Ancient DNA – unlocking plants’ fossil secrets

Introducing genetic and palaeogenetic approaches in plant palaeoecology and archaeology, Bordeaux, France, September 2003

Disillusion rather than enthusiasm had predominated among scientists after the first reports of successful but sometimes hardly repeatable retrieval of ancient DNA (aDNA) from very ancient specimens such as Miocene fossil leaf samples or even dinosaur eggs. In the past few years, however, scrupulous studies including a suite of controls for authenticity have provided the basis for regained confidence in the field of palaeogenetics. Furthermore, it has now been shown that animal and especially plant aDNA is not only present in fossil tissues but may even be indirectly retrieved from secondary samples such as fossil faeces (Poinar et al., 1998; Poinar et al., 2001) or sediments (Willerslev et al., 2003). The recent Bordeaux symposium on the use of plant aDNA provided a forum for this renewed enthusiasm and illustrated the substantial progresses that are currently being made. A particular focus was on the main European forest tree species during the Quaternary period – using aDNA as a link between phylogeographic and classical palaeoecological reconstructions (the topic of the FOSSILVA project, which brought together teams of geneticists and palaeoecologists, coordinated by Jacques-Louis de Beaulieu (CNRS – Marseille, France)).

'Retrieving DNA information from well preserved Quaternary fossil material is no longer a dream'

Reconciling neoecology and palaeoecology

Until recently, reconstructing past plant population dynamics has relied either on studies of genetic variation based on contemporary populations or on the fossil record. The two approaches are somewhat discordant. Neoecologists can investigate all aspects of an organism’s phenotype and genotype but have no direct access to population dynamics (although recent developments, such as the availability of population-based DNA sequence data, and approaches, such as phylogeography (Avise, 2000) and the coalescent (Kingman, 1982), have introduced a temporal dimension that was missing in former equilibrium-based analyses). Palaeoecologists, on the contrary, have direct access to ‘time’, but can rely only on phenotypes, and usually incomplete ones. Nevertheless, as a result of some of the first truly interdisciplinary efforts between
the two groups (Cwynar & MacDonald, 1987; Comps et al., 2001; Petit et al., 2002, 2003) our understanding of postglacial migration in plants has been rapidly improving. In May 2003, another conference organised by the Royal Society in London on the evolutionary legacy of the ice ages allowed disparate groups to examine the degree to which the alternation of cold and warm stages during the Quaternary had affected the evolution of both plants and animals. These and other recent advances have contributed to an emerging synthesis on the evolutionary significance of the ice ages.

Combining genetic and palaeoecological data

In recent years the use of pollen and macrofossil databanks has contributed significantly to a continuous increase in our knowledge of past plant distribution (Elenga et al., 2000; Litt et al., 2003) and ecology (Bennett, 1997). In Bordeaux, Eric Grimm (Illinois State Museum, USA) showed recent development achieved with the Global Pollen Database (GPD), which currently contains Quaternary pollen data from the Americas, Europe, Africa, Asia, and the Indo-Pacific regions. The objective is to assemble data from fossil deposits and modern surface samples into a relational database and to make them readily available to the scientific community (http://www.ngdc.noaa.gov/paleo/gpd.html). Broad-scale palaeoecological patterns inferred from such mapped pollen data have largely confirmed that plant communities are impermanent assemblages of taxa and often have no modern analogues, as first argued by Davis (1976). Simultaneously, during the past 10 yr, population genetics based on modern DNA techniques has shed new light on the postglacial migration history of many temperate plants. At the symposium, Giovanni G. Vendramin (CNR, Florence, Italy) presented an example of such studies. An impressively large dataset of the genetic variation in European broadleaved and coniferous species had been assembled using chloroplast markers and compared to pollen analytical data. The results showed how several temperate tree species had been similarly affected by successive ice ages and by the dissected geography of southern Europe, despite responding individually in terms of timing and routes of expansion (Petit et al., 2003). In particular, genetic diversity is higher in the central part of the range than in the Mediterranean part for a majority of the trees and shrubs investigated, even though the latter region had clearly provided areas suitable for long-term refugia. Presumably, admixture had occurred at intermediate latitudes at the confluence of the various colonisation routes emerging from the Mediterranean peninsulas. The presence of plant refugia in central and southern Europe during the last glacial maximum has long been a matter of debate, which inevitably highlights one of the main shortcomings of the use of fossil pollen to reconstruct past vegetation history: pollen found at a certain site can represent both local and long-distance dispersal by wind. This limits the spatial resolution of the method, especially when the pollen accumulation rate is slow, as was the case during the last glacial maximum. Hence, other analytical methods should be sought, such as analyses of plant macrofossil assemblages from lake sediments and of subfossil wood preserved in the record as macroscopic charcoal.

An example where all three proxies (genetic markers, pollen and macrofossils) had been integrated was provided for European beech (Fagus sylvatica) by Donatella Magri (University La Sapienza, Roma) and colleagues from the FOSSILVA project. Pollen data from several hundred lake and mire cores, mostly retrieved from the European pollen database, were combined with more scattered, but well-dated, macrofossil data and with an extensive genetic dataset obtained from over 600 forests distributed throughout the species range. The synthesis suggested a new scenario for postglacial recolonization by beech, and indicated that colonisation events predating the last glacial maximum need to be taken into account to explain the contemporary genetic patterns of this species. Obviously, questions remain and further confirmation will be needed that might necessitate resorting to aDNA.

