Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Amplification and Analysis of Miocene Plant Fossil DNA [and Discussion]

Edward M. Golenberg, T. A. Brown, J. L. Bada, P. Westbroek, M. J. Bishop and G. A. Dover

Phil. Trans. R. Soc. Lond. B 1991 **333**, 419-427
doi: 10.1098/rstb.1991.0092

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

This journal is © 1991 The Royal Society

Downloaded from rstb.royalsocietypublishing.org

Amplification and analysis of Miocene plant fossil DNA

EDWARD M. GOLENBERG

Department of Biological Sciences, Wayne State University, Detroit, Michigan 48202, U.S.A.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

SUMMARY

Ancient DNA has been extracted and sequenced from several animal and plant specimens. Previous considerations of the damage to ancient DNA have suggested that both the age and size of DNA fragments that can be retrieved and sequenced may be limited, the former to between several thousand and at most tens of thousands of years old, and the latter to at most a few hundred bases. A recent report of a 770 base pair (b.p.) sequence from the chloroplast gene *rbc*L from a Miocene *Magnolia latahensis* leaf indicates that both estimated limitations may be too conservative. Further work has indicated that analysis of Miocene fossil DNA can be replicated, and can, therefore, open up the prospects for future development of the field of molecular palaeontology.

Successful amplification of fossil DNA is sometimes confounded by factors inherent to fossil DNA or to samples with minimal amounts of target DNA. Techniques that alter denaturation, reduce inhibitors and the problem of contaminants, and repair DNA prior to polymerase chain reaction amplification can increase the probability of success.

1. INTRODUCTION

Evolutionary biology is the study of the patterns and processes of genetic changes in populations of organisms. It has been traditionally divided into three disciplines: palaeontology, systematics and evolutionary genetics. Of the three, evolutionary genetics has been the most isolated from the other two areas. The most prominent reason for this isolation is the lack of emphasis of historical perspective in the majority of population genetic studies. This is due in part to the concentration on studies of microevolutionary change in contemporaneous populations, particularly with emphases on elucidating adaptive against neutral evolution. Although these studies are important, the importance of the conditionality of prior states in determining evolutionary trajectories is rarely explicitly stated or seriously considered in population genetic studies.

With the development of molecular genetic techniques over the past ten years, systematists and evolutionary geneticists have entered into a rapprochement. Systematists appreciate the power of using either direct DNA sequences or restriction fragment analysis to generate a virtually unlimited data set of unequivocally genetic characters. In concert, theoretical population geneticists have used the models of genetic change to develop methods for phylogenetic inference. As a result, the study of the processes of molecular evolutionary change is of interest to both evolutionary genetics and molecular systematics.

Phylogenetic work on non-contemporaneous DNA was first advanced by the work of Higuchi *et al.* (1984). The authors cloned a fragment of DNA from a museum specimen and, perhaps of equal significance, they introduced the use of phylogenetic inference to establish the authenticity of the sequence. Further work on old and ancient animal DNA was advanced, and extensive characterization and sequencing of DNA of up to 4000 years old was achieved (Pääbo 1985*a*, 1989, 1990; Pääbo *et al.* 1988, 1989, 1990). In addition, recent work on extracting DNA from bones (Hagelberg *et al.* 1989) indicates that suitable material for ancient DNA analysis may be more readily available than previously thought.

Initial studies on ancient plant DNA were published at nearly the same time. Rogers & Bendich (1985) reported extracting nanogram amounts of restrictable DNA from plant tissues ranging in age from 22 to greater than 44600 years old. In contrast to the condition of ancient animal DNA, high molecular mass plant DNA was apparent in the samples. Later work by Helentjaris (1988) indicated that plant material from archaeological sites may also be amenable to DNA analysis.

2. BRIEF SUMMARY OF *MAGNOLIA* PAPER

These initial successes in the analysis of DNA extracted from old sources prompted an attempt at recovery and analysis of palaeontological leaf specimens from the Miocene fossil deposit at Clarkia, Idaho, U.S.A. (Golenberg *et al.* 1990). The Clarkia site has been well studied (Smiley 1985). The deposit is thought to be derived from the early Miocene, based primarily on floral composition (Smiley & Rember 1985). The flora of the Miocene community is representative of a warm, humid forest community, in contrast to the predominantly coniferous forests now present in this

Phil. Trans. R. Soc. Lond. B (1991) **333**, 419–427
Printed in Great Britain

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions: Biological Sciences.* www.jstor.org

Downloaded from rstb.royalsocietypublishing.org

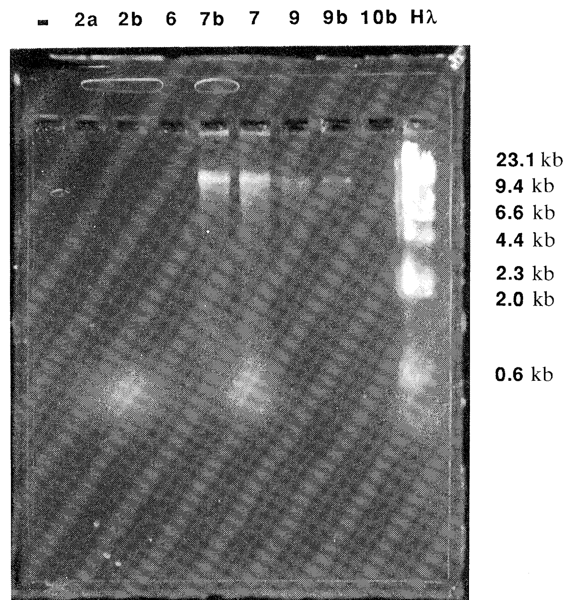


Figure 1. Fossil DNA extractions separated on an agarose gel. Lane 1 is a negative control extraction using the extraction buffer without sample. Lanes 2 and 3 are split samples from a fossil *Persea* leaf. Lane 3 is a sample from a fossil *Magnolia* leaf. Lanes 5 and 6 are split samples from a *Cocculus* leaf. Lanes 7 and 8 are split samples from a separate *Cocculus* leaf and lane 9 is a sample from holding buffer of a *Cocculus* seed sample before grinding of the tissue. Lane 10 contains lambda phage DNA digested with *Hin*d III as size markers.

area. The plant megafossils are chiefly leaf compression fossils, although reproductive parts and other vegetative parts are also found. In general, the cellular ultrastructure and some secondary metabolites are well preserved (Giannasi & Niklas 1985; Niklas *et al.* 1985). In particular, Niklas *et al.* (1985) found that chloroplasts were extremely well preserved, being found in over 90% of the cells examined in some species.

