Molecular Ecology (2005) doi: 10.1111/j.1365-294X.2005.02644.x

Ancient DNA from pollen: a genetic record of population history in Scots pine

L. PARDUCCI,*‡ Y. SUYAMA,† M. LASCOUX‡ and K. D. BENNETT*
*Department of Earth Sciences, Palaeobiology Program, Uppsala University, Villavägen 16, SE-752 36 Uppsala, Sweden, †Laboratory of Forest Ecology, Tohoku University, Naruko, Miyagi 989–6711 Japan, ‡Department of Evolution Genomics and Systematics, Evolutionary and Functional Genomics Program, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

Abstract

Assessments of plant population dynamics in space and time have depended on dated records of fossil pollen synthesized on a subcontinental scale. Genetic analyses of extant populations have revealed spatial relationships that are indicative of past spatial dynamics, but lack an explicit timescale. Synthesis of these data requires genetic analyses from abundant dated fossil material, and this has hitherto been lacking. Fossil pollen is the most abundant material with which to fill this data gap. Here we report genetic analyses of fossil pollen retrieved from Holtjärnen postglacial lake sediment in Sweden and show that plastid DNA is recoverable from Scots Pine and Norway spruce pollen grains that are 100 and 10 000 years old. By sequencing clones from two short plastid PCR products and by using multiple controls we show that the ancient sequences were endogenous to the fossil grains. Comparison of ancient sequences and those obtained from an extant population of Scots pine establishes the first genetic link between extant and fossil samples in this species, providing genetic continuity through time. The finding of one common haplotype present in modern, 100-year old and 10 000-year old samples suggests that it may have persisted near Holtjärnen throughout the postglacial period. This retrieval of ancient DNA from pollen has major implications for plant palaeoecology in conifer species by allowing direct estimates of population dynamics in space and time.

Keywords: ancient DNA, DNA sequence, microsatellite, palaeobiology, pollen, population history

Received 14 February 2005; revision accepted 13 May 2005

Introduction

In the last decade our understanding of plant population history has improved considerably as a result of interdisciplinary efforts between neoecologists and paleoecologists (e.g. Cwynar & MacDonald 1987; Petit *et al*. 2002, 2003). In addition, the availability of large DNA sequence data and the development of new analytical tools (Hein *et al*. 2005) have led to much stronger inferences on population histories. A major constraint, however, remains for population geneticists who must infer the past from modern molecular data.

Retrieving DNA information from well-preserved Quaternary fossil material should help bypass this limitation. In the last decade, the survival of ancient DNA (aDNA) in specimens up to several thousand years old has been established, and DNA information from specimens of even late Quaternary remains (up to 100 000 BP) has provided new insights on many evolutionary processes (Krings *et al*. 1997; Poinar *et al*. 1998; Leonard *et al*. 2000; Cooper *et al*. 2001; Lambert *et al*. 2002; Endicott *et al*. 2003; Shapiro *et al*. 2004). Nevertheless, methodological difficulties as well as the rarity of suitable well-preserved samples have often prevented a broader utility for aDNA studies. This is particularly true in plants, where well-preserved hard tissues, such as wood, are difficult to obtain in sufficient number, over large enough geographical areas. Moreover, because most wood cells lack viable nuclei and plastids, its genetic analysis is hampered by technical difficulties. Studies of plant aDNA have therefore lagged well behind those of animals, despite the wealth of subfossil material available for molecular studies, from wood to leaves and needles and even pollen grains (Suyama *et al*.

Correspondence: Laura Parducci, Fax: +46 16 4716424; E-mail: Laura.Parducci@ebc.uu.se

© 2005 Blackwell Publishing Ltd

1996; Dumolin-Lapegue *et al*. 1998; Deguilloux *et al*. 2002; Tani *et al*. 2003; Ziegenhagen *et al*. 2003). Recently, it has been shown that aDNA, particularly from plants, may even be indirectly retrieved from secondary samples such as fossil faeces (Poinar *et al*. 1998, 2001) or permafrost soils (Willerslev *et al*. 2003).

A crucial prerequisite for conclusive population-level studies with aDNA is the number of samples available and an extant DNA framework based on broad biogeographical studies, within which to locate the ancient samples. The sampling locations must also be carefully selected to find well-preserved study material.

We have focused on fossil pollen grains of vascular plants because these are the most abundant and among the best preserved remains for many groups of plants (Faegri & Iversen 1989). In particular, we have focused on pollen retrieved from Holocene lake sediments (up to 10 000 BP) because material from this period offers several advantages for aDNA studies. In Holocene lakes, pollen grains are found in high concentrations (> 100 000 grains/cm^3) and, due to the anaerobic conditions in which they have been maintained, they are also well preserved. Furthermore, because of their relatively young age, pollen grains are less affected by diagenesis and therefore more likely preserve DNA molecules. Another advantage conferred by this relatively young age for aDNA studies is that, by using appropriate molecular markers, lineages of Holocene specimens can be traced directly to living specimens. Thus extant sequences can easily be compared with the ancient ones and links can be made with extant population biogeography making a framework for the aDNA work. Finally, the fine degree of chronological precision that can be obtained in the Holocene enables the establishment of a detailed timescale and the application of dating methods, such as radiocarbon, can in some cases provide a high degree of accuracy requiring small amounts of material.

In principle, the analysis of pollen DNA from different ages and at different sites makes it possible to assess directly the history and the dynamics of ancient plant taxa both in space and time and to understand the microevolutionary processes that occurred over the recent past.

