and terrestrial palaeoecology relies predominantly on plant remains (e.g. pollen and spores). As a result, fossil plant material is abundant, and detailed palaeoecological information is available for hypothesis testing using aDNA.

Archaeogenetics is leading the way

One key area of research where plant aDNA studies proved to be particularly promising is the study of plant domestication. There have been several reports of aDNA isolated from archaeological remains of cultivated plants (Fig. 2a). Archaeological sites are rich sources of relatively well preserved and well-dated fossil remains usually not older than a few thousand years. In addition, archaeogenetics can rely on the many molecular markers developed for modern crops (Brown, 1999; Manen *et al.*, 2003).

Such studies have already contributed to a great extent to our understanding of issues such as the origin and the spread of agriculture (Freitas *et al.*, 2003), crop evolution (Jaenicke-Després *et al.*, 2003) or gene family evolution (Allaby *et al.*, 1999). For example, Jaenicke-Després *et al.* (2003) have recently provided evidence for early allelic selection of maize characters that are not directly observable on fossil specimens, such as plant architecture or starch characteristics. Furthermore, progress has been made in understanding the diagenesis of DNA or other biomolecules present in archaeological plant material. For example, DNA degradation during charring has been investigated experimentally (Chalfoun & Tuross, 1999; Threadgold & Brown, 2003) and results indicate the preservational limits of DNA targets that can be amplified from such remains. In the future, preserved samples such as wood (Dumolin-Lapègue *et al.*, 1999) or other construction material from buildings (Blatter *et al.*, 2002), but also from furniture, cultural, historical and archaeological artifacts (Burger *et al.*, 2000) or collections, might also provide research opportunities and applications in palaeogenetics.

Phylogeny and the analysis of antediluvian sequences

The use of plant aDNA for phylogenetic purposes has been usually limited to herbarium specimens not older than a few-hundred years (Savolainen *et al.*, 1995). During the late 1980s and early 1990s, the hunt for antediluvian plant DNA sequences (Golenberg *et al.*, 1990; Soltis *et al.*, 1992; Poinar *et al.*, 1993) had attracted much attention, paralleling similar work with extinct animals. Yet soon after, disillusion dominated the field because repeatability and authenticity of these reports were questioned (Cooper & Wayne, 1998; Gutiérrez & Marín, 1998). It is now commonly believed, albeit difficult to demonstrate, that DNA is completely degraded in samples older than one million years (Wayne *et al.*, 1999; Hofreiter *et al.*, 2001b; Pääbo *et al.*, 2004). The recent claim for the detection of plant aDNA from the Miocene (17–20 million yr before present;

Kim *et al.*, 2004) is unlikely to modify this opinion, in part because the same sequences have not been independently determined in another laboratory. Investigators studying such ancient samples will need to provide a considerable wealth of evidence to convince the scientific community about sequence authenticity.

Species identification in bulked samples

A recent trend in aDNA research is to identify species from sequences extracted and amplified from bulked fossil samples (e.g. from dung, intestinal contents or even raw sediments; Fig. 2b), in which individual remains, including those of plants, are not always precisely identified macroscopically. As an example, coprolites of herbivorous animals, representing fossilized fecal material, have been used as sources of plant aDNA (Poinar *et al.*, 2001; Hofreiter *et al.*, 2003) to infer diets as well as the vegetation present at the deposition time of the feces, up to 14 000 yr ago. In the same way, the middens of rodents were analysed to deduce the composition of Quaternary vegetation (Kuch *et al.*, 2002), whereas the contents of the intestine and colon of Ötzi, the Tyrolean iceman who lived some 5000 yr ago, were investigated to reconstruct his last meals (Rollo *et al.*, 2002). Recently, Willerslev *et al.* (2003) retrieved plant aDNA from a diversity of soil sediments, including permafrost dated to up to 400 000 yr before present, providing genetic information from this material in the absence of macrofossil evidence. In this paper, the authors argue that the sequences they amplified represent the oldest authenticated ancient DNA sequences known to date. The evaluation of more diverse sediments, including those from temperate settings, and the concomitant analysis of microbial communities might represent promising avenues for further development in plant palaeogenetics (Willerslev *et al.*, 2003). However, one should be aware of the possibility that DNA molecules may leak from decaying samples and subsequently be moved across sediment layers by water percolation, making authentication difficult.