DNA was extracted from leaf and reproductive materials scraped from the shale surface using the extraction procedure of Rogers & Bendich (1985) modified to include SDS instead of CTAB (Golenberg *et al.* 1990). Approximately one in ten extracts contained high molecular mass DNA that could be visualized on an agarose gel stained with ethidium bromide. Recent collections made in August 1990 reconfirmed both the extractability and frequency of these earlier findings. As seen in figure 1, samples containing DNA have high concentration of high molecular mass DNA (lanes 5, 6, 7, 8) along with a considerable amount of degraded DNA molecules. In contrast, no DNA can be visualized from samples in lanes 2, 3, 4 and 9. The absence of any noticeable DNA in most samples is contrary to the possibility of a contaminant that is systematically present during the extractions. In addition, the repeatability of positive results in extracts from divided samples (lanes 5, 6, 7, 8) suggests that the source of the DNA must be from the particular fossil material itself.

As a test of the authenticity of the extracted DNA, a 770 b.p. segment of the chloroplast gene, *rbc*L, was targeted for amplification. The rationale for using this particular gene is that chloroplast genomes occur in thousands of copies in leaves, the chloroplasts themselves are differentially well preserved in comparison with other subcellular structures (Niklas *et al.* 1985), chloroplast genes evolve slowly allowing for intra- and interfamilial phylogenetic comparisons, and a large data set of *rbc*L sequences is already available. Thus, by focusing on *rbc*L, we increase our chances of successful DNA extraction and authentication. An additional advantage of analysing *rbc*L is that it is not present in yeasts or in bacteria (with the exception of cyanobacteria). Therefore false positives from amplification of potential epiphytic DNA could be avoided.

An *rbc*L fragment was successfully amplified from a DNA sample extracted from a *Magnolia latahensis* leaf (Golenberg *et al.* 1990). The sequence was determined from both strands and from separate amplifications. Using the criteria outlined in Higuchi *et al.* (1984) the sequence was then compared with four other species, *Magnolia macrophylla*, *Liriodendron tulipifera*, *Persea americana*, and *Platanus racemosa*. Estimates of the number of nucleotide substitutions between the purported fossil *Magnolia* sequence and the four extant species indicated that the fossil was most similar to *M. macrophylla*, and then showed increasing numbers of substitutions to *L. tulipifera*, *P. americana* and *P. racemosa*, in that order, as would be expected from systematic considerations. There were 17 substitutions between the fossil and extant *Magnolia* samples, of which 13 are third-position synonymous substitutions. Twelve of the substitutions were transition substitutions. Both of these patterns of substitution biases correspond well to the transition : transversion and codon position substitution biases observed in *rbc*L (Clegg *et al.* 1986) suggesting that there is no obvious error introduced in the polymerase chain reaction replication. Having met these criteria and further phylogenetic analyses, we concluded that the amplified sequence was indeed a copy of the original Miocene gene fragment. This result indicates that the age of DNA that may be recovered and sequenced may be substantially extended.

3. CRITICISM OF PÄÄBO & WILSON

Pääbo & Wilson (1991) have recently reviewed the *Magnolia* work at Clarkia (Golenberg *et al.* 1990) and have suggested that it be received with caution until the results are further verified and duplicated. They state two reasons for their scepticism that informative DNA sequences may be obtained from several million-year-old material.

The first reason stems from *in vitro* studies of rates of spontaneous depurination (Lindahl & Nyberg 1972). Under hydrated conditions and physiological Mg^{2+} ionic strength, the rate of depurination is 4×10^{-9} per second at 70 °C and pH 7.4 (Lindahl & Nyberg 1972). Following depurination, the DNA chain is expected to break relatively rapidly (Lindahl & Andersson 1972). Extrapolating from the empirical rates, Pääbo & Wilson (1991) suggest a rate depurination of 6.33×10^{-6} sites per year. Based on this rate, and a reasonable assumption of copies of cpDNA molecules, no 800 b.p. fragments are expected to remain after

