PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Bacterial DNA in Clarkia Fossils [and Discussion]

Arend Sidow, Allan C. Wilson, Svante Paabo, S. Hummel, J. L. Bada, P. Westbroek, E. Hagelberg and G. B. Curry

Phil. Trans. R. Soc. Lond. B 1991 **333**, 429-433
doi: 10.1098/rstb.1991.0093

Bacterial DNA in Clarkia fossils

AREND SIDOW[1], ALLAN C. WILSON[1] AND SVANTE PÄÄBO[2]

[1] *401 Barker Hall, Division of Biochemistry and Molecular Biology, University of California, Berkeley, California 94720, U.S.A.*

[2] *Lehrstuhl für Allgemeine Biologie, Institute for Zoology, University of Munich, Luisenstrasse 14, W-8000 Munich 2, F.R.G.*

[No abstract supplied]

1. INTRODUCTION

For the growing number of investigators who study ancient DNA, the most stimulating paper published during 1990 reported the determination of a chloroplast DNA sequence from a plant compression fossil found in a Miocene deposit at Clarkia, Idaho (Golenberg *et al.* 1990). During August 1990, S. P. took part in an excavation at the Clarkia site organized by Dr C. J. Smiley of the Tertiary Research Center, University of Idaho. Analysis of extracts prepared from the plant remains shows that the high molecular mass (HMM) DNA that can be detected in some extracts from Clarkia fossils is mostly, if not exclusively, of bacterial origin. Further work aimed at the retrieval of plant sequences from these remains should therefore not be confined to extracts where HMM DNA is seen.

2. MATERIALS AND METHODS

(*a*) *Nucleic acids extraction*

Blocks of clay were split open and compression fossils in a good state of preservation were identified and taxonomically determined by Dr C. J. Smiley. Plant remains to be extracted were photographed and, within 5–10 min after excavation, ground together with dry ice in an acid-washed mortar. The powder was transferred to a vial containing an extraction buffer (100 mM Tris/Cl, pH = 8.0, 20 mM EDTA, 1.4 M NaCl, β-mercaptoethanol (0.2% by volume), 20 g l^{-1} CTAB). In some cases, an extraction buffer containing 20 g l^{-1} SDS instead of CTAB was used. After agitation the tube was incubated at 60 °C for 30 min. The sample was then stored until further processed in the laboratory by an extraction with chloroform : isoamylalcohol and isopropanol precipitation. The precipitated material was resuspended in 500 µl of double distilled water.

(*b*) *Enzymic amplification*

Primers used for amplification of chloroplast sequences were 5′-CGTTACAAAGGACGATGCTACCACATCGAG-3′ (*Den 1*), 5′-ATCATCACGTAGTAAATCAACAAAGCCTAAAGT-3′ (*Rub 2*) and 5′-GAAGTAAACATGTTAGTAACAGAACC-3′ (*Den 2*). The former two primers are identical in sequence to the ones used by Golenberg *et al.* (1990). Primer *Den 2* yields an amplification product of 122 base pairs (b.p.) in combination with primer *Den 1*. The amplifications were done in 25 µl volume as described (Pääbo *et al.* 1987) by using the amplification conditions described by Golenberg *et al.* (1990) in a Perkin-Elmer Thermal cycler, except that 40 instead of 30 cycles were performed. The 'universal' eubacterial primers RW01 and DG74 were kindly provided by Dr Diane Leong, Roche Diagnostics Research, Alameda, California, U.S.A. These primers show weak priming activity also to plant rRNA genes (S. Pääbo, unpublished observation). The primers NS1 and NS2 for the fungal 18*S* rRNA gene were kindly provided by Dr John Taylor, University of California, Berkeley, U.S.A. Both these primer pairs were used according to specifications provided by the investigators indicated.

(*c*) *Molecular cloning and sequencing*

For cloning experiments, the primers were kinased using standard procedures (Sambrook *et al.* 1990) After enzymic amplification, 1 U of Klenow DNA polymerase (USB Biochemicals) was added to the amplification reactions and incubated for 30 min at 25 °C. Portions of 10 µl were separated in a low melting agarose gel and the relevant bands were isolated. The products were ligated to *Sma*l-cut and dephosphorylated M13 mp10 and transformed into *E. coli* strain DH5a. Preparation of phage particles and of single-stranded DNA were done by using standard protocols (Sambrook *et al.* 1989). Sequencing was done with the Sequenase kit (USB Biochemicals).

(*d*) *Analysis of sequences*

Sequence alignment was done using the algorithm of Hein (1989*a*, *b*) and maximum likelihood analyses were done with the DNAML program of the PHYLIP package (Felsenstein 1981, 1989). The following options were set: F (frequency of bases estimated from the data), T (transitions assumed to be 1.5 times more common than transversions), U (likelihood-ratio test for user-specified trees; (Kishino & Hasegawa, 1989)).