These methods make use of available databases (e.g. EMBL or GENBANK) for sequence comparison (e.g. to reconstruct entire plant communities from the Quaternary). However, taxonomic resolution is limited by the number of species for which sequences are available in the databases. Systematic sequencing of complete local or regional floras using highly conserved primer pairs, such as *rbc*L (Chase *et al.*, 1993) or *trn*L-F (Taberlet *et al.*, 1991) of cpDNA, would enable researchers to exploit the full power of such approaches.

Studies based on bulked samples do not allow for an assessment of the authenticity of every sequence obtained. Instead, clusters of consensus sequences from several clones have to be used (Poinar *et al.*, 1998; Hofreiter *et al.*, 2000; Hofreiter *et al.*, 2003), resulting in reduced taxonomic resolution. As a consequence, it is only the overall reconstructed plant community composition that is used to evaluate if the results are reasonable. For example, dung segments of an extinct ground sloth from south-western USA (Poinar *et al.*, 1998) or Argentina (Hofreiter *et al.*, 2003) were analysed for chloroplast DNA and for pollen plus plant cuticle remains, respectively. In both regions, there was considerable overlap between genetic and morphological determinations, at least at the family level.

Towards population palaeogenetics

Recent attention has also been directed towards the study of aDNA diversity in wild plant species (i.e. DNA polymorphisms among and within populations) taking advantage of the knowledge from available phylogeographical survey). Such studies provide new insights into diachronical processes by considering more than just one point in time at a given location. For example, < 100-yr old herbarium material has been used to reconstruct the timing and route of invasion by a nonnative genotype of the common reed *Phragmites australis* into North America during the twentieth century (Saltonstall, 2002, 2003). Similarly, allelic diversity in nuclear microsatellites was determined from samples of the seagrass *Posidonia oceanica* up to 4000 yr old (Raniello & Procaccini, 2002). In another study, nucleotide diversity of several nuclear genes was estimated from DNA amplified from fossil heartwood of six trees of Japanese cedar *Cryptomeria japonica* that had been buried for about 3600 yr, and the results were compared with data from nearby present-day populations (Tani *et al.*, 2003). Ideally, for such population surveys, several samples suitable for aDNA analysis are required from each location and time period. Therefore, proving the authenticity of each of the many samples could rapidly become very limiting. To cope with high sample numbers, adequate replication of DNA extractions and PCRs could be performed while restricting cloning/sequencing to only a subset of samples (Taberlet *et al.*, 1996; Spencer & Howe, 2004).

Choosing appropriate techniques and markers

The use of real-time PCR in aDNA studies

Focusing on the population level requires careful evaluation of the molecular techniques used to answer each specific question. Real-time PCR might become a standard in the study of aDNA (Pruvost & Geigl, 2004), since it combines several advantages as outlined below.

Real-time PCR provides a reliable method to quantify the DNA concentration originally present in an extract of a target organism (Poinar *et al.*, 2003). By using highly specific primers and an oligonucleotide probe complementary to an internal segment of the target sequence, the amplification of any nontarget – ancient or recent – DNA present in the extract (e.g. resulting from microorganismal contamination) is excluded a priori. Like the primers, the probe anneals to the template sequence, but is successively cleaved by the polymerase during the elongation of the newly synthesized DNA strand. Probe cleavage emits a fluorescent signal that is recorded during

each amplification cycle. The original template DNA quantity added to the reaction mix can then be calculated. At the same time, it is possible to identify a single-nucleotide polymorphism (SNP) or a short insertion/deletion (indel) by adding two probes, labeled with different fluorescent dyes, which are complementary to either of the two sequence variants (allelic discrimination). Depending on which probe hybridizes to the target sequence, the respective dye, or both in case of heterozygosity when analysing a nuclear marker, emits a fluorescent signal. Thus, both organelle or nuclear markers can be screened (Alonso *et al.*, 2003). The SNPs or presence/absence of short indels can be identified, targeting short DNA fragments of about 60–300 nucleotides, a size compatible with degraded DNA.