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

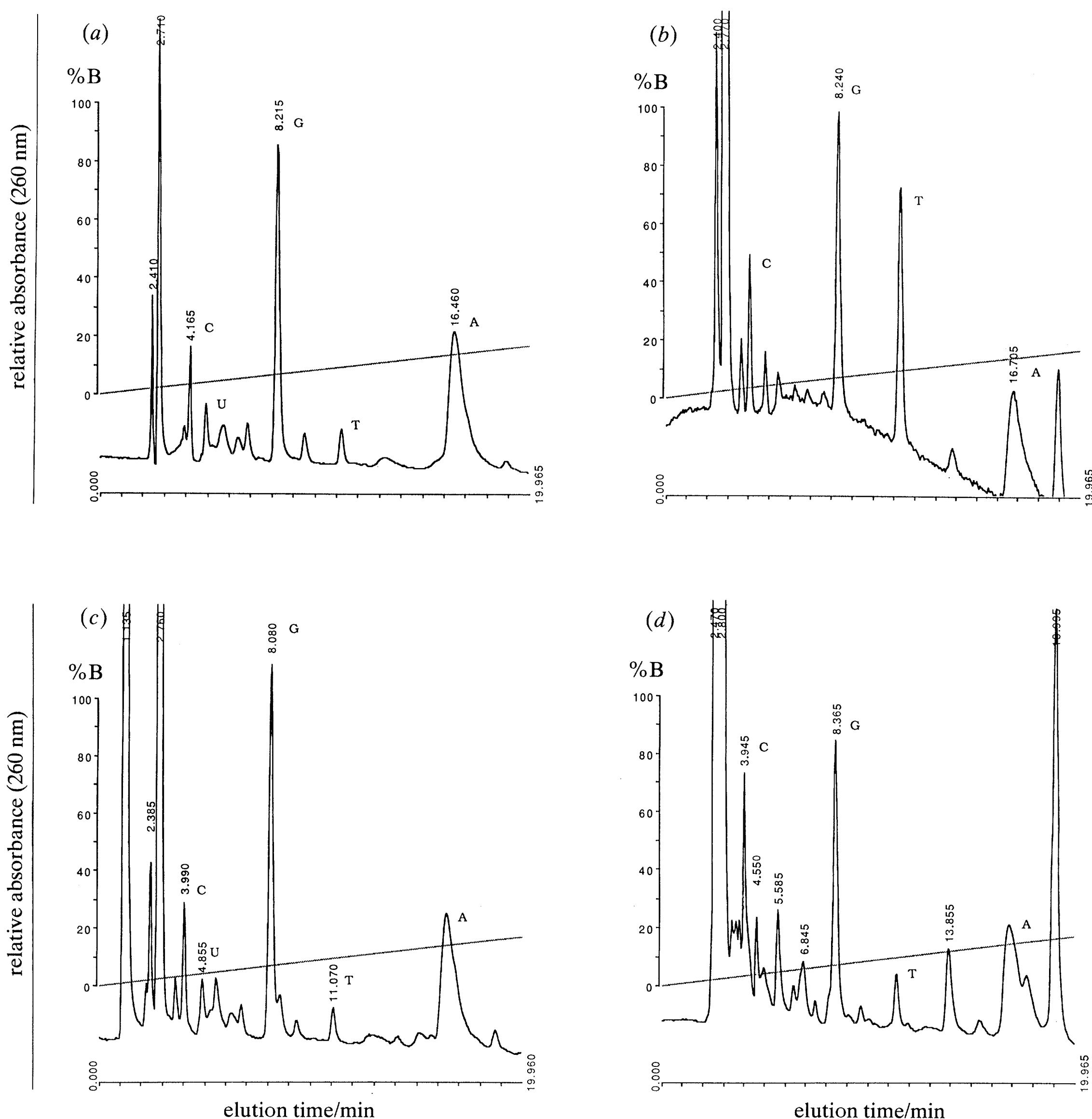


Figure 2. HPLC analysis of hydrolysed DNA. Columns, buffers and reaction conditions follow Pääbo (1985*b*, 1989). Elution times are given at a rate of (1 ml min^{-1}). Y axis is relative absorbance at 260 nm. Samples analysed are: (*a*) modern *Hordeum spontaneum*; (*b*) modern *Hordeum spontaneum* treated with RNase with the free nucleotides removed by passing through a G-25 Sephadex column before hydrolysis; (*c*) fossil *Cocculus* DNA; (*d*) fossil *Cocculus* DNA treated with RNase with the free nucleotides removed by passing through a G-25 Sephadex column before hydrolysis.

approximately 5000 years. Extending their analysis, it can be shown that, for an equal number of molecules, no 150 b.p. fragments should remain after 26700 years and no purines should remain at all after 4 million years (Ma).

The second reason for their scepticism comes from Pääbo & Wilson's reported inability to amplify sequences from Clarkia DNA extractions. They conclude, after amplifying bacterial sequences from Clarkia DNA extracts, that the source of the high molecular mass DNA visible in the extractions is eubacterial.

4. RESPONSE

Both of these concerns question the validity of the present work on the Miocene Clarkia DNA and the future of extended molecular palaeontological work. Of course, empirical confirmation must take precedence over theoretical objections. Yet some discussion of these theoretical objections may be of value in understanding some of the limits of molecular palaeontology in the future. Accordingly, I attempt to address these problems and further suggest practical steps that may be taken to improve the probability of success when amplifying fossil DNA.

In studying the rates of spontaneous depurination, Lindahl & Nyberg (1972) investigated DNA degradation under pH conditions ranging from 4.5 to 7.5 and at temperatures ranging from 80 to 45 °C. Under a constant pH of 5.0, the reactions decreased in a log-linear fashion with a decrease of temperature. In contrast, the change of rates of depurination with change of pH under a constant temperature of 70 °C was not log linear. Lindahl & Nyberg (1972) also found that while the presence of amines did not markedly reduce rates of depurination, depurination was slowed in solutions of higher ionic strength. It may

Downloaded from rstb.royalsocietypublishing.org

well be, therefore, that rates of depurination in the Clarkia leaves may not be estimated by simple linear extrapolation.

Niklas *et al.* (1985) suggested that the leaves at Clarkia underwent a series of transformations that may have been instrumental in their extraordinary preservation. A majority of leaves may have undergone natural dehydration before abscission resulting in the rupturing of the tonoplasts and a release of high concentrations of naturally occurring plant phenolics. Rehydration in tannin-rich waters may also have been instrumental in the preservation of the leaves. Niklas *et al.* (1985) also report that plants with normally high tannin contents were often found in the best state of preservation. Thus it is likely that the conditions in which the DNA was preserved are not similar to normal physiological conditions, and, therefore, depurination may have been slower than estimated. If these considerations prove to be important in DNA preservation, we might expect that plant material with high concentrations of phenolics and tannins may be a more likely source for future molecular palaeontological work than animal tissue.

A second consideration that may explain the inconsistencies between empirical results and theoretical expectations comes from the simple Poisson model used to estimate DNA decay (Pääbo & Wilson 1991). Single depurination events along a DNA dimeric polymer may not be sufficient to destroy the information on the molecule. A complete single strand of DNA is sufficient for double-stranded amplification. Therefore at least two hits, one on each strand, may be needed for disruption of the necessary information. In addition, it is possible that incomplete strands may act as primers on overlapping fragments and artificially recreate complete templates in a lag phase of amplification (Pääbo *et al.* 1989, 1990). This may also help to explain the large fragments amplified from the Clarkia deposit (Golenberg *et al.* 1990; this paper), and the smaller fragment reported amplified from a 40000-year-old mammoth (R. Higuchi & A. C. Wilson, unpublished results; see Pääbo 1989).