(*e*) *Southern transfer and hybridization*

Nucleic acids separated in agarose gels were depurinated by incubation of the gel in 0.25 M HCl for 10 min followed by denaturation, neutralization and transfer to nylon filters (PAL) according to described

protocols (Sambrook *et al.* 1989). Probes were labelled by random priming according to the instructions of the manufacturer (BRL). Hybridization was done overnight in 6 × SSC, 0.1 % SDS, 50 % deionized formamide, 5 × Denhardt's solution and 100 μg ml^{-1} denatured salmon sperm DNA (Sambrook *et al.* 1989). Filters were washed twice in 2 × SSC, 0.1 % SDS at ambient temperature for 10 min and twice at 55 °C in 0.2 × SSC for 30 min. Radioautograms were exposed overnight.

3. RESULTS

(*a*) *High molecular mass* (HMM) *DNA in extracts*

Golenberg *et al.* (1990) reported that approximately 10 % of the plant extracts prepared from Clarkia fossil plants contain HMM DNA that could be visualized on ethidium-stained agarose gels. These extracts were selected for enzymic amplification with primers specific for plant chloroplast sequences. We found that the observation that DNA is present in a fraction of the extracts can easily be replicated. The amount of DNA extracted from different plant remains varies dramatically. However, it always has an apparent molecular mass of over 10 kilobases and a fraction is totally excluded from the gel matrix (figure 1).

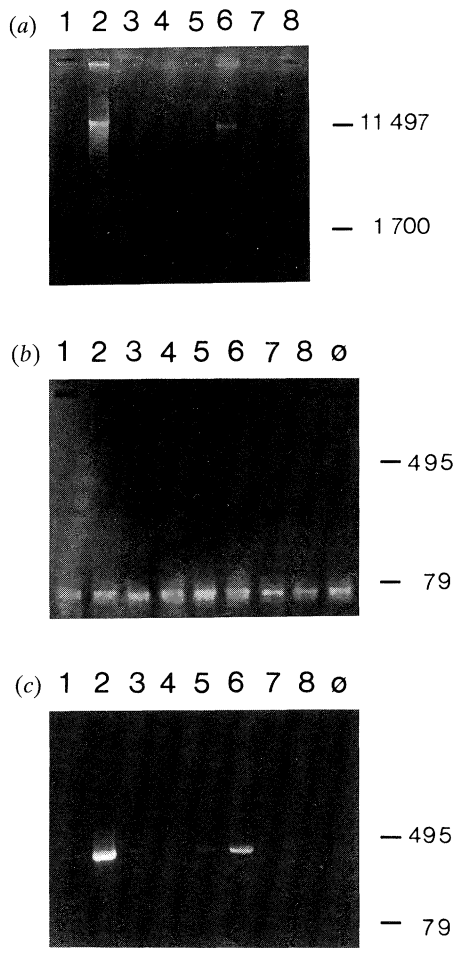


Figure 1. Electrophoretic analysis of extracts from Clarkia fossils and enzymic amplifications with plant and bacterial primers. (*a*) 10 μl of seven extracts from Clarkia samples (lanes 1–7) and one control extract (lane 8) were analysed by electrophoresis through a 0.8 % agarose gel and visualized by staining with ethidium bromide. The extracts were: 1, *Salix*; 2, *Magnolia* or *Persea*; 3, clay next to number 2; 4, *Magnolia*; 5, *Betula*; 6, *Quercus* (acorn); 7, *Quercus*; 8, extraction control. (*b*) Amplifications from the above extracts using primers specific for the chloroplast *rbc*L gene (*Den 1* and *Rub 2*). (*c*) Amplifications from the same extracts using primers designed to amplify a segment of the bacterial 16*S* rRNA gene. Lanes where no extract was added to the amplification reaction are indicated by Ø. Numbers indicate the migration positions of molecular size markers.

(*b*) *Amplifications using chloroplast primers*

We attempted to amplify chloroplast DNA sequences using the published primers (Golenberg *et al.* 1990) that are expected to yield a product of 790 b.p. In addition, a primer designed to yield a product of 122 b.p. in combination with one of the published primers was used. These primers work well on contemporary DNA extracted from *Magnolia* species as well as a monocot (*Zea mays mays*, data not shown). By using various amplification protocols, including one identical to the protocol reported by Golenberg *et al.* (1990) to amplify the Miocene *M. latahensis* sequence in 30 cycles, we have failed to produce any specific amplification products detectable on ethidium-stained gels even after 40 cycles (figure 1). Obviously, this just reflects our inability to do these experiments successfully in the extracts that we have prepared. It does, however, show that the success rate of amplification from Clarkia fossils is not very high.