Pollen contains haploid DNA and is the means of its dispersal to the haploid DNA of the ovules. The outer walls of pollen grains are extremely resistant to chemical and physical attack and may, in suitable environments such as the accumulating anaerobic sediments at the bottom of small lakes, be preserved indefinitely. In flowering plants, pollen, when shed from the flower, consists of either two or three haploid cells. Both types of pollen possess a large vegetative cell enclosing, either a single or two generative cells. The vegetative cells comprise the bulk of the pollen grain cytoplasm, including numerous plastids and mitochondria, and are responsible for the development of the pollen tube and the delivery of the generative cells to the embryo sac together with the nuclear haploid DNA. Organellar DNA, present in plastids and mitochondria, is therefore present in multiple copies in the vegetative cells of pollen of all flowering plants. Nagata *et al*. (1999) showed that during pollen maturation, in accordance with the mode of inheritance of the organelles (paternal or maternal), there is a selective increase or decrease of the amount of organellar DNA in the mature generative cells. As a result, pollen grains of species with paternal inheritance of plastid DNA (like the majority of conifers) are particularly rich of organellar DNA at maturation.

DNA amplification of plastid DNA regions from modern pollen is demonstrated in species with maternal inheritance of plastid DNA, such as *Hordeum vulgare* and *Secale triticum* (Petersen *et al*. 1996); presumably from the plastids present in the vegetative cell. DNA amplification from fossil pollen has also been demonstrated. In an early study Suyama *et al*. (1996) succeeded in amplifying a short region of chloroplast DNA (cpDNA) from four pollen grains of *Abies* spp. older than 100 000 years BP collected from a Quaternary peat at Kurota Lowland, Fukui, Japan. Experience with this and other unsuccessful attempts to amplify DNA from old pollen material of conifer species enabled us to design a protocol that would maximize the chance of success in Scots pine.

Materials and methods

Extant material

Fresh needles from 30 Scots pine (*Pinus sylvestris* L.) adult trees were collected from an extant population located around Holtjärnen, a small lake situated in central Sweden (60°39′N, 15°56′E). We avoided working with fresh pollen of Scots pine because it was a potential source of contamination through air and clothes in the successive aDNA analyses. Needles were stored at −20 °C until use and DNA was extracted using DNeasy Plant Mini Kit (QIAGEN) and directly used for polymerase chain reaction (PCR).

DNA extraction was performed in an isolated room dedicated to extant DNA work (see *PCR contamination*) after the analyses on ancient material were completed.

Ancient material

We choose to work with ancient material from lake sediments with optimal continuous preservation conditions and with a pollen type that was abundant and easily identifiable from other pollen types present in the sediment. Pollen grains of Scots pine and Norway spruce (*Picea abies* L.) were retrieved from a 2.5-m-long Holocene (0–11 000 BP) core retrieved from Holtjärnen.

The pollen grain of *P. sylvestris* consists of a body about 50 μm long with two laterally placed large bladders. The

© 2005 Blackwell Publishing Ltd, *Molecular Ecology*, 10.1111/j.1365-294X.2005.02644.x

pollen grain of *Pi. abies* is easily distinguished from pine by its larger body size, commonly > 75 µm, and smooth transition between bladders and body.

The core was analysed for fossil pollen and macro remains, and an age depth model was constructed based on six radiocarbon determinations of terrestrial macro remains (Giesecke & Bennett 2004). All the laboratory and field analyses, including drilling the core, were carried out during winter, to avoid periods of pollen release, using rigorous methods to avoid contamination (see *PCR contamination*). The initial PCR experimentation on fresh pollen of different *Abies* species, as well as the preliminary screening of primers on extant DNA was performed in Japan.

Small sediment samples (5 g wet weight) were removed with a sterile scalpel in a clean-air room from two positions in the core at 7.00 and 9.45 m depth (100 years and 10 000 years old, respectively). The upper 2 cm of the surface sediment were discharged to avoid contamination during sampling, and the samples were stored in sterile plastic bags in a refrigerator (+5 °C). Subsamples were sent to Japan for independent replication of the results. Approximately 0.5 g (wet weight) of sediment were dissolved in sterile distilled water, sequentially sieved between 180 and 40 µm, and stored at +5 °C for no more than a week. Between 10 to 15 drops of this solution were deposited onto disposable glass slides for evaluation through microscope (Olympus BH 2) with magnifications ×40 and ×100. A total of 10 pollen grains were selected each day and transferred by mouth pipetting on a new slide using specifically modified glass micropipettes. Briefly, the technique involved the heating of a standard glass pasteur pipette, stretching it out and then quickly breaking off the tip to obtain a sharp end. In this way we obtained pipettes with a diameter of around 150 µm that were successively sterilized before use. Each grain was repeatedly washed in sterile distilled water, transferred to new slides and moved to a spatially isolated area of the laboratory devoted to the PCR analysis of low-copy-DNA. Here, through a Leica Wild M3C microscope (magnification ×6.4 and ×10) each grain was again washed in 15 different drops of sterile distilled water and transferred with 1 µL of water to a DNA-free PCR tube. When possible, i.e. when the grain was sufficiently large and surrounded by less than 1 µL of water, the grain was directly crushed in the PCR tube using sterile plastic pipette tips. For each grain, contamination by extant DNA was monitored using a PCR blank that included all PCR reagents and 1 µL from the last drop of water used for washing the grain.

Choice of the marker

We choose to amplify microsatellite regions from the plastid DNA. The chances of recovering plastid DNA from ancient grains are in fact greater than recovering nuclear DNA because plastids are present in multiple copies in the vegetative cells of the grains and because of the extra protection of the plastid membrane. Plastid DNA has also the advantage of being more conserved in its priming sites and the availability of a huge database makes it easy to design primers and chose appropriate DNA regions for different hypothesis testing. Because the rate of nucleotide substitution of plastid DNA is slow we searched for regions of the genome that were evolving fast enough to show variation within the short time frame of Holocene. The search was facilitated by the availability of the sequence of the entire chloroplast genome of *Pinus thunbergii* (Wagasuki *et al*. 1994), in which several microsatellite regions have been already detected (Powell *et al*. 1995).