Further evidence of the condition of the Clarkia DNA is now available (E. M. Golenberg, G. M. Li & L. Romano, unpublished results). Samples of DNA extractions from the Clarkia fossils and from modern barley (*Hordeum spontaneum*) were treated with formic acid at 170 °C and then analysed using high-performance liquid chromatography (HPLC) following the methods of Pääbo (1985*b*, 1989). The peaks of the nitrogenous bases were first identified by injecting pure bases and later confirmed by coinjections with samples. Cytosine, uracil, guanine, thymine and adenine were eluted at 3.9, 4.8, 8.1, 11.1 and 16.4 min respectively (flow rate 1 ml min^{-1} in a linear gradient of 1 to 10% acetonitrile). Both modern and fossil DNA samples were distinguished by identifiable peaks for all five nitrogenous bases (figure 2*a*, *c*). This result is somewhat biased in the fossil sample because of extraction with carrier tRNA as a coprecipitant, although the relative amounts of uracil to thymine and to cytosine are very similar in the modern barley sample which was not extracted with tRNA. A further treatment of the sample with RNase and separation in a G-25 Sephadex column removed the RNA before formic acid treatment and HPLC analysis (figure 2*b*, *d*). In comparison with the modern sample, the fossil DNA has a decrease in thymine and a slight decrease in guanine. Also, an undefined peak at 13.8 min increased in the fossil samples. However, it is clear that there is no drastic depurination in the fossil samples. These same samples contained plant DNA as is evidenced by the ability to amplify *rbc*L fragments from them (see below).

Pääbo & Wilson's second reason for scepticism arises from their inability to amplify chloroplast sequences from the fossil sample and suggest that the DNA is eubacterial in origin. Amplification of fossil plant DNA is more difficult than amplifying fresh animal or bacterial DNA (see next section). An absence of polymerase chain reaction (PCR) products in a reaction reflects more on these difficulties than on the existence of template DNA. Similarly, contaminating DNA is a constant worry in all PCR reactions. The occurrence of a false positive, therefore, reflects more on the cleanliness of the PCR experiment than on the existence or not of true target DNA. Scanning electron microscopy of several leaves from the Clarkia remains did not show any evidence of epiphytic bacterial growth (W. Thompson, person communication).

5. PROBLEMS AND SOLUTIONS FOR AMPLIFICATION OF FOSSIL DNA MATERIAL

Amplification of specific DNA fragments of *Taq* polymerase-mediated PCR can fail even under the best conditions. The use of fossil DNA extracts as template material can be expected to introduce further complications that will reduce the likelihood of successful amplification reactions. Analysis of the factors that complicate the amplification of fossil DNA can be useful for suggesting modifications of standard PCR techniques that would increase success rates.

Extracts of fossil DNA present two groups of factors that may adversely affect successful PCRs. The first group of factors results from the minimal amounts of DNA extracted from the sample material and would be shared by comparably minute amounts of freshly prepared DNA. These factors include increased susceptibility to false positive amplifications resulting from trace amounts of contaminating DNA, and inefficient priming as a result of incomplete denaturation of template molecules. The second group of factors is specific to ancient and fossil DNA and is caused by the presence of co-extracted polymerization-inhibiting substances and from the natural degradation of the sample DNA. I treat each of these factors separately.

(*a*) *Contaminating DNA*

The potential for contaminating DNA to serve as a template for the polymerase chain reaction has been noted in many sources (e.g. Pääbo *et al.* 1989). The extent to which this is a problem in standard PCRs is related to the ratio of renaturation of primers to legitimate against illegitimate target sequences. This

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

ratio varies according to the similarity of the primer sequence to the target DNA and to the illegitimate, contaminating DNA, and according to relative ratios of two classes of template DNA molecules. The degree of similarity of the primer sequence to the target DNA and therefore the ability of the target DNA to outcompete illegitimate DNA molecules for primers is independent of the age of the DNA and is solely a function of the ability to design sequence or species-specific primers. In contrast, the second factor, that of the ratio of legitimate to illegitimate sequences, is indirectly a function of the age of the sample DNA. DNA extractions from often limited source material and the natural degradation of the DNA result in very limited amounts of complete target molecules. As such, even minute amounts of contaminating DNA that would be swamped out in reactions with extant DNA, can successfully compete with target fossil DNA. The observations of some workers of a tendency to find false products in later cycles of the PCR ($\geqslant 40$) is consistent with this interpretation.

Much has been written concerning the necessity for clean techniques in preparing PCR reactions (Kwok & Higuchi 1989). For the reasons given above, these admonitions should be adhered to especially closely by researchers working with fossil DNA. However, beyond standard clean techniques, a procedure in which batch solutions are exposed to short-wave UV irradiation before the addition of the double-stranded target DNA and the polymerase enzyme is very effective in removing the problem of contaminating DNA (Sarkar & Sommer 1990). UV irradiation causes the formation of cross linkages in double-stranded DNA and thereby reduces its effectiveness as a template for a PCR. The single-stranded oligonucleotide primers are not significantly damaged by exposure to UV irradiation. Any secondary introduction of contaminating DNA can only come in pipetting the target DNA and the polymerase enzyme. Use of negative controls can effectively monitor systematic contamination from these sources. Any consistent evidence for contaminating DNA that does not appear in the negative control reactions will suggest that the source tube of the fossil DNA is contaminated.

(*b*) ***Small quantities of target DNA***

Small quantities of target DNA do not simply reduce the quantity of the final amplified product in a linear manner. Additional considerations arise, such as that of contamination discussed above. Polymerase chain reactions must be viewed as a series of linked reactions acting upon a population of target molecules. It is clear that successful polymerization is predicated upon successful priming of the target molecules with the synthetic oligonucleotide primers. Successful priming is predicated upon both sequence similarity and successful denaturation of target molecules. It is expected that none of the three reactions, polymerization, priming and denaturation, are completely efficient; that is, only some fraction of the total population of molecules will denature, and only a fraction of these will be successfully primed. Under standard PCR conditions, the number of target molecules is large so that incomplete denaturation and priming will not noticeably affect the success of the experiment. In contrast, when the number of target molecules is small, inefficient denaturation may effectively prevent polymerization resulting in little or no product. A simple method to adjust for this problem is to increase denaturation time in the early cycles to 5 min. This longer exposure to temperatures of 94 °C increases the probability that each DNA molecule will denature, and thus act as a template for polymerization.

It is worth noting here that these suggested longer denaturation times are in addition to a preincubation at 94 °C for 2 min before the thermal stable polymerase is added. The preincubation prevents initial polymerization of unwanted sequences because of mispriming under relaxed conditions as the DNA sample is brought up to 94 °C in the first cycle (0 cycle false polymerization). False products produced in the early cycles may serve as substrate for polymerization as the cycling continues. These products are especially noticeable in single-stranded amplifications where they appear as smears of both higher and lower molecular mass around the desired product. These products do not commonly interfere in sequencing reactions when sequencing primers internal to the single-stranded DNA are used; however, by competing for primers in the PCR, they reduce the quantity of desired single-stranded product.