However, such an allelic discrimination approach has some limitations when working with aDNA. Base modifications, which may by detected by a thorough, but laborious, PCR amplification and cloning/sequencing approach, could lead to wrong allele assignment. Using uracil *N*-glycosylase (UNG; Longo *et al.*, 1990) could serve as a loophole (Hofreiter *et al.*, 2001a). This enzyme removes uracil from DNA molecules, which produces an abasic site and thus leads to a break in the DNA strand. It may be routinely applied in real-time PCR, where thymine (dTTP) is replaced with uracil (dUTP) in the PCR mix for incorporation in the elongating DNA strand. At the start of the PCR cycling, UNG activity destroys any formerly amplified DNA molecules before being heat-inactivated – a positive side-effect in aDNA work because it helps to prevent cross-contamination with PCR products from earlier amplifications. However, UNG equally affects aDNA templates that contain uracil instead of cytosine because of deamination (Lindahl, 1993). While UNG treatment might thus prevent the potential genotyping of a wrong probe-compatible sequence, it may also destroy too many of the few authentic DNA molecules left in the extract of a fossil sample. As a consequence, preliminary real-time PCR without UNG could be run in order to evaluate the likelihood of finding sufficient templates to risk partial decomposition.

This technique not only corroborates credibility of the results obtained, but also allows for high-throughput genotyping since no post-PCR treatment, such as gel electrophoresis, is necessary. The PCR tubes remain closed throughout the analysis, thereby minimizing the risk of contamination predominantly by PCR products. Because PCRs beginning with < 1000 template molecules are likely to provide inaccurate results (Cooper & Poinar, 2000; Hofreiter *et al.*, 2001b), the quantification of the initial amount of DNA in the sample could help evaluate the authenticity of the sequences obtained. Ultimately, however, additional evidence is necessary for distinguishing aDNA from recent, contaminating DNA, and at least part of the samples have to be cloned and sequenced (Cooper & Poinar, 2000).

Potential applications of real-time PCR analyses in aDNA studies are abundant. In general, the combination of universal primers (e.g. *rbc*L, *trn*L-F) and species-specific probes represents a promising strategy. For example, postglacial migration of plant species, or even colonization routes of particular genotypes or lineages, could be integrated over space and time, particularly where present-day patterns of genetic variation based on easily accessible polymorphisms are clear-cut. Suitable model systems are manifold. White oaks in Europe exhibit ancient divides between several maternal lineages that survived in different glacial refugia, characterized by SNPs or small indels in cpDNA (Ferris *et al.*, 1993; Dumolin-Lapègue *et al.*, 1999; Deguilloux *et al.*, 2002; Petit *et al.*, 2002). These cpDNA variants could be traced across chronosequences (i.e. the changes over time at a given location). Similarly, Norway spruce (*Picea abies*) shows a clear differentiation between two mitochondrial lineages, distinguished by several SNPs (Sperisen *et al.*, 2001). These lineages point to separate glacial histories. There is further evidence for two SNP-based mitotypes occurring in populations of the Carpathian mountains (F. Gugerli *et al.* unpublished), a region considered to be a glacial refugium for *P. abies*. A third example is provided by a SNP found in a cpDNA gene of extant silver fir (*Abies alba*) that shows a cline across central Europe (Liepelt *et al.*, 2002).

The choice among genomes and genes

As in any population genetic survey, selecting the appropriate molecular markers to address a particular question is essential. The choice depends on the type of question as well as the spatial and temporal resolution envisaged. While no general recommendation can be given, cpDNA and mtDNA represent promising targets because they are present in multiple copies within each plant cell. Consequently, they are most likely to be retrieved from fossil tissues. In the case of aDNA, the choice of markers is limited by constraints inherent in the use of highly degraded DNA. It has been demonstrated that hydrolytic and oxidative damage will degrade aDNA to short fragments no longer than 200 nucleotides (Yang, 1997; Pääbo *et al.*, 2004). Targeting such short fragments will decrease the information content of each analysis, compared with those typically employed with contemporary material. The inverse relationship between amplification success rate and the size of the amplicon has been confirmed with DNA isolated from dry wood, with effects still detectable for fragments shorter than 100 nucleotides (Deguilloux *et al.*, 2002). The same study confirmed that sequences present in multiple copies per cell (i.e. cpDNA, mtDNA and ribosomal DNA sequences) yield more reliable amplification compared with single-copy nuclear sequences. This is relevant since uniparentally inherited markers, such as cpDNA or mtDNA, are generally more strongly structured in space than are nuclear markers and have consequently been extensively used in phylogeographic studies (Petit *et al.*, 2003, 2005). Instead, microsatellite markers are prone to slippage during amplification, a problem that could be exacerbated when working with highly degraded templates, making the cloning of each

genotype a crucial and necessary step to support allelic data. Surprisingly, anonymous markers such as random amplified polymorphic DNA (RAPDs) have also been used in a few instances, although these techniques are known to be prone to inconsistencies, even when working with DNA isolated from fresh tissues. Advantages and limitations of particular marker types should be considered when planning palaeogenetic studies.