Amplification, cloning and sequencing

Previous studies showed that aDNA is often degraded to an extent that amplification of fragments longer than 500 base pairs (bp) is very difficult to achieve (Pääbo 1989). Initially we used different primers and tested amplifications of different lengths on pollen of Scots pine from 10 000 BP and results showed that the longest fragment that we were able to amplify was around 200 bp. Therefore, we choose to amplify two short, yet variable, DNA regions of the plastid genome (220 and 105 bp) from ancient pollen. These two DNA regions included two variable microsatellite regions (Vendramin *et al*. 1996) composed of mononucleotide repeats of 2–14 bp in length (T/A/G and C/T stretches, in the two fragments, respectively). From the extant material we amplified a longer region of 770 bp that included the whole 220-bp fragment (Fig. 1). The five primer pairs, designed based on the *P. thunbergii* cpDNA sequence (Wagasuki *et al*. 1994) were the following: A_f 5′-TCGCGATTACATAA-CCAGATG-3′ and A_r 5′-GACCCTGTTGTTCCGTTCTC-3′ (770 bp); B_f 5′-GCTTATGGCATTGTTGATGT-3′ and B_r

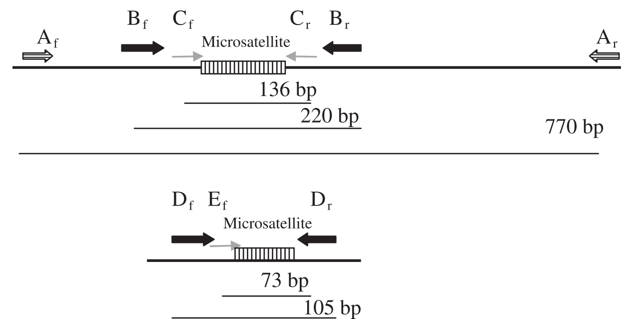

Fig. 1 Schematic representation of the two plastid DNA regions analysed in fossil and modern samples of *Pinus sylvestris* and *Picea abies*. See Materials and methods for names and DNA sequences of the primers.

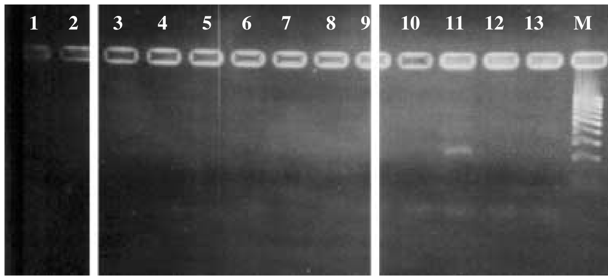

Fig. 2 Agarose gel showing a 220-bp-long amplification product obtained from a 100-year-old pollen grain (lane 11) using primer pair B_f/B_r. *M*: molecular size marker (Gene Ruler 100-bp DNA Ladder, MBI Fermentas). Lanes 3, 5, 7, 9, 11 and 13: amplifications from six different pollen grains. Lanes 2, 4, 6, 8, 10 and 12: corresponding negative controls. Lane 1: negative control with all PCR reagents.

5′-TGGGCATTCTAGCTGTATTG-3′ (220 bp); C_f 5′-GCGGA-AGGTTGATCTCATAGC-3′ and C_r 5′-GTATGCGGAA-TCAACTGGTTC-3′ (136 bp); D_f 5′-CTTGGATGGAATA-GCAGCC -3′ and Dr 5′-GGAAGCGCATTAAGGTCATTA-3′ (105 bp); E_f 5′-GACCCAACTCAGTAAATCC-TC-3′ and D_n 5′-GGAAGCGCATTAAGGTCATTA-3′ (73 bp).

In pollen, primary amplifications were carried out using primer pairs B_f/B_r for the 220-bp fragment and primer pairs D_f/D_r for the 105-bp fragment (Fig. 1). The PCR mixture of 15 μL included 15 mM Tris-HCl (pH 8.0), 50 mM KCl, 1.5 mM $MgCl_2$, 0.1 mM each dNTPs, 0.25 μM each primers and 0.25 unit/mL of Ampli*Taq* Gold (Applied Biosystems). We used GeneAmp PCR system 2400 and 9600 (PE Applied Biosystems) with the following cycling conditions: 94 °C for 10 min (hotstart), 45 cycles of 94 °C for 20 s, 55 °C for 30 s, and a final step of 72 °C for 7 min. After PCR, 4 μL of the product reaction was electrophoresed in 2% agarose gel, stained with ethidium bromide and visualized by UV transillumination. From the PCR products that were visible as single bands on agarose gel (Fig. 2), 10 μL were vacuum dried, resuspended in 3 μL of sterile distilled water. Direct ligation of primary amplification products succeeded only with one fragment (sequence 03–39). The rest of the primary PCR products were used as a template, in the second stage of the nested amplification, without further purification. Secondary amplifications were performed with primer pairs C_f/C_r and E_f/D_r in 50-μL amplification reactions [15 mM Tris-HCl (pH 8.0), 50 mM KCl, 1.5 mM $MgCl_2$, 0.1 mM each dNTPs, 0.25 μM each primers and 0.25 unit/mL of Ampli*Taq* Gold (Applied Biosystems)] performing 25 cycles as in the primary amplifications reactions. Secondary PCR products were purified using the GFX PCR DNA Purification Kit (Amersham Biosciences) and 4 to 8 μL used for direct dideoxy sequencing using primer pairs C_f/C_r and E_f/D_r and PRISM Dye Terminator Kit (Applied Biosystems). Each sequence was read from both strands.