(*c*) ***Specificity of primers***

A common reason for the failure of PCRs or any DNA polymerization reaction is the incompatibility of the 3′ end of the synthetic primers to the target DNA. This should be one of the first considerations for negative results in either sequencing or polymerase chain reactions of fresh extant DNA. It is equally necessary to explore such possibilities when trying to amplify fossil or ancient DNA, instead of prematurely assigning the problem to the template DNA. Design of primers should take into consideration expected rates of sequence divergence and, therefore, be based on known sequences from extant species within a reasonable systematic distance. Conserved regions from such sequences should be identified by sequence comparisons and should be preferred as candidates for primers.

Ancient DNA is expected to have the additional problem of being largely degraded. This aggravates the problem of primer design because one cannot assume *a priori* that a target molecule of the desired length will be present in sufficient number to give a positive result. Thus, amplification experiments should be designed to use a series of primers bracketing various lengths and positions in the gene. Figure 3 illustrates such an approach using DNA extracted from *Cocculus* sp. Sets of primers from positions 1–724, 209–724, 478–1201 (not shown) and 724–1366 (not shown) in the chloroplast gene, *rbc*L, produced negative results. Primers bracketing positions 478–955 and 478–724 (not shown in figure) produced positive results.

Downloaded from rstb.royalsocietypublishing.org

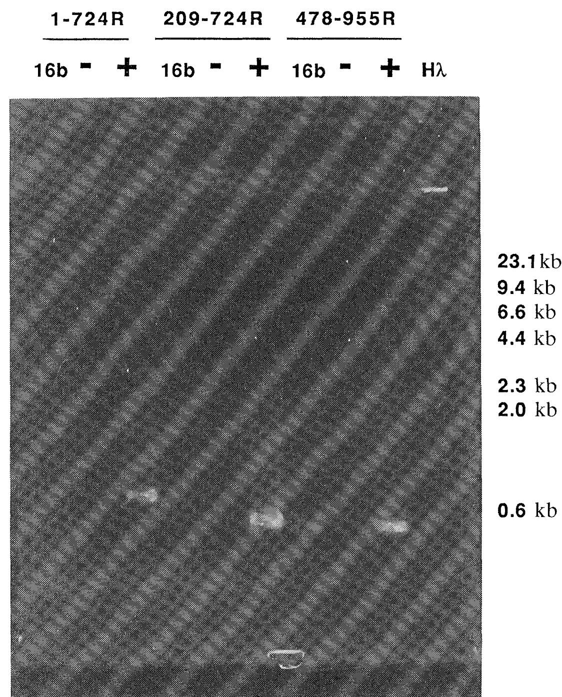


Figure 3. PCR amplification products separated on an agarose gel. Lanes 1, 4, and 7 are products of reactions using fossil *Cocculus* DNA as a template. Lanes 2, 5, and 8 are products from a negative control using extractions from extraction buffer without sample as a template. Lanes 3, 6, and 9 are products from positive control reactions using modern *Hordeum spontaneum* DNA as template. All primers used are based on the *Liriodendron tulipifera rbc*L sequence. Lanes 1 to 3 are from reactions using primers LtrbcL 1 (ATG TCA CCA CAA ACA GAG ACT AAA GC) and LtrbcL 724R (CTT CGC ATG TAC CTG CAG TAG C); lanes 4 to 6 are from reactions using primers LtrbcL 209 (GGA CCG ATG GAC TTA CCA GCC TTG ATC G) and LtrbcL 724R; lanes 7 to 9 are from reactions using primers LtrbcL 478 (GAT AAA TTG AAC AAG TAT GGT CGT CC) and LtrbcL 955R (GAG TGA ATA TGA TCT CCA CCA GAC ATA CG). The reactions were run in 100 µl volumes using 50 pmol of each primer with the standard (Promega Taq polymerase) reaction buffer. Before the addition of the template DNA and enzyme, the solutions were exposed to short-wave UV light for 3–5 min. Two µl (approximately 2 to 5 ng) of template DNA were used. Before addition of enzyme, the solution was incubated at 94 °C for 2 min. The enzyme was added while maintaining the solutions at 80 °C or higher. The heat settings were as follows: four cycles of 94 °C (5 min), 38 °C (2 min), 74 °C 3 (min); followed by 11 cycles of 94 °C (1.5 min), 42 °C (2 min), 74 °C (3 min); followed by 20 cycles of 94 °C (1.5 min), 48 °C (2 min), 74 °C (3 min) and then 15 min at 74 °C.

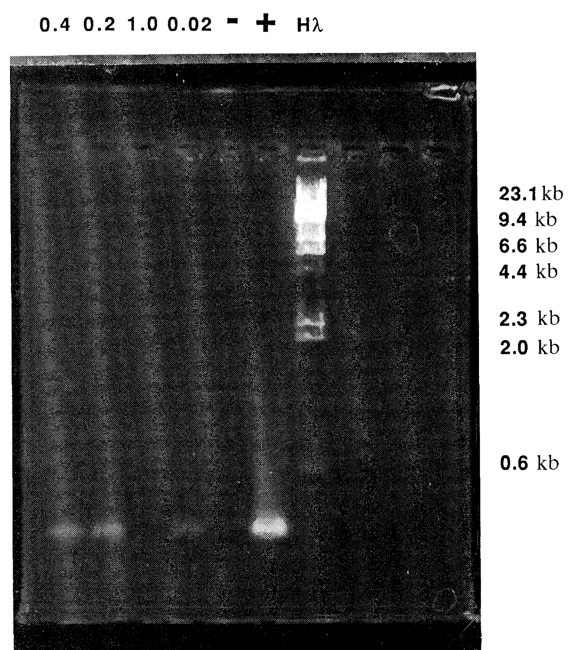


Figure 4. PCR amplification products from a dilution series experiment separated on an agarose gel. A sample of fossil *Cocculus* DNA was estimated to contain approximately 60 ng DNA per µl. Lane 1 used 0.4 µl template, lane 2 used 0.2 µl template, lane 3 used 1.0 µl template, lane 4 used 0.02 µl template, lane 5 used a negative control extract as a template and lane 6 used a positive control modern *Hordeum spontaneum* DNA as a template. All reactions used primers LtrbcL 478 and LtrbcL 724R. Other reaction conditions are identified to those described in the legend to figure 3.