Which plant parts are likely to contain aDNA?

Theoretically, plants, with their long-term storage capability, should preserve aDNA well. As for studies using fresh material, young, meristematic tissues with a comparatively high density of cells could prove particularly suitable for palaeogenetic studies. Such samples have a higher chance of containing nondegenerated copies of target DNA fragments, in particular from the numerous organelle genomes per cell. Similarly, tissues with protective coverings, such as seeds and fruits, might be promising targets. In trees, much subfossil or fossil material is available for molecular studies, from wood to leaves and needles (Deguilloux *et al.*, 2002, 2003), fruits or seeds (Brown, 1999; Ziegenhagen *et al.*, 2003), or even single pollen grains (Suyama *et al.*, 1996). Pollen is one of the major sources of palaeobotanical information, but it could also be useful for aDNA molecular analysis. The outer walls of pollen grains are extremely resistant to chemical and physical attack. In suitable environments, such as the accumulating anaerobic sediments at the bottom of small lakes, they might be preserved indefinitely. However, to date, there has been only one published report of successful DNA amplification from fossil pollen: in 1996, *rbc*L sequences were claimed to have been amplified from isolated Japanese *Abies* pollen grains dating from over 100 000 yr ago (Suyama *et al.*, 1996). In principle, bulked pollen samples might also be analysed, provided that cloning procedures are used to separate individual sequences. Plant samples from drilled cores, however, are particularly prone to contamination with both ancient (but from a different time period) or recent pollen. Thus, care should be taken to unearth fossil plant tissue outside the flowering season. Similarly, samples for DNA extraction are best retrieved in a clean-air room under elevated pressure, and care should be taken to avoid collecting material from the sample surface.

Choice of sampling location

Sampling locations should be carefully selected, bearing in mind the chance of finding sufficient numbers of well-conserved fossils. DNA persistence is greatly reduced when exposed to high temperatures (Wayne *et al.*, 1999). Thus, the most promising grounds are found in cold areas, where tissue is likely to be well preserved at low temperature (e.g. frozen in permafrost soil) (Hofreiter *et al.*, 2001b; Willerslev *et al.*, 2003). However, very dry environments, such as deserts, have also yielded high-quality aDNA sources (Poinar *et al.*, 2003). This evidence suggests that temperature might be effective towards DNA degeneration only indirectly, through its influence on water availability, since degradation of DNA involves hydrolytic processes (Willerslev *et al.*, 2004).

In formerly forested areas, sediments are most likely to contain macro-remains of woody plant species, particularly trees. The wealth of phylogeographic information currently available for temperate forest trees (Tomaru *et al.*, 1998; Cannon & Manos, 2003; Petit *et al.*, 2003) could serve as a starting point for aDNA studies. The same holds for plants from the Arctic (Abbott & Brochmann, 2003) and the Alps (Stehlik, 2003).

Preservation of macromolecules as an indication for the presence of aDNA

Even in a priori suitable sites, aDNA appears to be retrievable from only a small fraction of the sample material (Deguilloux *et al.*, 2002). Therefore, it would be desirable to classify samples according to their prospect of containing DNA molecules before starting genetic analyses (Bada *et al.*, 1999). Macroscopic sample evaluation often appears insufficient and sample age *per se* is not unambiguously indicative of the preservational state of a tissue. Rather, it is the condition of a sample at the time of burial, its environment during early storage and the fossilization processes involved that have the most significant effects on DNA preservation (Yang, 1997). Still, there is a time effect of sample age on PCR success, as seen in a study in which wood samples were retrieved along the cross-section of a tree trunk. Younger tissues from the outer part of the trunk were more likely to yield DNA for PCR amplification than were samples from older wood from the stem's heart, and there was a better chance of obtaining longer amplified DNA fragments compared with the older samples (Deguilloux *et al.*, 2002).