(*c*) *Amplification using bacterial primers*

As large amounts of DNA can be detected in some of these extracts, we concluded that the HMM DNA could not represent plant DNA that is in a form that allows the amplification of chloroplast sequences. We therefore attempted to amplify DNA sequences from various classes of micro-organisms, using primers to conserved parts of genes for nuclear ribosomal RNA. Primers known to allow amplification from a wide variety of fungal species invariably failed to yield amplification products. However, primers designed to amplify an approximately 370 b.p.-long piece of bacterial ribosomal RNA genes yielded amplification products from the extracts where HMM DNA was present (figure 1). After only 20 cycles, large amounts of product were detected and the amounts of product correlated approximately with the amounts of HMM DNA present in the extracts. In addition, small amounts of product could be detected in many extracts where no DNA could be seen in agarose gels and in some extracts prepared from clay from the same strata where compression fossils were found. After more cycles of amplification, these products were easily detected. Extraction controls and buffer controls were negative.

(*d*) *Sequencing of amplified DNA*

Direct sequencing of the amplification products after the generation of single-stranded DNA by asymmetric PCR showed that several sequences were present in the products. The amplification products from three

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

```
N. tabacum chloroplast   GACGTCAAGTCATCATGCCCCTTATGCCCTTGGCGACACACGTGC-TACAATGGCCGGGACAAAGGGTCGCGATCCCGCGAGGGTGAGCTAACCCCAAAA
Anacystis nidulans       .....................CAT...G...T........A.........CT.C.....GC.A.A.....AG........TGA...A..T.T.CC..
Streptomyces coelicolor  .....................T.T.G...TG.....................T....T.A.CT.....A....A...TG....G..T.T.....
Pseudomonas aeruginosa   ..................G......C.G..AG...T.....................T...T.........T..C.AG.......T.G.......T..T..
Clones 1 and 2           ....C......C.....G........GGTAG...TT.......AA.........T...T...G.....A..C.A.........-G....C..T....G..
Clones 7,9,11,16 and 20  ..........C.....G........GGTAG...TT.......CA.........T...T...G.....T..C.AG........TG....C..T....G..
Clones 8,12,15 and 18    ..........C.....G......C.GG..G...T.....................G.T.....T...AT..T.AGGG....CCC.TC..A..T.T.....
Clones 24 and 27         ..G.....................C.T.TAG...TG.....A............G...T...G....CT..TGGG........CG....C..T...T...
Clones 28,30 and 31      ..........C.....G......C.GG..G...T.....................G.T.....T...A...T.AGGG..A.CCC.TC..A..T.T.....
Clone 17                 ............GT...G......C.A..AG...TG.......A.C.......C..A.T...G.AT.ATC...AT....A...TG...GA..T.T.....
Clone 25                 .....................G......T.T.G...TG.....................T....T....T.....A...T....TG....G..T.T.....
Clone 35                 ..........C.....G......C.G..AG...T........A..........T...T...G.....T....AA........TG....C..T....G..
Clone 37                 A..........C.....G........GGTAG...TT.......AA.........T...T...G.....T..C.A.........-G....C..T....G..

tab  ACCCGTCCTCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATCGCTAGTAATCGCCGGTCAGCCATACGGCGGTGAATTCGTTCCCGGG-CCTTGTACACACCGCCCGT
ana  ...G.GG.........A.....................................G.....................A.......-....T.........A.....................
str  .G...GT..............GG.T...........A.CC.......T....G...............A.A.....AT.G.T.........A.....................
pse  .A...AT.GT...C.....C....T...........A....G.....T.....................TGAA....A-..GTCA........A...C.....................
cl1  .G...AT.GT...C.......G..T...........A..C.......T.....................G.A.....-..GTC.........A.....................
cl7  .G...AT.GT...C.......TTCT...........AGA.......T.....................G.A.....-..GTC.........A...........T.....................
cl8  .G...-T..............GGCT...........AGCC.......TT....................TG.A.....A-CG.CA........A.....................
c24  .-...CT.........A.....................................TT...G.T......T...C.G.A.....AC.G.C.G.......A.....................
c28  .G...-T..............GGCT...........AGCC.......TT....................TG.A.....A-.G.CA........A.....................
c17  ...T.-T......C.......G..T...........A..C.......TT....................TA.A.....A-CG.TA........A.....................
c25  .G...GT..............GG.T...........A.CC.......T....G..T...........A.A.....AG.G.T.........A.....................A
c35  .G...AT.C....C.....C.A..T...........A..T.G.....T....-..............G.A.....T..G.C.........A.....................
c37  .G...AT.GT...C.....C.G..T...........A..C.G.....T.....................G.A.....-..GTC.........A...........G.....................

tab  CACACTATGGGAGCTGGCCATGCCCGAAGTCGTTACCTTAACCG-C--AAGG--GGGGGATGCC-GAAGCGGGGCTAGTGACTGGAGTGAAGTCGTAACAAGGTAGCCGTACTGGAAGG
ana  .....C....A..T.........................C.......TTCGC...A.....GC...GA.G.TA.....GA.......G.....................C......
str  ...GTC.C.AA..TC..TA.CA.......C..G.GG.CC....CCTT.GT..GA...A.C..T.GA.G.T...A..G.C..T...GAC.....................C......
pse  .....C.......TG..TTGCT..A.....A.C..GTC.....................G..AC.G.TA.CACG.A.T.AT.CA.......G.....................GG....CC
cl1  .....C.......TG..TTCCA..A.....A....G.C.....................A...C...TA.CACG.T...AT.CA.......G.....................TC......
cl7  .....C.......TG.AATC....A.....A.G..G.C.....................G...C.C.TA.CACG.T...TT.CA.......G.....................GG....CC
cl8  .....C.......T....TT.A..T....A..G.G.GC.....-.GC...T..GA..CAGCCG.CACG.TA...TC..C......G.....................GG-...CC
c24  ..A..C.C.AA..TC...A.CA.......C..G.GG.C.....-.GC......AA..A.CC.T.GA.G.T....TC.....T...G..T.....................C......
c28  .....C.......T...TTT.A..T....A..G.G.GC.....-.GC......G..CA.CC.G.CACG.TA...TC..C......G.....................GG.-..CC
c17  .....C.C.....T..ATTG.A..A.....A...GAGC.....T.TT.T....AA.CA.GCTA.CA.G.TAT..T.G..A.....G.....................C..GG....CC
c25  ...GTC.C.AA..TC..TA.CA.......C..G.GG.CC....C.TT.GT...A...A.C..T.GA.G.T...A..G.C..T...GAC.....................T...TC......
c35  .....C.......TGA.TTGCT..A....C.....GTC........A..G...G.AAC..C.A.CACG.A.T..T.CA.......G.....................TC......
c37  .....C.......TG.ATAT.A..A....CA.G..G.......T.T..CG...A...C.C.TG..ACG.TA..AT.CA.......G.....................---.GGAA
```