(*d*) *Inhibiting factors*

Compounds that coprecipitate with DNA have been noted in the past to interfere with amplification of DNA sequences. Pääbo (1990) suggested either adding bovine serum albumin to overcome the inhibition of the polymerase enzyme or diluting the samples to reduce the effects of inhibiting agents.

Special considerations must be made when working with fossil plant material. First, depending on the species, plant tissues can contain large amounts of secondary metabolites that will be extracted along with the DNA. For example, species from the genus *Quercus* have high concentrations of inhibitors that prevent successful amplification irrespective of the condition of the DNA. Second, clay particles that are associated with the compression fossils may be inadvertently transferred in the process of DNA extraction. These particles may interfere with successful amplification by chelating the free Mg^{2+}.

Both conditions may be successfully counteracted by suitable dilution of the template DNA, but this will have conflicting effects on amplification. Reduction of inhibitor concentration will increase the rate of successful amplification, whereas reduction of the number of target DNA copies will decrease the concentration of products. Initial dilution series experiments may prove helpful in optimizing conditions for particular samples. In figure 4, PCR products from different concentrations of the fossil *Cocculus* DNA using *Liriodendron tulipifera*-based *rbc*L primers have been separated on an agarose gel. Concentration effects are evident, running from no product using undiluted template to reduced product in reactions containing the smallest initial concentrations of template. If DNA concentrations are initially too low to permit sufficient dilution, titration experiments with varying concentrations of $MgCl_2$ may prove useful.

(*e*) *DNA damage and repair*

The relative difficulty in amplifying fossil DNA seems to contradict the apparent high concentration of

Downloaded from rstb.royalsocietypublishing.org

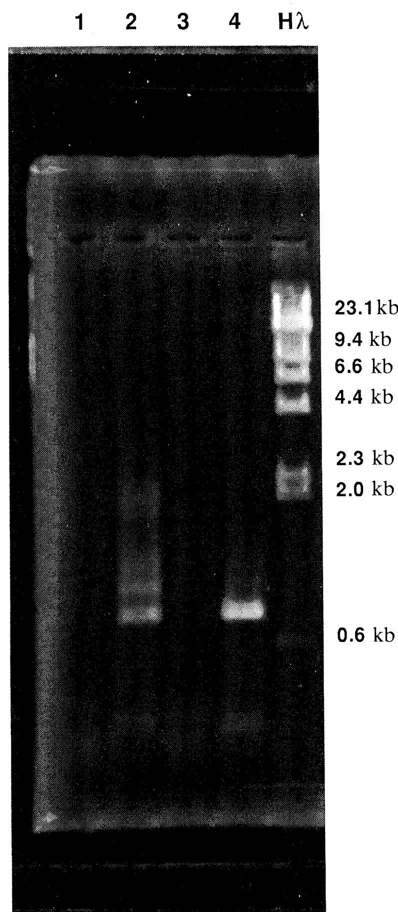


Figure 5. PCR amplification products testing pretreatment of fossil DNA. Lane 1 contains the products of the negative control, lane 2 the products of a positive control using modern *Hordeum spontaneum* DNA for a template, lane 3 the products of a reaction using untreated fossil *Cocculus* DNA as a template, lane 4 the products of a reaction using treated *Cocculus* DNA as a template. The reactions were initially run using primers LtrbcL 1 and LtrbcL 1201R (CCT AAA GTT CCT CCA CCG AAC TG). A portion (2 μl) of the reactions was then used in a second amplification using primers LtrbcL 209 and LtrbcL 955R. Reaction conditions are as above. Pretreatment of the fossil DNA includes a 20 min incubation of DNA with Klenow fragment of DNA polymerase I in the presence of dNTPs and buffer. The solution is heated to 70 °C to denature the enzyme followed by incubation at 37 °C with T_4 polynucleotide kinase and the addition of buffer for 1 h. The treatment is completed by addition of T_4 ligase incubated at room temperature for 2 h. The DNA is cleaned by a phenol–chloroform extraction and excess dNTPs are removed by passing through a G-50 Sephadex column.

high molecular mass DNA molecules. A possible explanation for this may involve the stabilizing structure of double-stranded DNA. Highly nicked strands of DNA may be held together by complementary fragments that span the nicked regions on the opposite strand. Heat denaturation would cause such molecules to break up into small single-stranded fragments thereby reducing the number of complete target fragments available. Attempts at amplification of such material will end with negative results.

This problem of fragmented single-stranded DNA can be addressed by prior *in vitro* repair of the template before PCR amplification. Rogan & Salvo (1991) suggested that the template DNA can be treated with the Klenow fragment of DNA polymerase I in an excess of dNTP to fill in gaps caused by nicks in the DNA. This may be successful in treating DNA of intermediate age, but very old and fossil DNA may be so highly nicked that even after filling in gaps the fragments may be too small to act as a template in amplification. To overcome this problem, I extended Rogan & Salvo's suggestion by coupling the Klenow-mediated polymerase reaction with a T_4 polynucleotide kinase reaction followed by a T_4 DNA ligase reaction. The rationale behind this treatment is that gaps may be filled in by the polymerase but would still be unlinked. The kinase reaction would assure that a 5′ phosphate was present and then the adjacent fragments would be ligated, similar to ligation of sticky ends when cloning. The results of such prior treatment are shown in figure 5. A 747 b.p. fragment is shown amplified in the treated fossil *Cocculus* sample (lane 4) but absent in the untreated sample from the same extraction.

It should be noted that illegitimate ligation between random fragments may occur and result in chimeric target sequences. This should not be a significant problem, however, as blunt end ligations are much less efficient than staggered or sticky end ligations. Secondarily, if coding regions of known expected length are targeted, fragments of anomalous lengths may be identified in advance.

6. CURRENT RESEARCH ON THE CLARKIA MIOCENE DNA IN OTHER LABORATORIES

Along with the results presented herein demonstrating the replication of amplification of Miocene Clarkia samples, a number of independent laboratories have initiated work on the Miocene Clarkia DNA. Some of these have offered to share their findings although, before publication, each result should be viewed as work in progress with final verification coming before publication.