Several measures relying on the preservation of biomolecules other than nucleic acids have been proposed in order to estimate whether the presence of aDNA is likely in a given fossil sample (Bada *et al.*, 1999), and such approaches should be encouraged. For example, Poinar *et al.* (1996) introduced amino acid racemization for this purpose. Racemization refers to the change of the three-dimensional structure of amino acids from one form into a mirror form (isomer), resulting in measurable change in optical behavior. A sufficiently low ratio of the D-form vs the L-isomer of specific amino acids is indicative of the preservation of other macromolecules, such as DNA (Poinar *et al.*, 1996). Such information supports the authenticity of subsequent PCR-based findings and precludes the risk of circular argumentation when relying only on DNA-based evidence (Bada *et al.*, 1999). Nevertheless, the analyses of such alternative biomolecules require the same precautions, with respect to potential contamination from contemporary tissue, as aDNA studies, and positive results will not guarantee successful aDNA amplification.

Further directions in aDNA research

Despite persistent difficulties and technical bottlenecks, advances during recent years have contributed to the regained confidence and credibility in palaeogenetic studies and have opened up a vast field of research. Related applications, such as forensics, traceability and certification of diverse plant products, including food, drink (e.g. wine), fibers or wood (Manen *et al.*, 2003; Deguilloux *et al.*, 2004; Fladung *et al.*, 2004), conservation genetics (as in the control of the autochthony of a population by comparison with the genetic structure of its precursors), and even palaeoclimatology (Jahren *et al.*, 2004) might help to further establish the field by providing additional impetus and support. Critical judgement of one's own achievements and results throughout the entire analysis, from sampling to data analysis, is mandatory. Thorough peer evaluation might further contribute to the acceptance of new findings. We believe that there are still many unexplored opportunities for palaeogenetic studies in plants, provided that all research groups involved realize that significant investment is needed in terms of preparatory work with material, development of specific markers, laboratory equipment and isolation, methodological discipline and integrated expertise from many fields such as palaeontology, molecular biology, biochemistry, geology and ecology.

Acknowledgements

We thank Keith Bennett, Robert Blatter, Marie-France Deguilloux, Susan Hoebee, Rolf Holderegger, and the anonymous referees for critically reading and improving earlier versions of the article. Financial support was granted through the EU-funded project Fossilva (EVK2-1999-00015P; F. G., R. J. P.), the Swiss Federal Office for Education and Science (BBW 99.0689-2), the Swiss National Science Foundation (3100A0-101218; F. G.), the Laboratoire de Recherches sur les Monuments Historiques (R. J. P.), the Swedish Research Council, the Swedish Foundation for International Cooperation in Research and Higher Education (STINT) and the Carl Trygger Foundation (L. P.).

References

Abbott RJ, Brochmann C. 2003. History and evolution of the arctic flora: in the footsteps of Eric Hultén. *Molecular Ecology* **12**: 299–313.

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.

Alonso A, Martín P, Albarrán C, García P, Primorac D, García O, Fernández de Simón L, García-Hirschfeld J, Sancho M, Fernández-Piqueras J. 2003. Specific quantification of human genomes from low copy number DNA samples in forensic and ancient DNA studies. *Croatian Medical Journal* **44**: 273–280.

Austin JJ, Smith AB, Thomas RH. 1997. Palaeontology in a molecular world: the search for authentic ancient DNA. *Trends in Ecology and Evolution* **12**: 303–306.

Bada JL, Wang XYS, Hamilton H. 1999. Preservation of key biomolecules in the fossil record: current knowledge and future challenges. *Philosophical Transactions of the Royal Society of London, Series B* **354**: 77–86.

Banerjee M, Brown TA. 2002. Preservation of nuclear but not chloroplast DNA in archaeological assemblages of charred wheat grains. *Ancient Biomolecules* **4**: 59–63.

Bennett KD. 1997. *Evolution and ecology: the pace of life*. Cambridge, UK: Cambridge University Press.

Blatter RHE, Jacomet S, Schlumbaum A. 2002. Spelt-specific alleles in HMW glutenin genes from modern and historical European spelt (*Triticum spelta* L.). *Theoretical and Applied Genetics* **104**: 329–337.

Brewer S, Cheddadi R, de Beaulieu JL, Reille M, Data Contributors. 2002. The spread of deciduous *Quercus* throughout Europe since the last glacial period. *Forest Ecology and Management* **156**: 27–48.