In addition, 3 μL of the purified secondary products were ligated into a pGEM-T vector (Promega) in the presence of 3 units/μL of T4 DNA ligase at 16 °C for 12 h. *Escherichia coli* DH5α cells were transformed by electroporation using 5 μL of the ligation and grown in 1 mL SOC medium for 90 min at 37 °C. The cells were pelleted by centrifugation at 1000× *g* for 10 min, resuspended in 200 μL of SOC medium before plating on selective IPTG/X-gal agar plates. White colonies were screened in 10-μL PCRs using M13 universal primers (procedures as in secondary amplifications). After 1 min at 95 °C, 25 cycles of PCR were carried out (20 s at 95 °C, 20 s at 50 °C, 1 min and 30 s at 72 °C) and clones with inserts of expected size were screened by agarose gel electrophoresis. Cells from colonies that contained inserts of the correct size were grown in LB medium at 37 °C between 12 and 16 h and plasmid DNA was extracted using QIAprep Spin Miniprep Kit (QIAGEN). Seven microlitres of DNA were used for direct dideoxy sequencing with M13 universal primers using PRISM Dye Terminator Kit (Applied Biosystems).

Sequencing reactions were cleaned for dye-terminator removal with Dye Ex 2.0 Spin Kit (QIAGEN) and resolved using a laser-induced fluorescence capillary electrophoresis system (ABI PRISM 310 Genetic Analyser, PE Applied Biosystems). Alternatively, sequencing reactions were also loaded onto 6% denaturing polyacrylamide gels and analysed on an ABI PRISM 377 Genetic Analyser (PE Applied Biosystems).

A minimum of six clones per fragment amplified were aligned and compared with the sequences obtained from direct dideoxy sequencing (forward and reverse sequences) (see Supplementary material). The alignments were searched for ambiguities and substitutions using the software SEQUENCER version 4.1.4 and a consensus sequence was obtained in each alignment.

Amplifications from fresh plant material were conducted in a separated area isolated from those with work involving ancient DNA (see PCR contamination). The reactions were performed on 50-μL volumes using 25 cycles and cycling conditions as in secondary amplifications. The PCR products were purified and DNA sequences were derived by direct dideoxy sequencing (procedures as above). Each sequence was read from both strands.

Sequence analysis of the 220-bp region

Sequences analysis was carried out on the 220-bp fragments obtained from samples from three geological periods. Consensus sequences of the internal microsatellite regions were used to construct a minimum-spanning network using the parsimony criterion (Templeton *et al*. 1992) as implemented in TCS alpha, version 1.0 (Clement *et al*. 2000).

Table 1 Success rate of amplification for the 220-bp and the 105-bp fragments in 100-year-old and 10 000-year-old pollen of *Pinus sylvestris* and in 100-year-old pollen of *Picea abies*

220 bp Sample age	Specimen	Sequences/no. of samples	Success rate (%)	Haplotypes/no. of sequences	Unique haplotypes
100 BP	*P. sylvestris*	7/500	1.4	4/7	1
10 000 BP	*P. sylvestris*	4/301	1.3	3/4	1
Total		**11/801**	**1.3**		
Modern	*P. sylvestris*	30/30*	100	12/30	10
100 BP	*Pi. abies*	2/100	2.0	2/2	

105 bp Sample age	Specimen	Sequences/no. of samples	Success rate (%)	Haplotypes/no. of sequences	Laboratory
100 BP	*P. sylvestris*	1/120	1.7	1/1	Sweden
100 BP	*P. sylvestris*	3/114		3/3	Japan
10 000 BP	*P. sylvestris*	2/120	1.2	2/2	Sweden
10 000 BP	*P. sylvestris*	3/288		2/3	Japan
Total		**9/642**	**1.4**		

*DNA sequences obtained from needles.

The network collapsed identical sequences into haplotypes and calculated their frequencies in the sample. An absolute distance matrix was then calculated for all pairwise comparisons of haplotypes and the probability of parsimony was calculated for pairwise differences until the probability exceeded 0.95. The number of mutational differences between haplotypes associated with the probability just before this 95% cut-off was then the maximum number of mutational connections between haplotypes justified by the parsimony criterion.

Results

Amplification and sequence variation

Over 1400 fossil pollen grains from Scots pine were analysed in Sweden and Japan from sediments from two geologic periods (10 000 and 100 years ago) and we obtained a total of 20 positive amplifications (14 for the 220-BP fragment and 9 for the 105-BP fragment). In addition, we examined also 100 pollen grains of Norway spruce retrieved from the 100-year-old sediment (this pollen type was absent in older positions in the core (Giesecke & Bennett 2004) and obtained two positive amplifications (Table 1). Three amplifications obtained from pollen of Scots pine (100 000-year old) appeared after electrophoresis as fragments of dubious size and as expected clones from these fragments yielded sequences of unknown origin. These amplifications were therefore discarded from the successive analyses leaving only seven positive products from the 100 BP period. The remaining PCR products yielded only clones with sequences from Scots pine or Norway spruce and no amplifications were found in the corresponding negative controls.

In the Supplementary material we present the DNA sequences obtained from ancient pollen in this study and we report all clones and corresponding consensus DNA sequences for the internal variable region of the 220-bp fragments. We found a variable number of T/A/G and C/T repeats in the internal microsatellite regions among Scots pine and Norway spruce sequences. On the other hand, sequences flanking the microsatellites were highly homologous in all fragments, i.e. very few substitutions and no insertion/deletions were detected. This confirms the high level of conservation of the chloroplast genome of conifers and the usefulness of plastid microsatellites for detecting variation in this genome. Only in two cases there was disagreement between the sequence scored from the clones and the two obtained from direct sequencing (01–48 and 27–36 in Supplementary material). In general, the amplification products were composed of two classes of sequences. One included the majority of clones with identical sequence (from which the consensus sequence was inferred) and a minor class with between one and three sequences with a number of repeats differing from the consensus sequence.