Douglas and Pamela Soltis of Washington State University have reported successful amplification and sequencing of a fragment of the chloroplast gene *rbc*L from *Taxodium*. The sequence clusters well with modern *Metasequoia*, as would be expected from *a priori* systematic considerations. Similarly, Holly A. Wichman and Meredith Hamilton of the University of Idaho have reported amplifying a fragment of *rbc*L from a sample extracted from a Miocene *Quercus* sample. As with the *Cocculus* amplifications reported herein, final verification of the origin of the amplified fragment must await the sequencing of the DNA.

In the investigation of the distribution and evolution of copia-like retrotransposons in plants, Michael Cummings from Harvard University, Daniel Voytas and Steven Rodermel of Iowa State University, and Andrzej Konieczny and Frederick Ausubel of Massachusetts General Hospital have looked for retrotransposon sequences in modern and fossil material. Initial results indicate amplification products in six of ten fossil samples. Preliminary sequencing has been done

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

on Miocene *Liriodendron* and *Platanus* amplified products. If these initial findings are verified, this will be an indication that small copy number nuclear genes may be recovered.

In conclusion, recent results indicate that Miocene DNA can be extracted and analysed to elucidate problems in molecular evolution. The richness of the Clarkia site, opening up the possibility of research in rates and modes of evolution, fossil identification and molecular systematics, and patterns of biogeographic change, is potentially just a sampling of the material and research that will be done in the future at other, perhaps even older sites. The work accomplished at Clarkia will hopefully close the circle in the study of evolutionary biology and begin an active dialogue between palaeontologists and evolutionary geneticists.

I acknowledge the help and inspiration of Michael T. Clegg and David E. Giannasi in the work on the Miocene Clarkia plant DNA. I am indebted to C. J. Smiley, Yang Hong, Mary Durbin and David Henderson for their help in the field at Clarkia. I thank Francis and Vicki Kinbaum of Clarkia, Idaho, and the Tertiary Research Center at the University of Idaho for access to the fossil site and the material therein. I also thank my colleagues Douglas Soltis, Holly Wichman, Michael Cummings and Svante Pääbo for sharing their unpublished results, and William Moore for reading an earlier version of this manuscript.

REFERENCES

Clegg, M. T., Ritland, K. & Zurawski, G. 1986 Processes of chloroplast DNA evolution. In *Evolutionary processes and theory* (ed. S. Karlin & E. Nevo), pp. 275–294. New York: Academic Press.

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: Pacific Division of the American Association for the Advancement of Science.

Golenberg, E. M., Giannassi, 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.

Hagelberg, E., Sykes, B. & Hedges, R. 1989 Ancient bone DNA amplified. *Nature, Lond.* **342**, 485.

Helentjaris, T. 1988 *Maize Genet. Coop. News Lett.* **62**, 104–105.

Higuchi, R., Bowman, B., Friedberger, M., Ryder, O. A. & Wilson, A. C. 1984 DNA sequences from the quagga, an extinct member of the horse family. *Nature, Lond.* **312**, 282–284.

Kwok, S. & Higuchi, R. 1989 Avoiding false positives with PCR. *Nature, Lond.* **339**, 237–238.

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

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

Niklas, K. J., Brown, R. M. Jr & Santos, R. 1985 Ultrastructural states of preservation in Clarkia angiosperm leaf tissues: implications on modes of fossilization. In *Late Cenozoic history of the Pacific Northwest* (ed. C. J. Smiley), pp. 143–160. San Francisco: Pacific Division of the American Association for the Advancement of Science.

Pääbo, S. 1985*a* Molecular cloning of ancient Egyptian mummy DNA. *Nature, Lond.* **314**, 644–645.

Pääbo, S. 1985*b* Preservation of DNA in ancient Egyptian mummies. *J. Archeol. Sci.* **12**, 411–417.

Pääbo, S. 1989 Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1939–1943.

Pääbo, S. 1990 Amplifying ancient DNA. In *PCR protocols: a guide to methods and applications* (ed. M. A. Innis, D. H. Gelfand, J. J. Sninsky & T. J. White), pp. 159–166. New York: Academic Press.

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., Higuchi, R. G. & Wilson, A. C. 1989 Ancient DNA and the polymerase chain reaction. *J. biol. Chem.* **264**, 9709.

Pääbo, S., Irwin, D. M. & Wilson, A. C. 1990 DNA damage promotes jumping between templates during enzymatic amplification. *J. biol. Chem.* **265**, 4718–4721.

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

Rogan, P. K. & Salvo, J. J. 1991 Molecular genetics of pre-Columbian South American mummies (In preparation.)

Rogers, S. O. & Bendich, A. J. 1985 Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Pl. molec. Biol.* **5**, 69–76.

Sarkar, G. & Sommer, S. S. 1990 More light on PCR contamination. *Nature, Lond.* **343**, 27.

Smiley, C. J. (ed.) 1985 *Late Cenozoic history of the Pacific Northwest.* San Francisco: Pacific Division of the American Association for the Advancement of Science.

Smiley, C. J. & Rember, W. C. 1985 Physical settings of the Miocene Clarkia fossil beds, northern Idaho. In *Late Cenozoic history of the Pacific Northwest* (ed. C. J. Smiley), pp. 11–31. San Francisco: Pacific Division of the American Association for the Advancement of Science.

Discussion

T. A. Brown (*Department of Biochemistry and Applied Molecular Biology, UMIST, U.K.*). Has Professor Golenberg sequenced the PCR products from fossil DNA repaired by Klenow polymerase treatment?

E. M. Golenberg. No. I have not sequenced these products as yet.

T. A. Brown. Some retrotransposons in plants have very large copy numbers, e.g. del of lily has a copy number of greater than 1000.

E. M. Golenberg. The retrotransposons amplified from plant fossils are of a type that have low copy numbers in modern plants.

J. L. Bada (*Scripps Institution of Oceanography, University of California at San Diego, U.S.A.*). The preservation of DNA for periods of tens of millions of years seems inconsistent with our understanding of the geochemical stability of proteins. The phosphodiester bond is much less stable than the peptide bond. In general we believe that proteins and peptides are unlikely to survive more than a few million years on the Earth's surface. Professor Golenberg's example that the 'protein rich' fish are poorly preserved supports this. This generates a paradox of why there should be some sort of process that preserves labile DNA and not protein.

E. M. Golenberg. Yes, the paradox remains. It is worth

Downloaded from rstb.royalsocietypublishing.org

noting, however, that the plant material is much better preserved. It is important to determine which factors, that are particular to plant tissue, promote preservation, whether it be the cell wall, the chloroplast membrane, the phenolics and tannins, or the process of cell senescence and death.