Figure 2. Alignment of nine 16*S* rRNA gene sequences amplified from Clarkia extracts. Amplification products from three extracts were cloned and 40 clones were sequenced. Twenty clones contained inserts with identifiable sequences. Clones 1–2 are from extract 1 (*Salix*), clones 7–24 from extract 2 (*Magnolia* or *Persea*) and clones 25–37 from extract 6 (*Quercus*). They represented nine different sequences. The sequences were aligned (Hein 1989*a*, *b*) with 14 published bacterial and chloroplast 16*S* rRNA sequences. This alignment is shown for the nine cloned sequences as well as the four sequences chosen as references in the phylogenetic analysis (figure 3). The decision on which positions to include in the analysis was based on the entire alignment of 14 published sequences. Included positions are underlined.

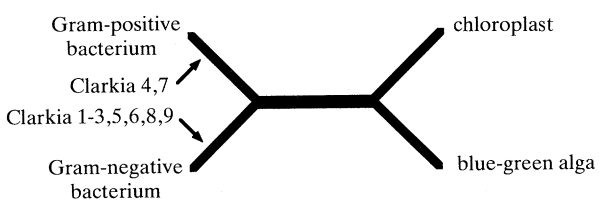


Figure 3. Phylogenetic analysis of the Clarkia sequences. Four reference sequences, one chloroplast (*Nicotiana tabacum*), one blue-green alga (*Anacystis nidulans*), one Gram-negative (*Pseudomonas aeruginosa*) and one Gram-positive bacterium (*Streptomyces coelicolor*) are joined in a network. The nine Clarkia sequences were added individually to the network and, in each case, their best position was determined by a maximum likelihood analysis (Felsenstein 1981, 1989). The best assignments of the nine sequences are shown. To assess the statistical significance of each assignment, we tested it against the other four possible assignments with a likelihood ratio test (Kishino & Hasegawa 1989). Grouping with the chloroplast sequence was ruled out significantly ($p < 0.05$) for all nine Clarkia sequences. Equally unlikely is the assignment of the sequences to the algal branch. In some cases, the placement of a Clarkia sequence on another bacterial line is not statistically ruled out. The sequences appear by numbers that correspond to the order with which they appear in figure 2.

Clarkia plant extracts were therefore cloned in M13 and a total of 40 clones were sequenced. Twenty of these proved to contain full length inserts of ribosomal origin. They represented a total of nine independent sequences which are presented in figure 2. Five of these sequences are represented by two or more identical clones, whereas the other four sequences were found only once.

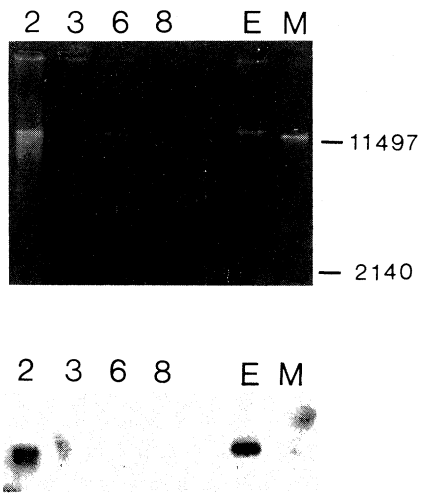


Figure 4. Hybridization of the amplification product from extract 2 to four Clarkia extracts and bacterial DNA. Portions of the indicated extracts as well as approximately 300 ng of DNA extracted from *E. coli* (lane E) and *Magnolia stellata* (lane M) were transferred from an agarose gel (upper panel) to a filter and hybridized to the radioactively labelled amplification from extract 2 (lower panel). The numbers indicate migration positions of molecular mass markers.