Different factors influenced the rate of success during amplification from pollen of the two periods. The length of the amplified fragment was one of them. Our initial tests on 100 000-year old Scots pine pollen revealed that the maximum amplifiable length was around 200 bp. PCR products significantly longer than 200 bp were too difficult to obtain from ancient pollen and enhanced the chance of representing contaminating DNA. Nevertheless, we found

Table 2 Haplotype distribution based on the 220-bp fragments obtained from ancient and extant samples of *Pinus sylvestris* (haplotypes A–P) and *Picea abies* (haplotypes Q and R) and on the105-bp fragment obtained from ancient pollen of *P. sylvestris* (haplotypes S–W). The DNA sequences and the length refer to the internal variable region only (microsatellite)

	DNA sequence									No. of sequences			
Haplotypes	T	C	T	A	G	A	G	A	Length	Modern‡	100 years ago	10 ka	*N*
A	2	2	3	11	12	0	0	2	32	3			3
B	2	2	4	10	12	0	0	2	32	3			3
C	2	2	4	10	13	0	0	2	33	3			3
D	2	2	3	11	13	0	0	2	33	3			3
E	2	2	4	10	10	0	0	2	30	3			3
F	2	2	3	10	14	0	0	2	33	1			1
G	2	2	2	12	12	0	0	2	32	1			1
H	2	2	3	11	10	0	0	2	30	2			2
I	2	2	4	10	11	0	0	2	31	3	4		7
J	2	2	4	11	10	0	0	2	31	2			2
K	2	2	3	10	13	0	0	2	32	1			1
L	2	2	3	10	12	0	0	2	31	5	1	2	8
M	2	2	2	10	12	0	0	2	30		1		1
N	2	2	2	10	11	0	0	2	29		1		2
O	2	2	2	9	10	0	0	2	27			1	1
P	2	2	3	10	9	1	2	2	31			1	1
Total										**30**	**7**	**4**	**41**
Q	5	1	2	11	6	3	3	3	34		1		1
R	5	1	2	11	5	3	3	3	33		1		1
Total											**2**		2
	C	T	A	T									
S	10	10	1	9					30		1* + 1†		2
T	10	11	1	9					31		1†		1
U	12	11	1	9					33		1†		1
V	10	7	1	9					27			1*	1
W	10	8	1	9					28			1*	1
X	10	9	1	8					28			2†	2
Y	10	10	1	9					30			1†	1
Total											**4**	**5**	**9**

DNA sequences obtained in *Sweden and in †Japan; ‡DNA sequences obtained from needles.

that the rate of success during amplifications from pollen from the earlier period was not particularly higher when we changed from the 220-bp to the 105-bp fragment, increasing only from 1.4% to 1.7%. Similarly, with pollen from the oldest period (10 000 BP), higher success rates associated with amplification of shorter fragments were not specifically observed. On the other hand we observed that the best results were obtained when pollen grains were successfully crushed prior to amplification. This may have been the cause of higher success rates obtained with pollen from spruce (2%) which was larger and easier to crush in the PCR tubes.

Comparison of ancient and modern sequences

Sequence comparison among all ancient and extant specimens of Scots pine and Norway spruce gave a total number of 25 haplotypes. Their distribution among all samples analysed is presented in Table 2. In particular, sequence comparison of the ancient Scots pine sequences obtained from the two geologic periods (seven from 100 BP and three from 10 000 BP) with modern homologous DNA regions obtained from 30 extant individuals, revealed 16 different haplotypes (haplotypes A–P in Table 2). The genetic relationships found among these 16 haplotypes are visualized in the minimum-spanning network presented in Fig. 3. In the network the haplotypes occupy nodes and each branch represents a single nucleotide change detected in the microsatellite region.

Discussion

DNA preservation in the fossil pollen

Here we report genetic analyses of fossil pollen retrieved from postglacial lake sediment in Sweden and show that plastid

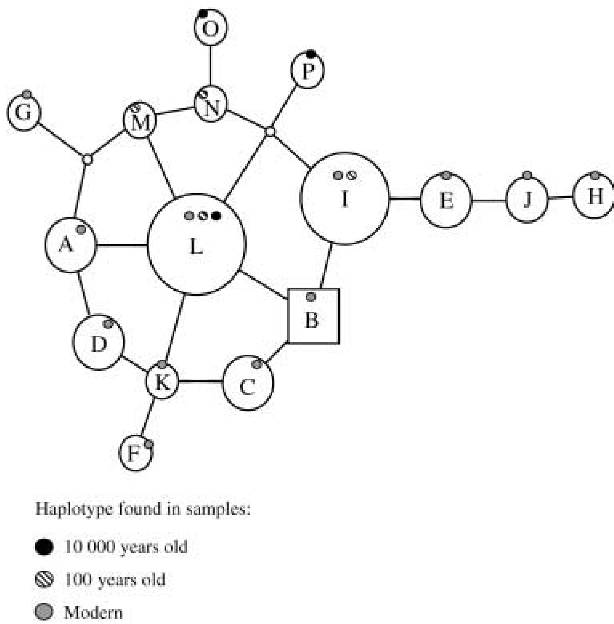

Fig. 3 Minimum-spanning network constructed with the 220-bp cpDNA fragment showing the relationships among the 16 haplotypes found in fossil and modern specimens of Scots pine. Each line in the network represents a single mutational change between haplotypes. The small open circle represents a haplotype not present in the sample but inferred to be intermediate between the two closest haplotypes. The sizes of the circles are proportional to the haplotype frequencies. The haplotype in a square has the biggest outgroup weight.