Brown TA. 1999. How ancient DNA may help in understanding the origin and spread of agriculture. *Proceedings of the Royal Society of London, Series B* **354**: 89–98.

Burger J, Hummel S, Herrmann B. 2000. Palaeogenetics and cultural heritage. Species determination and STR-genotyping from ancient DNA in art and artefacts. *Thermochimica Acta* **365**: 141–146.

Cannon CH, Manos PS. 2003. Phylogeography of the Southeast Asian stone oaks (*Lithocarpus*). *Journal of Biogeography* **30**: 211–226.

Chalfoun DJ, Tuross N. 1999. Botanical remains: utility in protein and DNA research. *Ancient Biomolecules* **3**: 67–79.

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu Y-L, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedrén M, Gaut BS, Jansen RK, Kim K-J, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang Q-Y, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn GH Jr, Graham SW, Barrett SCH, Dayanadan S, Albert VA. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* **80**: 528–580.

Cooper A, Poinar HN. 2000. Ancient DNA: do it right or not at all. *Science* **289**: 1139.

Cooper A, Wayne R. 1998. New uses for old DNA. *Current Opinion in Biotechnology* **9**: 49–53.

Deguilloux M-F, Pemonge M-H, Petit RJ. 2002. Novel perspectives in wood certification and forensics: dry wood as a source of DNA. *Proceedings of the Royal Society of London, Series B* **269**: 1039–1046.

Deguilloux M-F, Pemonge M-H, Bertel L, Kremer A, Petit RJ. 2003. Checking the geographical origin of oak wood: molecular and statistical tools. *Molecular Ecology* **12**: 1629–1636.

Deguilloux M-F, Pemonge M-H, Petit RJ. 2004. DNA-based control of oak wood geographic origin in the context of the cooperage industry. *Annals of Forest Science* **61**: 97–104.

Dumolin-Lapègue S, Pemonge M-H, Gielly L, Taberlet P, Petit RJ. 1999. Amplification of oak DNA from ancient and modern wood. *Molecular Ecology* **8**: 2137–2140.

Elenga H, Peyron O, Bonnefille R, Jolly D, Cheddadi R, Guiot J, Andrieu V, Bottema S, Buchet G, de Beaulieu J-L, Hamilton AC, Maley J, Marchant R, Perez-Obiol R, Reille M, Riollet G, Scott L, Straka H, Taylor D, van Campo E, Vincens A, Laarif F, Jonson H. 2000. Pollen-based biome reconstruction for southern Europe and Africa 18000 yr BP. *Journal of Biogeography* **27**: 621–634.

Ferris C, Oliver RP, Davy AJ, Hewitt GM. 1993. Native oak chloroplasts reveal an ancient divide across Europe. *Molecular Ecology* **2**: 337–344.

Fladung M, Nowitzki O, Ziegenhagen B, Markussen T. 2004. Identification of transgenes from wood of genetically transformed poplar trees. *Journal of Wood Science and Technology* **38**: 207–215.

Freitas FO, Bendel 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. *Journal of Archaeological Science* **30**: 901–908.

Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, Henderson D, Zurawski G. 1990. Chloroplast DNA sequence from a Miocene magnolia species. *Nature* **344**: 656–658.

Gutiérrez G, Marín A. 1998. The most ancient DNA recovered from an amber-preserved specimen may not be as ancient as it seems. *Molecular Biology and Evolution* **15**: 926–929.

Hofreiter M, Poinar HN, Spaulding WG, Bauer K, Martin PS, Possnert G, Pääbo S. 2000. A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology* **9**: 1975–1984.

Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S. 2001a. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research* **29**: 4793–4799.

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. 2001b. Ancient DNA. *Nature Reviews Genetics* **2**: 353–359.

Hofreiter M, Betancourt JL, Sbriller AP, Markgraf V, McDonald HG. 2003. Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Cura, Neuquen Province, southwest Argentina. *Quaternary Research* **59**: 364–378.

Hummel S. 2003. *Ancient DNA typing: methods, strategies and applications.* Berlin, Germany: Springer.

Jaenicke-Després V, Buckler ES, Smith BD, Gilbert MTP, Cooper A, Doebley J, Pääbo S. 2003. Early allelic selection in maize as revealed by ancient DNA. *Science* **302**: 1206–1208.