Ancient DNA – potential and pitfalls

A major constraint remains for population geneticists, who must extrapolate from modern molecular data. However, retrieving DNA information from well-preserved Quaternary fossil material is no longer a dream and should therefore allow this limitation to be bypassed. The survival of aDNA in specimens up to several thousand years old is now well established, and retrieving DNA from even older remains of the late Quaternary (up to 100 000 yr ago) has added further insights to the study of many evolutionary processes (Krings et al., 1997; Poinar et al., 1998; Leonard et al., 2000; Cooper et al., 2001; Barnes et al., 2002; Lambert et al., 2002).

Unfortunately, for various reasons, studies of plant aDNA have lagged behind those of animals and humans. Despite the wealth of plant subfossil material available for molecular study, from wood to leaves and needles or even single pollen grains (Suyama et al., 1996; Deguilloux et al., 2002, 2003; Ziegenhagen et al., 2003), aDNA appears to be retrievable in only a small fraction of the samples investigated. Furthermore, because of the minute amounts and degraded nature of aDNA there is a constant risk of contamination. The need for authentication became clear in the mid-1990s when a number of high-profile studies were shown to be nonreproducible, and in successive years standard criteria had to be developed to determine the ancient origin of DNA sequences (Handt et al., 1994; Cooper & Poinar, 2000) (Box 1).

DNA preservation in ancient specimens

One important property of aDNA is its fragmented nature, mostly caused by the hydrolysis of the DNA phosphodiester bonds and the N-glycosyl bonds. Typically, only a small proportion of fossil specimens contain DNA that can be amplified.
by PCR. In such cases, rapid screening methods can be useful to identify the large fraction of samples that are so badly preserved that there is little hope of retrieving any aDNA. In particular, amino acid analyses (counts of the total amount of amino acids preserved in a specimen, amino acid composition and extent of racemization) have proven to be very useful proxies for monitoring DNA preservation in fossil samples (Poinar et al., 1996). Results from such studies have suggested that the DNA preservation is mainly linked to the temperature and its constancy at a site, rather than to its age. An additional useful way to check the quality of old DNA sequences is the estimation of the number of template DNA molecules from which the PCR starts. Robert Blatter (Max Plank Institute, Leipzig, Germany) presented two studies, on 100-yr-old museum bones from 42 orang-utans and from 20 000-yr-old ground sloth coproliths, where different fragments of the mitochondrial rDNA region were quantified by RT-PCR to identify specimens suitable for population genetic studies. These methodologies are very promising and can be adapted to studies of other specimens, including plants.

**New uses for ancient DNA**

An intriguing potential use of aDNA research is in revealing the genetic record of temporal changes in populations over millennia time-scales. So far, genetic changes within populations have been reconstructed using only current patterns of genetic diversity. Given the dynamics of climate change and migration within continents, extrapolations based on such data are difficult. Ancient DNA can provide a direct record of the tempo and mode of genetic change within populations, and thus a means of testing and refining existing population models. The deposits with the largest potential are probably those of the Arctic permafrost, of high altitude caves, and of other cold and arid environments. Unfortunately, however, opportunities for these types of studies on aDNA are offered only when a sufficient number and distribution of fossil samples are available. As an example, in a recent study on fossil bones of Adélie penguins from Antarctica, Lambert et al. (2002) succeeded in measuring the rate of nucleotide evolution in the mitochondrial DNA by analysing an adequately large number of fossil samples. This study was possible because of the particular aspect of the life history of the animals studied and of the extreme environmental conditions in which the bones had been preserved. In plants, the retrieval and possibilities for analysis of well preserved fossilised hard tissues such as wood are very rare compared to the retrieval of bones in animals. Moreover, as explained at the symposium by Birgit Ziegenhagen (Phillips University of Marburg, Germany), the molecular analysis of woody, sclerchymatic or dry tissues from plants is methodologically very difficult, even when the tissues are sampled from extant specimens. In this context, Yoshihisa

---

**Box 1 Authenticity of ancient DNA sequences**

The extraction and preparation of the PCR required for amplification of ancient DNA (aDNA) must be done in a laboratory that is rigorously separated from work involving modern DNA, equipment should be regularly treated with bleach and UV irradiation, and protective clothing should always be worn (Handt et al., 1994; Höss et al., 1994). In addition to these basic lab procedures, a number of criteria have been developed to determine the ancient origin of DNA sequences (Cooper & Poinar, 2000):