DNA is recoverable and usable from pollen grains from 10 000 BP. Sample age per se is not unambiguously indicative of the preservational state of an ancient tissue; it is rather the sample's original condition and its environment during early storage that seem to have the most significant effects on DNA preservation (Golenberg 1991; Yang 1997). Accordingly, we obtained similar success rates of amplification in pollen of different ages and in general, better amplifications were obtained with short fragments (< 200 bp) while longer fragments were too difficult to obtain and increased the chance of representing contaminating DNA. The procedure of crushing the grains also appeared to be indispensable if positive results are desired from a large numbers of samples.

PCR contamination

Ancient DNA is extremely susceptible to contamination with extraneous extant DNA, and studies must therefore demonstrate appropriate authentication procedures (Cooper & Poinar 2000). The low number of preserved DNA molecules present in the ancient sample usually poses a serious problem because any single intact contaminant DNA molecule may out-compete the degraded ancient molecules (Pääbo 1989). Another factor complicating amplification from aDNA sequences is associated with a PCR beginning from a small number of DNA molecules, meaning that if nucleotide misincorporations due to diagenetic damage in aDNA occur during the initial PCR cycles they will be represented in a large fraction of the molecules in the final amplification product (Höss *et al*. 1996).

In this study we followed several criteria in order to control DNA contamination from both extant and ancient DNA.

1. The primers used worked specifically on the plastid DNA of Pinaceae species and the sequences of the two amplified regions could tell apart different species within this family (i.e. *Pinus sylvestris* and *Picea abies*). If erroneous amplifications started from DNA present in the sediments from species other than *P. sylvestris* and *Pi. abies*, one would be alerted to the problem because either they would yield distinct sequences (DNA from closely related species), or no sequences at all (DNA form distant species).
2. To further exclude the possibility that the results were due to amplification from extant DNA from *P. sylvestris* and *Pi. abies* or from PCR amplicons, we worked in physically isolated areas. Manipulations of fossil samples prior to PCR were performed in a separated room isolated from the one with work involving modern DNA. Handling of pollen grains and PCRs were set up in a second area devoted to the preparation of low-copy-DNA PCR where extant DNA and/or PCR products were never introduced. Here, laboratory equipment and reagents were regularly sterilized and dedicated protective clothing was worn. Finally, a third room was dedicated to PCR work involving extant DNA and to sequencing work. We also used negative controls throughout our experimentation (during first and second round of PCR) and when positive amplification was obtained from pollen, the corresponding amplification control was also screened, i.e reamplified following the same procedures as with pollen. No PCR products were obtained from negative controls. Positive controls were avoided because they provided a high contamination risk.
3. To investigate if the variation seen in the ancient sequences was due to PCR misincorporations and to reliably count the number of nucleotide stretches in the regions analysed, the PCR products were cloned in a plasmid vector. Clones were aligned and compared to the two sequences obtained from direct dideoxy sequencing (forward and reverse). Nucleotide changes present in only few clones were likely to be due to misincorporations by the DNA polymerase during PCR and therefore such sequences were not considered in the analysis. Each alignment was searched for ambiguities and substitutions and

consensus sequences were used in the successive analyses.

4 Finally to test whether DNA amplification from pollen was reproducible in another laboratory, and to investigate whether the results were due to laboratory-specific artefacts and/or contamination, independent analyses of the 100 BP fragment were repeated using the same procedures in a second laboratory in Japan on pollen isolated from the same sediments. Amplifications and sequence analyses were performed by different persons in the two laboratories.

Authenticity of the ancient sequences and consistency of the results

Authentication of the ancient sequences and the consistency of the results came from different lines of evidence.

Analysis of pollen from different species. The method was initially designed for working with pollen of Scots pine and was successively tested also on pollen from Norway spruce. Amplifications from both pollen types gave DNA sequences that discriminated between the two taxa. This showed the consistency of the method and its potentiality on conifer species with characteristics similar to Scots pine and Norway spruce (i.e. with paternal inheritance of plastid DNA).

Independent reproduction of the results. The achievement of comparable result from analogous remains from the same sediment (see above) was also a valuable method to verify the authenticity of the ancient sequences. It is unlikely in fact that the amplifications obtained in this study started from contaminating modern DNA of the two respective species.

Moreover, in the duplication experiment with the 105-bp fragment, different persons independently obtained sequences of Scots pine in the two different laboratories in Japan and Sweden, using pollen grains isolated from the same sediment samples (100 and 10 000 BP) and following identical procedures. Clearly, because each PCR used single pollen grains, this reproduction of results could not be 'true replication' because the amplifications obtained in the two laboratories were not from the same pollen grain and therefore may not necessarily have had the same DNA sequence. Nevertheless, the PCR products obtained in the two different laboratories yielded sequences of Scots pine demonstrating that results were reproducible and not due laboratory-specific artefacts or contamination.