(*e*) *Phylogenetic analysis of sequences*

The sequences were aligned (Hein 1989*a*, *b*) with 14 bacterial and chloroplast 16*S* rRNA sequences. Figure 2 presents the cloned sequences as well as the sequences of four reference species; one plant (*Nicotiana tabacum*), one blue-green alga (*Anacystis nidulans*), one Gram-negative (*Pseudomonas aeruginosa*) and one Gram-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

positive bacterium (*Streptomyces coelicolor*). The blue-green alga was included in the analysis to avoid fortuitous groupings between bacterial sequences and the plants.

The four reference sequences were joined according to their known relationship (Woese 1987) in a network that contains four branches leading to the terminal taxa and one central branch (figure 3). The nine Clarkia sequences were then in turn added to the five different branches present in the network and the likelihoods of the five topologies were calculated and compared for each sequence (Kishino & Hasegawa 1989). In figure 3, the best assignments of the nine sequences are shown. In all cases, the placement shown is statistically better ($p < 0.05$) than a grouping with the chloroplast or the *Anacystis* sequences. It is thus clear that the sequences analysed are of bacterial origin. Furthermore, the frequency distribution of the sequences indicates that there are many more bacterial species present in these Clarkia specimens than the ones we have presented here.

(*f*) *Origin of the amplified sequences*

To find out whether the amplified sequences derive from the HMM DNA seen in agarose gels, portions of some extracts, on which the amplifications were done, were separated by electrophoresis and transferred to a filter. The amplification product from one extract used for cloning was radioactively labelled and hybridized to the filter (figure 4). The amplification product hybridized to the HMM DNA of the extract from which it is derived as well as to *E. coli* DNA. However, it did not hybridize to a plant DNA (*Magnolia stellata*). This confirms that the sequences amplified from this extract are of bacterial origin and are derived from the visible HMM DNA.

4. DISCUSSION AND CONCLUSION

The possibility exists that the bacterial sequences that we have amplified from the HMM DNA are of Miocene age. However, we believe that this is unlikely. One reason for this is that the error rate in the amplifications, which can be calculated from the cloned sequences (less than in 500 bases), is approximately equal to that observed in amplifications from fresh animal tissues (A.S., unpublished observations) and considerably lower than the inferred error rate in amplifications from genomic DNA extracted from 4500-year-old maize (P. Goloubinoff, personal communication). The template DNA for this amplification thus seems to be largely intact. We consider this an unlikely state of affairs after 16–20 Ma (Pääbo & Wilson 1991). An alternative possibility is that bacteria have survived in the fossiliferous strata in Clarkia and that the sequences are thus derived from living bacteria that have been isolated since the Miocene. This possibility deserves further investigation. However, the most likely possibility may be that bacteria recently penetrated part of the fossil bed, perhaps from the exposed edge of the excavation site. Whatever the origin of the bacteria in the Clarkia fossils, our results show that further attempts to amplify plant sequences from Clarkia fossils should not be limited only to extracts where HMM DNA can be visualized in agarose gels. The possibility of finding plant sequences may be equally high in extracts where no DNA is detected by such techniques.

We thank Dr Jack Smiley for invaluable help and support during the field work, Mr and Mrs Kienbaum and the Tertiary Research Center for access to the fossil site, Dr Tom White, Dr John Taylor and Dr Diane Leong for providing primers, Dr Michael Clegg and Dr Ed Golenberg for sharing their experience and expertise, and Ms. Gertraud Feldmaier-Fuchs for expert technical assistance. This work was supported by grants from the NSF and NIH to Dr Allan Wilson and by the University of Munich.

REFERENCES

Felsenstein, J. 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. molec. Evol.* **17**, 368–376.

Felsenstein, J. 1989 Phylogenetic Inference Programs (PHYLIP) Manual 3.2, University of Washington, Seattle and University Herbarium, University of California, Berkeley.

Giannasi, D. E. & Niklas, K. J. 1985 The paleobiochemistry of fossil angiosperm floras. Part I. Chemosystematic aspects. In *Late Cenozoic history of the Pacific Northwest* (ed. C. J. Smiley), pp. 161–174. San Francisco: AAAS.

Golenberg, E. M., Giannasi, D. E., Clegg, M. T., Smiley, C. J., Durbin, M., Henderson, D. & Zurawski, G. 1990 Chloroplast DNA sequence from a Miocene *Magnolia* species. *Nature, Lond.* **344**, 656–658.