- **Physically isolated work area** To avoid contamination, it is essential that, before the amplification stage, all aDNA research is carried out in a dedicated, isolated environment. A building in which large amounts of the target DNA are routinely amplified should be banned.
- **Control amplifications** Multiple extraction and PCR controls must be performed to detect sporadic or low-copy number contamination, although carrier effects do limit their efficacy. All contaminated results should be reported, and positive controls should generally be avoided, as they increase contamination risk.
- **Appropriate molecular behaviour** PCR amplification strength should be inversely related to product size (large 500- to 1000-bp products are unusual) and sequences should make phylogenetic sense.
- **Reproducibility** Results should be repeatable from the same, and different, DNA extracts of a given specimen.
- **Cloning** Direct PCR sequences must be verified by cloning amplified products to determine the ratio of endogenous to exogenous sequences, or damage-induced errors. Overlapping fragments are desirable to confirm that sequence variation is authentic and not the product of errors introduced when PCR amplification starts from a small number of damaged templates.
- **Independent replication** Intra-laboratory contamination can only be discounted when separate samples of a specimen are extracted and sequenced in independent labs.
- **Biochemical preservation** Indirect evidence for DNA survival in a specimen can be provided by assessing the total amount, composition, and relative extent of diagenetic change in amino acids and other residues.
- **Quantitation** The copy number of the target DNA should be assessed using quantitative RT-PCR. When the number of starting templates is low (< 1000), it may be impossible to exclude the possibility of sporadic contamination, especially for human DNA studies.
- **Associated remains** Evidence that DNA has survived in associated material from other species is critical supporting evidence. Such remains could also make good negative controls for PCR amplifications of the target species.
Suyama (Tohoku University, Sendai, Japan) presented a new and promising approach for the analysis of plant aDNA. Together with Laura Parducci (Uppsala University, Sweden), they have recently developed a method for the amplification and sequencing of short chloroplast DNA regions from single pollen grains of conifers isolated from Holocene lake sediments. If reproducible and applicable to different plant species, the method will allow direct estimates of the population dynamics in space and time and could reveal new details of the historical patterns of evolutionary change in plant species.

**Ancient DNA extracted from soil**

Recently, a team at Copenhagen University (Denmark) led by Eske Willerslev and Anders Hansen succeeded in extracting DNA from Siberian permafrost sediments (Fig. 1) and from soil sampled in temperate caves from New Zealand, ranging in age from 10 000 yr to 400 000 yr (Willerslev et al., 2003). As Willerslev explained in Bordeaux the team was initially looking for bacterial DNA but found out that they could recover fragments of plant chloroplast DNA in soil samples from permafrost cores. Eventually they were able to identify DNA sequences from at least 19 different plant taxa (angiosperms, gymnosperms and mosses) as well as from various animals including mammoth, bison and horse. Much of the plant DNA probably derived from roots, which would have been well protected under the frozen ground. The data presented suggested dramatic changes in the taxonomic diversity and composition of Beringian vegetation during the Quaternary. Perennial, dry, temperate cave sediments in New Zealand also yielded sequences of extinct biota, including two species of ratite moa and 29 plant taxa characteristic of the prehuman environment. Importantly, the latter result demonstrated that DNA, under particular conditions such as dry environments, could be preserved even in unfrozen conditions.

**Perspectives**

In plants, strict adherence to all the criteria for the authentication of aDNA is difficult and, in some cases, impossible to achieve – this was clear from many of the studies presented at Bordeaux. For example, the criteria of reproducibility, biochemical preservation and DNA quantitation are not applicable when studying single-pollen grains of plant specimens, because the fossil material is depleted with the first PCR analysis. Nevertheless, it is critical that the highest possible standards are applied if plant aDNA research is to remain credible. The Bordeaux meeting has highlighted the power of aDNA analysis as a technique – exemplified by the cross-over between palaeogenetics and palaeoecology, in which aDNA analysis is enabling new and sophisticated methods for hypothesis testing and model validation. The scientific rewards to both disciplines look set to follow the promising start of aDNA research.

**Acknowledgements**

The symposium was organised by Rémy Petit (INRA – Cestas, France), Marie-France Deguilloux (University of Bordeaux) and Jacques-Louis de Beaulieu (CNRS – Marseille) – we regret that only a selection of the research presented could be mentioned in the space of this report, but see http://www.pierroton.inra.fr/Paleo/index.htm for all abstracts. Sponsorship was provided by the Conseil Régional d’Aquitaine, INRA, CNRS and the New Phytologist Trust. FOSSILVA was supported by the Commission of the European Communities (project EVK2-1999–00015P). L. P. thanks the Swedish
Research Council and the Stint Foundation for support and K. D. Bennett (Uppsala University, Sweden) for helpful comments.

Laura Parducci1,2,* and Rémy J. Petit3

1Conservation Biology and Genetics (EBC), Uppsala University, Norbyvägen 18/D, S-752 36, Sweden;
2Palaeobiology program, Department of Earth Sciences, Uppsala University, Villavägen 16, SE-752 36 Uppsala, Sweden;
3Institut National de la Recherche Agronomique, UMR Biodiversité, Génomes et Écosystèmes, F-33612 Cestas, France

*Author for correspondence: tel +46 18 4716414; fax +46 18 4716414; email Laura.Parducci@ebc.uu.se

References


Key words: ancient DNA, palaeoecology, palaeogenetics, archeology, pollen database, PCR, DNA sequence.