Analysis of the mutational patterns of the analysed fragments. Knowledge of the mutational patterns exhibited in a specific genome can also aid determination of authenticity of ancient sequences. If one understands the mutation pattern for the analysed DNA fragment, it should be possible to discern random errors from natural substitutions. In our study, the number of repeated nucleotides in the microsatellite varied in increments or decrements of one (stepwise-mutation model, SSM), as generally observed in plastid DNA microsatellites. This should be interpreted as though the variability observed in the microsatellite was due to natural substitutions rather than the misincorporation of nucleotides during PCR. In addition, the results showed that the overall size of the microsatellite region increased through time, from the oldest (10 000 BP) to the younger (100 BP and modern) populations (Table 1). The tendency for new mutations to cause an increase in allele size is a well-documented type of bias observed in microsatellite mutation. This phenomenon has been reported both in nuclear (Karhu *et al*. 2000; Vigouroux *et al*. 2003) and in chloroplast microsatellites (Petit *et al*., unpublished). In addition, post-mortem stability has also been documented for nuclear microsatellites showing that, in humans, the size of these repeated sequences remains stable after tissue decomposition (Hoff-Olsen *et al*. 2001).

Haplotype analysis. One clear advantage of working with young material (up to 100 000 BP) is that homologous DNA sequence information from modern specimens is available. Homologous sequences from extant organisms can then be compared with ancient sequences and the comparison should make phylogenetic sense. Such a comparison is essential in the context of any aDNA studies to know whether the PCR results are from a real plant/animal rather from some associated and as yet unsequenced microbe present in the ancient remains.

The network presented in Fig. 3 reveals a genetic continuity through time in the three Scots pine populations from Holtjärnen. It contains a small number of common haplotypes surrounded by others that are at one- or two-point mutations distant. There is one most common haplotype (haplotype L) that is present in all three periods: 10 000 BP, 100 BP, and present. The high frequency of another common haplotype (haplotype I) suggest that it may also have persisted for a long time near Holtjärnen. These two haplotypes have also more mutational connections in the network, suggesting that they may have given rise to the other haplotypes in the network. The analysis of 10 000-year old pollen reveals also the presence of two unique haplotypes (haplotype O and P). Although only four sequences were available from this geological period, their absence within the larger group of modern and 100-year-old samples implies that they probably went extinct sometime after 10 000 BP at Holtjärn. Haplotype P is particularly interesting because is the most divergent (sequence 03/39 in Supplementary material). The uniqueness of these two ancient haplotypes and their grouping in the network with related younger living specimens suggest the possibility

that they both were ancestors of the living Scots pine individuals.

Molecular behaviour during amplification. An inverse correlation between amplification efficiency and length of the amplification is a very simple indicator of the extent of degradation present in an aDNA template (Pääbo *et al*. 2004). In this study, shorter fragments were more readily amplified than longer ones while we did not succeed in amplifying fragments much longer than 220 bp.

We have shown that aDNA information can be retrieved from sediment fossil pollen as old as 10 000 years. This will allow us to obtain direct estimates of the genetic changes that occurred between populations of different ages of both Scots pine and Norway spruce. The method has been tested on two conifer species that show paternal inheritance of plastid DNA and should also be tested on species showing maternal inheritance of this genome (like the majority of angiosperms) to verify its general applicability in plants. Nevertheless it has major implications in palaeoecology and offers a number of potential uses for studying the ecology and the evolution of populations of the two species studied here as well as of related species of the Pinaceae family. There is an excellent record of the distribution and abundance of conifers through the identification of fossil pollen in sequences of peat and lake sediments (Berglund *et al*. 1996). In addition, the resolution in such records is such that it is often not possible to make morphologically based identifications below the generic level. By choosing adequately variable plastid DNA regions, it should now be possible to relate DNA from ancient and extant conifer specimens and distinguish taxa for which morphological identification is not possible or difficult.

We speculate that it should even be possible to analyse multiple regions at the same time in one pollen grain using multiple primers during the same PCR. In this way information from multiple plastid regions can provide a more detailed record of the tempo and mode of genetic change within populations from different ages.

The method presented here offers the possibility of directly observing mutation patterns and evolutionary rate changes of these fast-evolving DNA regions on a geological timescale; an opportunity open only when a sufficient number and distribution of fossil samples is available for analysis. A recent study of nucleotide evolution in the mitochondrial DNA from fossil bones of Adélie penguins from Antarctica is a good example of such a study in an animal group (Lambert *et al*. 2002): the possibility is now open for such work with conifers.

Acknowledgements

We thank, Linda Albinsson for modern Scots pine sequences, Rémy J. Petit for critically reading earlier version of the manuscript, Jacob Höglund for sequencing facilities, Gunnilla Engström for assistance during sequencing and Anita Matsson for valuable advices for pollen manipulation. The work was supported by grants from the Swedish Research Council (grants G507-266/2001 and G5103-20005375/2000; LP, ML and KDB), the Swedish Foundation for International Cooperation in Research and Higher Education (STINT) (LP and YS), the Carl Trygger Foundation (LP) and from grants-in-aid for Exploratory Research (No. 15658047) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (YS).

Supplementary material

The supplementary material is available from http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC2644/MEC2644sm.htm

Table S1 The DNA sequences derived from amplifications of the 220-BP and 150-BP plastid fragments from fossil pollen of two geologic periods (100 and 10 000 BP). Only the internal regions including the microsatellite are presented. Designations in the left column consist of two letters (PS: *Pinus sylvestris*, PA: *Picea abies*, PT: *Pinus thunbergii*) indicating the species followed by a number indicating the PCR number. Amplifications of the 105-BP fragment were performed at the university in Sweden (S) and repeated at the university in Japan (J). The homologous *Pinus thunbergii* sequence is given above as a reference (accession number D17510).

Table S2 The DNA sequences of clones derived from amplifications of the 220-BP fragment amplified in 13 *P. sylvestris* pollen grains (seven from 100 BP and four from 10 000 BP) and in two *P. abies* pollen grains (100 BP). Only the internal regions including the microsatellite are presented (circa 124 BP). Asterisks indicate sequences obtained from direct sequencing using forward and reverse primers. Numbers in the left column represent PCR number and letters indicate clone number. Below the clones the inferred consensus sequence is given for each alignment.