Hein, J. 1989*a* A new method that simultaneously aligns and reconstructs ancestral sequences for any number of homologous sequences, when the phylogeny is given. *Molec. Biol. Evol.* **6**, 649–668.

Hein J. 1989*b* A tree reconstruction method that is economical in the number of pairwise comparisons used. *Molec. Biol. Evol.* **6**, 669–684.

Kishino, H. & Hasegawa, M. 1988 Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. molec. Evol.* **29**, 170–179.

Lindahl, T. & Andersson, A. 1972 Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. *Biochemistry* **11**, 3618–3623.

Lindahl, T. & Nyberg. B. 1972 Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11**, 3610–3618.

Niklas, K. J. & Brown, R. M. Jr 1981 Ultrastructural and paleobiochemical correlations among fossil leaf tissues from the St. Maries River (Clarkia) area, Northern Idaho, U.S.A. *Am. J. Bot.* **68**(3), 332–341.

Pääbo, S., Gifford, J. A. & Wilson, A. C. 1988 Mitochondrial DNA sequences from a 7000-year-old brain. *Nucl. Acids Res.* **16**, 9775–9787.

Pääbo, S. & Wilson, A. C. 1991 Miocene DNA sequences – a dream come true? *Curr. Biol.* **1**, 45–46.

Sambrook, J., Fritsch, E. F. & Maniatis, T. 1989 *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor Laboratory Press.

Smiley, C. J. (ed.) 1985 *Late Cenozoic history of the Pacific Northwest*. San Francisco: AAAS.

Woese, 1987 Bacterial evolution. *Microbiol. Rev.* **51**, 221–271.

Discussion

S. HUMMEL (*Institut für Anthropologie der Universität Göttingen, F.R.G.*). Professor Pääbo mentioned that in the Moa bird

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

from New Zealand the bone DNA was found to be in a better state of preservation than the DNA in the soft tissue remains. Has this finding to be assumed as a single case or might it be systematic? Are there explanations for the better preservation of the bone DNA?

S. Pääbo. This was only observed in once experiment. It is possible that the preservation of bone DNA may be due to the binding of DNA to hydroxyapatite. It does appear that enzymic decay is not as fast in bone as in soft tissue.

S. Hummel. How does Professor Pääbo know that the sequence changes he sees are not the result of base damage?

S. Pääbo. Base changes resulting from damage are not seen when the amplification product is directly sequenced because direct sequencing produces a sequence that is averaged over all molecules amplified. Incorrect and unambiguous sequences would only be produced by direct sequencing if the amplification started from one single molecule or one single jumping event between two or more molecules. Such errors would be detected when the PCR was repeated from the same extract or another extract from the same specimen. If the amplification product is cloned and single cones sequenced a high error rate is seen that reflects PCR artefacts introduced by damage in the template DNA. A further source of confidence in the ancient sequences is that the variants observed or close relatives of them exist in present-day populations and that the differences observed fall within the spectrum of variation known to exist for the molecule in question.

J. L. Bada (*Scripps Institution of Oceanography, University of California at San Diego, U.S.A.*). In the HPLC analyses of the bases in ancient DNA, are the other peaks diagenetic products of A, G, C and T?

S. Pääbo. We assume that the additional peaks that appear in the chromatograms from ancient DNA extract are modified bases, predominantly from pyrimidines.

P. Westbroek (*Geobiochemistry Unit, Department of Biochemistry, University of Leiden, The Netherlands*). Is Professor Pääbo discussing rare success events, or is success frequent? It would be good to know this for those with an interest in the field and planning new experiments.

S. Pääbo. For soft tissue remains that macroscopically look well preserved it is possible to amplify mitochondrial sequences of 100 b.p. in most cases. For bones our experience to date is that success is much more rare.

E. Hagelberg (*MRC Molecular Haematology Unit, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, U.K.*). Concerning Professor Pääbo's 'criteria of authenticity' applied when amplifying soft tissue remains of human origin: what are these criteria, and how do they rule out artefactual results?