Jahren AH, Petersen G, Seberg O. 2004. Plant DNA: a new substrate for carbon stable isotope analysis and a potential paleoenvironmental indicator. *Geology* **32**: 241–244.

Kim S, Soltis DE, Soltis PS, Suh Y. 2004. DNA sequences from Miocene fossils: an *ndh*F sequence of *Magnolia latahensis* (Magnoliaceae) and an *rbc*L sequence of *Persea pseudocarolinensis* (Lauraceae). *American Journal of Botany* **91**: 615–620.

Kuch M, Rohland N, Betancourt JL, Latorre C, Steppan S, Poinar HN. 2002. Molecular analysis of a 11 700-year-old rodent midden from the Atacama Desert, Chile. *Molecular Ecology* **11**: 913–924.

Lascoux M, Palmé AE, Cheddadi R, Latta RG. 2004. Impact of Ice Ages on the genetic structure of trees and shrubs. *Proceedings of the Royal Society of London, Series B* **359**: 197–207.

Lee AB, Cooper TA. 1995. Improved direct PCR screen for bacterial colonies – wooden toothpicks inhibit PCR amplification. *Biotechniques* **18**: 225–226.

Liepelt S, Bialozyt R, Ziegenhagen B. 2002. Wind-dispersed pollen mediates postglacial gene flow among refugia. *Proceedings of the National Academy of Sciences, USA* **99**: 14590–14594.

Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* **362**: 709–715.

Litt T, Schmincke HU, Kromer B. 2003. Environmental response to climatic and volcanic events in central Europe during the Weichselian Lateglacial. *Quaternary Science Reviews* **22**: 7–32.

Longo MC, Berninger MS, Hartley JL. 1990. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene* **93**: 125–128.

Manen J-F, Bouby L, Dalnoki O, Marinval P, Turgay M, Schlumbaum A. 2003. Microsatellites from archaeological *Vitis vinifera* seeds allow a tentative assignment of the geographical origin of ancient cultivars. *Journal of Archaeological Science* **30**: 721–729.

McGraw JB. 1993. Ecological genetic variation in seed banks. IV. Differentiation of extant and seed bank-derived populations of *Eriophorum vaginatum*. *Arctic and Alpine Research* **25**: 45–49.

Morris AB, Baucom RS, Cruzan MB. 2002. Stratified analysis of the soil seed bank in the cedar glade endemic *Astragalus bibullatus*: evidence for historical changes in genetic structure. *American Journal of Botany* **89**: 29–36.

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

Parducci L, Petit RJ. 2004. Ancient DNA – unlocking plants' fossil secrets. *New Phytologist* **161**: 335–339.

Petit RJ, Csaikl UM, Bordács S, Burg K, Coart E, Cottrell J, van Dam B, Deans JD, Dumolin-Lapègue S, Fineschi S, Finkeldey R, Gillies A, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Mátyás G, Munro RC, Olalde M, Pemonge M-H, Popescu F, Slade D, Tabbener H, Taurichini D, de Vries SGM, Ziegenhagen B, Kremer A. 2002. Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management* **156**: 5–26.

Petit RJ, Aguinagalde I, de Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**: 1563–1565.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG. 2005. Comparative organisation of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* doi: 10.1111/j.1365–294X.2004.02410.x.

Poinar HN, Cano RJ, Poinar GO. 1993. DNA from an extinct plant. *Nature* **363**: 677.

Poinar HN, Höss M, Bada JL, Pääbo S. 1996. Amino acid racemization and the preservation of ancient DNA. *Science* **272**: 864–866.

Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz BA, Bland H, Evershed RP, Possnert G, Pääbo S. 1998. Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* **281**: 402–406.

Poinar HN, Kuch M, Sobolik KD, Barnes I, Stankiewicz AB, Kuder T, Spaulding WG, Bryant VM, Cooper A, Pääbo S. 2001. A molecular analysis of dietary diversity for three archaic Native Americans. *Proceedings of the National Academy of Sciences, USA* **98**: 4317–4322.

Poinar HN, Kuch M, McDonald G, Martin P, Pääbo S. 2003. Nuclear gene sequences from a late Pleistocene sloth coprolite. *Current Biology* **12**: 1150–1152.