References

Berglund BE, Birks HJ, Ralska-Jasiewiczowa M, Wright HE (1996) *Palaeoecological Events During the Last 15 000 Years: Regional Syntheses of Palaeoecological Studies of Lakes and Mires in Europe*, 4th edn. Wiley, Chichester.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

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

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature*, **409**, 704–707.

Cwynar LC, MacDonald GM (1987) Geographical variation of lodgepole pine in relation to population history. *American Naturalist*, **129**, 463–469.

Deguilloux M-F, Pemonge MH, Petit RJ (2002) Novel perspectives in wood certification and forensics: dry wood as a source of DNA. *Proceedings of the Royal Society of London. SeriesB, Biological Sciences*, **269**, 1039–1046.

Dumolin-Lapegue S, Pemonge MH, Gielly L, Taberlet P, Petit RJ (1998) Amplification of oak DNA from ancient and modern wood. *Molecular Ecology*, **8**, 2137–2140.

Endicott P, Gilbert MTP, Stringer C *et al.* (2003) The genetic origins of the Andaman Islanders. *American Journal of Human Genetics*, **72**, 178–184.

Faegri K, Iversen J (1989) *Textbook of Pollen Analysis*. Wiley, Chichester.

Giesecke T, Bennett KDB (2004) The Holocene spread of *Picea abies* (L.) Karst. in Fennoscandia and adjacent areas. *Journal of Biogeography*, **31**, 1523–1548.

Golenberg EM (1991) Amplification and analysis of Miocene plant fossil DNA. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **333**, 419–427.

Hein J, Schierup MH, Wiuf C (2005) *Gene genealogies, variation and evolution*. A Primer in Coalescent Theory. Oxford University Press, Oxford and New York.

Hoff-Olsen P, Jacobsen S, Megåv B, Olaisen B (2001) Microsatellite stability in human post-mortem tissues. *Forensic Science International*, **119**, 273–278.

Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, Pääbo S (1996) DNA damage and DNA sequence retrieval from ancient tissue. *Nucleic Acids Resource*, **24** (7), 1304–1307.

Karhu A, Dieterich JH, Savolainen O (2000) Rapid expansion of microsatellite sequences in pine. *Molecular Biology and Evolution*, **17**, 259–265.

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S (1997) Neanderthal DNA sequences and the origin of modern human. *Cell*, **90**, 19–30.

Lambert DM, Ritchie PA, Millar CD, Holland B, Drummond AJ, Baroni C (2002) Rates of evolution in ancient DNA from Adelie penguins. *Science*, **295**, 2270–2273.

Leonard JA, Wayne RK, Cooper A (2000) Population genetics of ice age brown bears. *Proceedings of the National Academy of Sciences, USA*, **97**, 1651–1654.

Nagata N, Saito C, Sakai A, Kuroiwa H, Kuroiwa T (1999) The selective increase or decrease of organellar DNA in generative cells just after pollen mitosis one controls cytoplasmic inheritance. *Planta*, **209**, 53–65.

Pääbo S (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences, USA*, **86**, 1939–1943.

Pääbo S, Poinar H, Serre D, Jaenicke-Despr'es V *et al.* (2004) Genetic analyses from ancient DNA. *Annual Reviews of Genetics* 10.1146/annurev.genet. 37.110801.143214.

Petersen G, Johansen B, Seberg O (1996) PCR and sequencing from a single pollen grain. *Plant Molecular Biology*, **31**, 189–191.

Petit RJ, Brewer S, Bordacs S *et al.* (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management*, **156**, 49–74.

Petit RJ, Aguinagalde I, de Beaulieu JL *et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Poinar HN, Hofreiter M, Spaulding WG *et al.* (1998) Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science*, **28**, 402–406.

Poinar HN, Kuch M, Sobolik KD *et al.* (2001) A molecular analysis of dietary diversity for three archaic Native Americans. *Proceedings of the National Academy of Sciences, USA*, **98**, 4317–4322.

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski J (1995) Polymorphic simple sequence regions in chloroplast genomes: Applications to the population genetics of pines. *Proceedings of the National Academy of Sciences, USA*, **99**, 7759–7763.

Shapiro B, Drummond AJ, Rambaut A *et al.* (2004) Rise and fall of the Beringian steppe bison. *Science*, **306**, 1561–1565.

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, Genetics and Systematics*, **71**, 145–149.

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

Templeton AR, Crandall KA, Sing CF (1992) A cladistic–analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA-sequence data. 3. Cladogram estimation. *Genetics*, **132**, 619–633.

Vendramin GG, Lelli L, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. *Molecular Ecology*, **5**, 595–598.

Vigouroux Y, Matsuoka Y, Doebley J (2003) Directional evolution for microsatellite size in maize. *Molecular Biology and Evolution*, **20**, 1480–1483.

Wakasugi T, Tsudzuki SI, Shibata M, Sugiura M (1994) A physical map and clone bank of the black pine (*Pinus thunbergii*) chloroplast genome. *Plant Molecular Biology*, **12**, 227–241.

Willerslev E, Hansen AJ, Binladen J *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**, 791–795.

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, 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.

Laura Parducci is a researcher at Uppsala University working with plant ancient DNA and genetic variation in natural plant populations. Yoshihisa Suyama is a molecular biologist working at Tohoku University interested in ecological genetics, molecular systematics and phylogeny of forest trees species. Martin Lascoux is a population geneticist working at Uppsala University interested in the impact of past demography and selection on gene genealogies of plant species. Keith D. Bennett is professor in Quaternary Geology at Uppsala University interested in Holocene-scale biogeographic patterns, their development, and the significance for these results for macroevolutionary patterns.