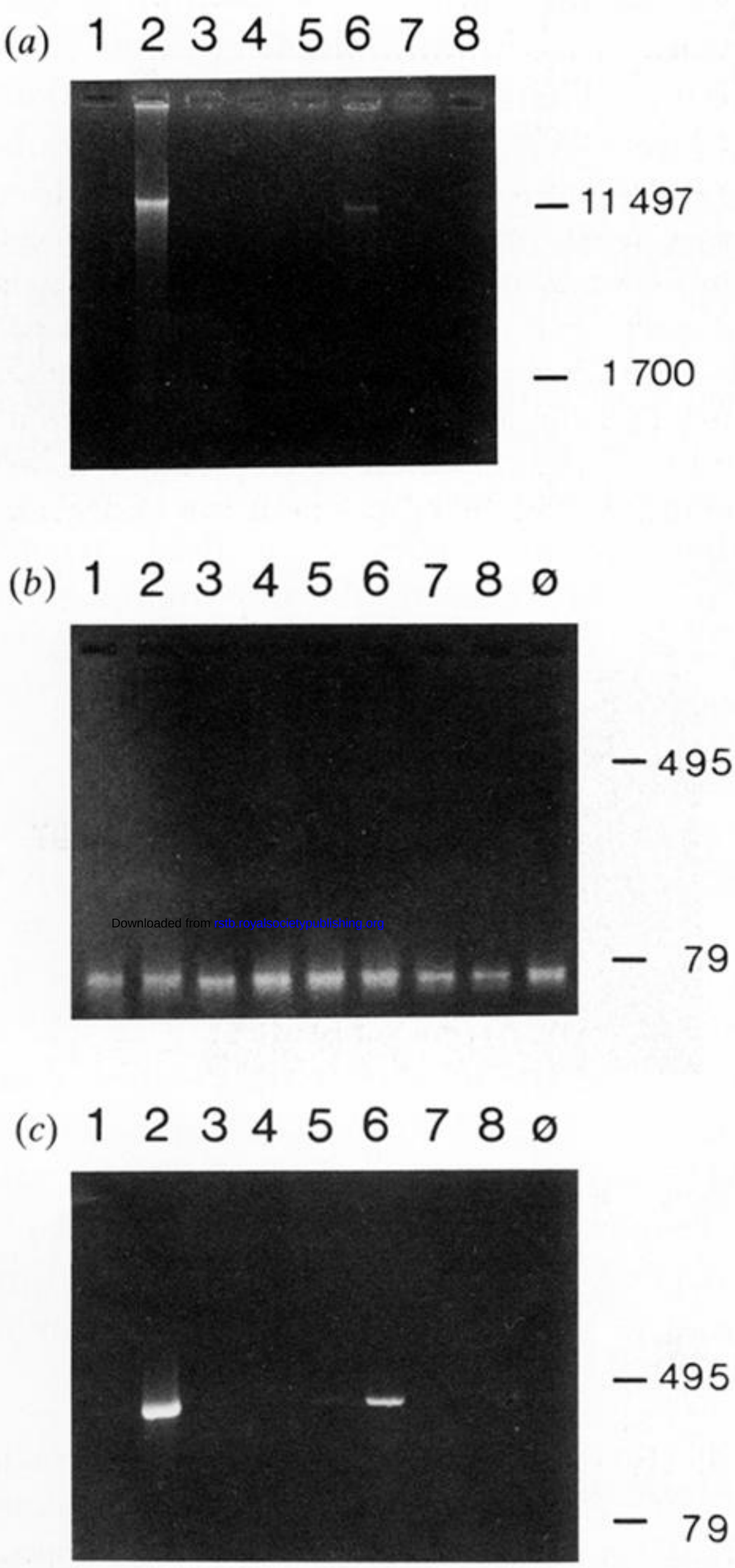
S. Pääbo. These have been published in *J. biol. Chem.* **264**, 9707–9712 (1989) and *PRC protocols: a guide to methods and applications*, pp. 159–166, Academic Press (1990). In short they include: (i) amplification from extraction controls, i.e. extracts done without any tissue present; (2) PCR controls, i.e. amplifications without any extract added; (iii) several independent extracts from the same individual giving identical sequences; (iv) unambiguous sequence reactions and (v) for our material we also often find it useful to do amplifications with a series of primer pairs yielding products of different lengths. The damaged state of the old DNA causes an inverse correlation between amplification efficiency and size of the amplification product.

E. Hagelberg. Professor Pääbo's comments on the inability to repeat Golenberg's chloroplast DNA work are analogous to his criticism of bone DNA work before he was able to repeat the work himself.