Porsild AE, Pharington CR, Mulligan GA. 1967. '*Lupinus arcticus*' Wats. grown from seeds of Pleistocene age. *Science* **158**: 113–114.

Pruvost M, Geigl E-M. 2004. Real-time quantitative PCR to assess the authenticity of ancient DNA amplification. *Journal of Archaeological Science* **31**: 1191–1197.

Raniello R, Procaccini G. 2002. Ancient DNA in the seagrass *Posidonia oceanica*. *Marine Ecology – Progress Series* **227**: 269–273.

Rollo F, Ubaldi M, Ermini L, Marota I. 2002. Ötzi's last meals: DNA analysis of the intestinal content of the Neolithic glacier mummy from the Alps. *Proceedings of the National Academy of Sciences, USA* **99**: 12594–12599.

Saltonstall K. 2002. Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. *Proceedings of the National Academy of Sciences, USA* **99**: 2445–2449.

Saltonstall K. 2003. Microsatellite variation within and among North American lineages of *Phragmites australis*. *Molecular Ecology* **12**: 1689–1702.

Savolainen V, Cuénoud P, Spichiger R, Martinez MDP, Crèvecoeur M, Manen J-F. 1995. The use of herbarium specimens in DNA phylogenetics – evaluation and improvement. *Plant Systematics and Evolution* **197**: 87–98.

Schneller JJ. 1998. How much genetic variation in fern populations is stored in the spore banks? A study of *Athyrium filix-femina* (L.) Roth. *Botanical Journal of the Linnean Society* **127**: 195–206.

Shen-Miller J. 2002. Sacred lotus, the long-living fruits of China Antique. *Seed Science Research* **12**: 131–143.

Soltis PS, Soltis DE, Smiley CJ. 1992. An *rbc*L sequence from a Miocene *Taxodium* (bald cypress). *Proceedings of the National Academy of Sciences, USA* **89**: 449–451.

Spencer M, Howe CJ. 2004. Authenticity of ancient-DNA results: a statistical approach. *American Journal of Human Genetics* **75**: 240–250.

Sperisen C, Büchler U, Gugerli F, Mátyás G, Geburek T, Vendramin GG. 2001. Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. *Molecular Ecology* **10**: 257–263.

Stehlik I. 2003. Resistance or emigration? Response of alpine plants to the ice ages. *Taxon* **52**: 499–510.

Suyama Y, Kawamuro K, Kinoshita I, Yoshimura K, Tsumura Y, Takahara H. 1996. DNA sequence from a fossil pollen of *Abies* spp. from Pleistocene peat. *Genes and Genetic Systems* **71**: 145–149.

Taberlet P, Cheddadi R. 2002. Quaternary refugia and persistence of biodiversity. *Science* **297**: 2009–2010.

Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**: 1105–1109.

Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**: 3189–3194.

Tani N, Tsumura Y, Sato H. 2003. Nuclear gene sequences and DNA variation of *Cryptomeria japonica* samples from the postglacial period. *Molecular Ecology* **12**: 859–868.

Threadgold J, Brown TA. 2003. Degradation of DNA in artificially charred wheat seeds. *Journal of Archaeological Science* **30**: 1067–1076.

Tomaru N, Takahashi M, Tsumura Y, Takahashi M, Ohba K. 1998. Intraspecific variation and phylogeographic patterns of *Fagus crenata* (Fagaceae) mitochondrial DNA. *American Journal of Botany* **85**: 629–636.

Wayne RK, Leonard JA, Cooper A. 1999. Full of sound and fury: the recent history of ancient DNA. *Annual Review of Ecology and Systematics* **30**: 457–477.

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

Willerslev E, Hansen AJ, Poinar HN. 2004. Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology and Evolution* **19**: 141–147.

Yang H. 1997. Ancient DNA from Pleistocene fossils: preservation, recovery, and utility of ancient genetic information for quaternary research. *Quaternary Science Reviews* **16**: 1145–1161.

Ziegenhagen B, Guillemaut P, Scholz F. 1993. A procedure for mini-preparations of genomic DNA from needles of silver fir (*Abies alba* Mill.). *Plant Molecular Biology Reporter* **11**: 117–121.

Ziegenhagen B, Liepelt S, Kuhlenkamp V, Fladung M. 2003. Molecular identification of individual oak and fir trees from maternal tissues of their fruits or seeds. *Trees* **17**: 345–350.