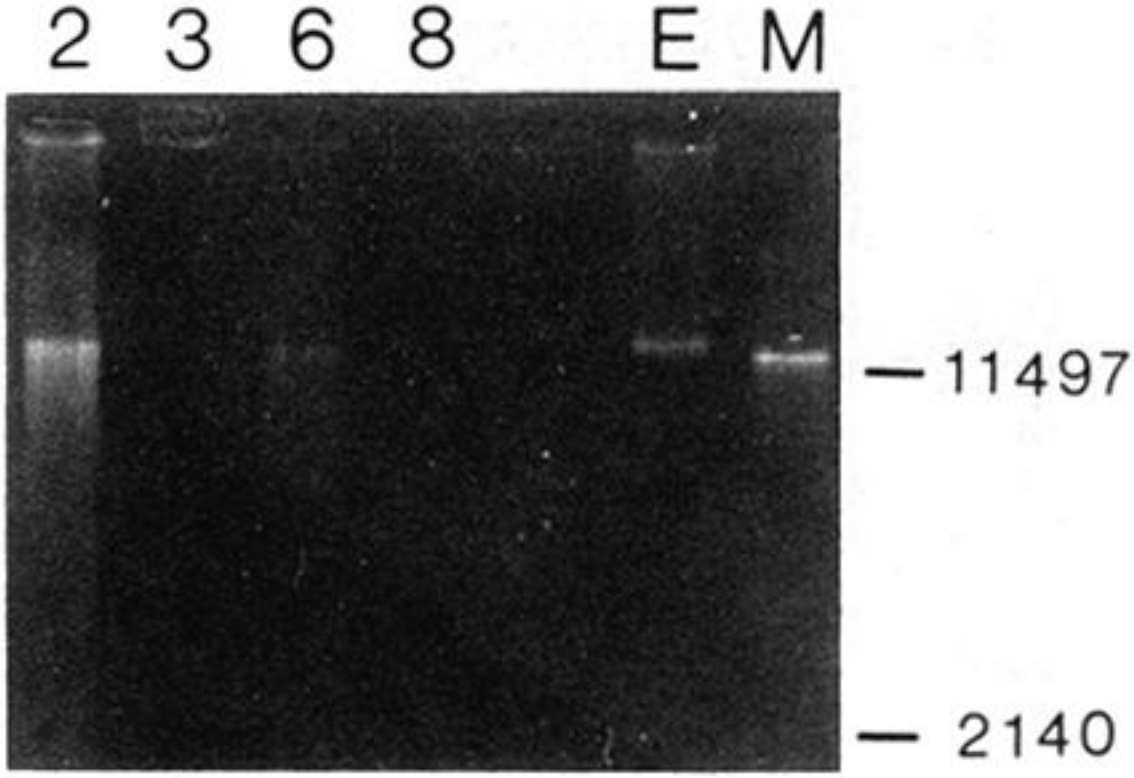
S. Pääbo. I think that it is important that we discuss our work in an open and constructive way.

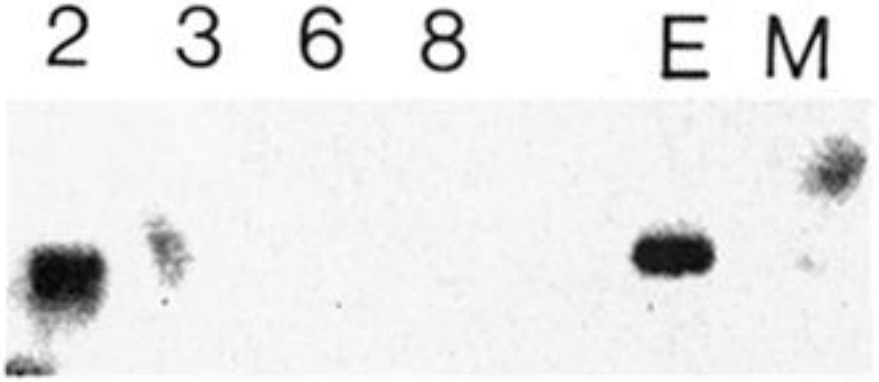
G. B. Curry (*Department of Geology and Applied Geology, University of Glasgow, U.K.*). The last discussion has highlighted the fact that sequencing ancient DNA is a relatively new field, and one which is providing many difficult problems for geneticists interested in this work. From the geological perspective there seems to be very little hard evidence on the fossilization conditions likely to be most conducive to the survival of ancient DNA, especially as in some cases DNA has been recovered from rocks in which the state of morphological preservation is poor. It does seem clear that it is crucial to concentrate, at least initially, on genes that are well studied from living organisms to avoid the numerous complications and potential pitfalls associated with trying to draw phylogenetic information from the small segments of DNA sequences available from the fossil record.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES



gure 1. Electrophoretic analysis of extracts from Clarkia ssils and enzymic amplifications with plant and bacterial imers. (*a*) 10 µl of seven extracts from Clarkia samples anes 1–7) and one control extract (lane 8) were analysed by ectrophoresis through a 0.8 % agarose gel and visualized by aining with ethidium bromide. The extracts were: 1, *Salix*; *Magnolia* or *Persea*; 3, clay next to number 2; 4, *Magnolia*; *Betula*; 6, *Quercus* (acorn); 7, *Quercus*; 8, extraction control.) Amplifications from the above extracts using primers ecific for the chloroplast *rbc*L gene (*Den 1* and *Rub 2*). (*c*) mplifications from the same extracts using primers designed amplify a segment of the bacterial 16*S* rRNA gene. Lanes here no extract was added to the amplification reaction e indicated by Ø. Numbers indicate the migration positions molecular size markers.





gure 4. Hybridization of the amplification product from tract 2 to four Clarkia extracts and bacterial DNA. ortions of the indicated extracts as well as approximately 0 ng of DNA extracted from *E. coli* (lane E) and *Magnolia llata* (lane M) were transferred from an agarose gel (upper nel) to a filter and hybridized to the radioactively labelled nplification from extract 2 (lower panel). The numbers dicate migration positions of molecular mass markers.