P. Westbroek (*Department of Biochemistry, University of Leiden, The Netherlands*). DNA is a polyanion that obviously can bind to mineral surfaces and be stabilized. Would Professor Golenberg agree that to study the behaviour in such potentially preservational microenvironments may be the most significant task in the short term for this field?

E. M. Golenberg. Yes, the interaction of DNA with minerals may be important in the preservation of the DNA. Whether such minerals can stabilize the DNA within the cell structures certainly would be of great interest.

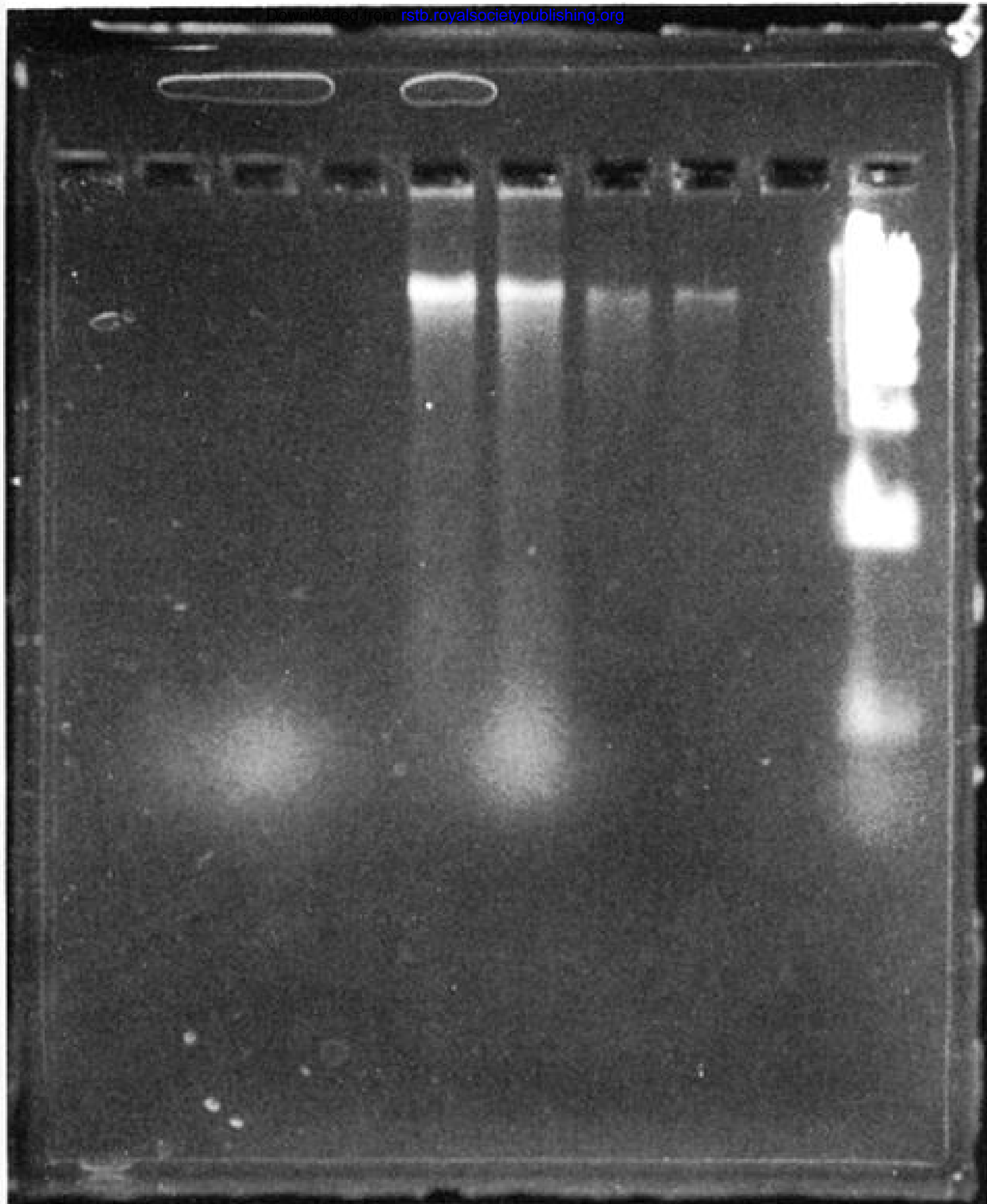
M. J. Bishop (*MRC Molecular Genetics Unit, Hills Road, Cambridge, U.K.*). Since 1965 attempts have been made to include time in descriptions of molecular evolution for the purpose of reconstructing phylogeny. 'Molecular clocks' are metaphors for models of molecular evolution as stochastic processes for the changes in DNA or protein sequences. The simplest models are based on a Poisson process for nucleotide changes. We now know that there are strong but poorly understood constraints on DNA sequence overlapping dinucleotide composition even for different regions of the same genome (the isochores). Models can be refined as our knowledge improves. The stumbling-block to date has been that there was no way to test if the data fit the model being used because of the lack of historical sequence information. The ability to obtain DNA sequences from fossil material will enable us directly to study the processes of molecular evolution which are going on in different parts of genomes. Similarly, human population genetics performed by the comparison of DNA sequences from historic and prehistoric bone material will revolutionize our understanding of the origins and migrations leading to modern man.

G. A. Dover (*Department of Genetics, University of Cambridge, U.K.*). I agree with both the introductory remarks made by Professor Golenberg on the need to bring population genetics closer to palaeontology and systematics, and with the point of discussion raised by Dr Martin Bishop that the uncovering of very old 'dead' DNA might ultimately lead to a better understanding of the genomic and populational processes and constraints at play on the evolution of DNA. To do this, however, it would be advisable to choose for study an evolutionary series of intermediate stages between ancient and modern taxa, if this is at all possible; and then to monitor the molecular evolution of at least two parallel genetic systems in such a series. Although a lot of useful public attention is drawn to studies of quagga horses, Tasmanian wolves, frozen mammoths, Egyptian mummies and other exotica, such findings do not necessarily in themselves tell us much about the fine-grained processes at work in molecular evolution and organismal divergence. We need to move away from the 'stamp collecting' phase of molecular palaeontology to a more systematic survey.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

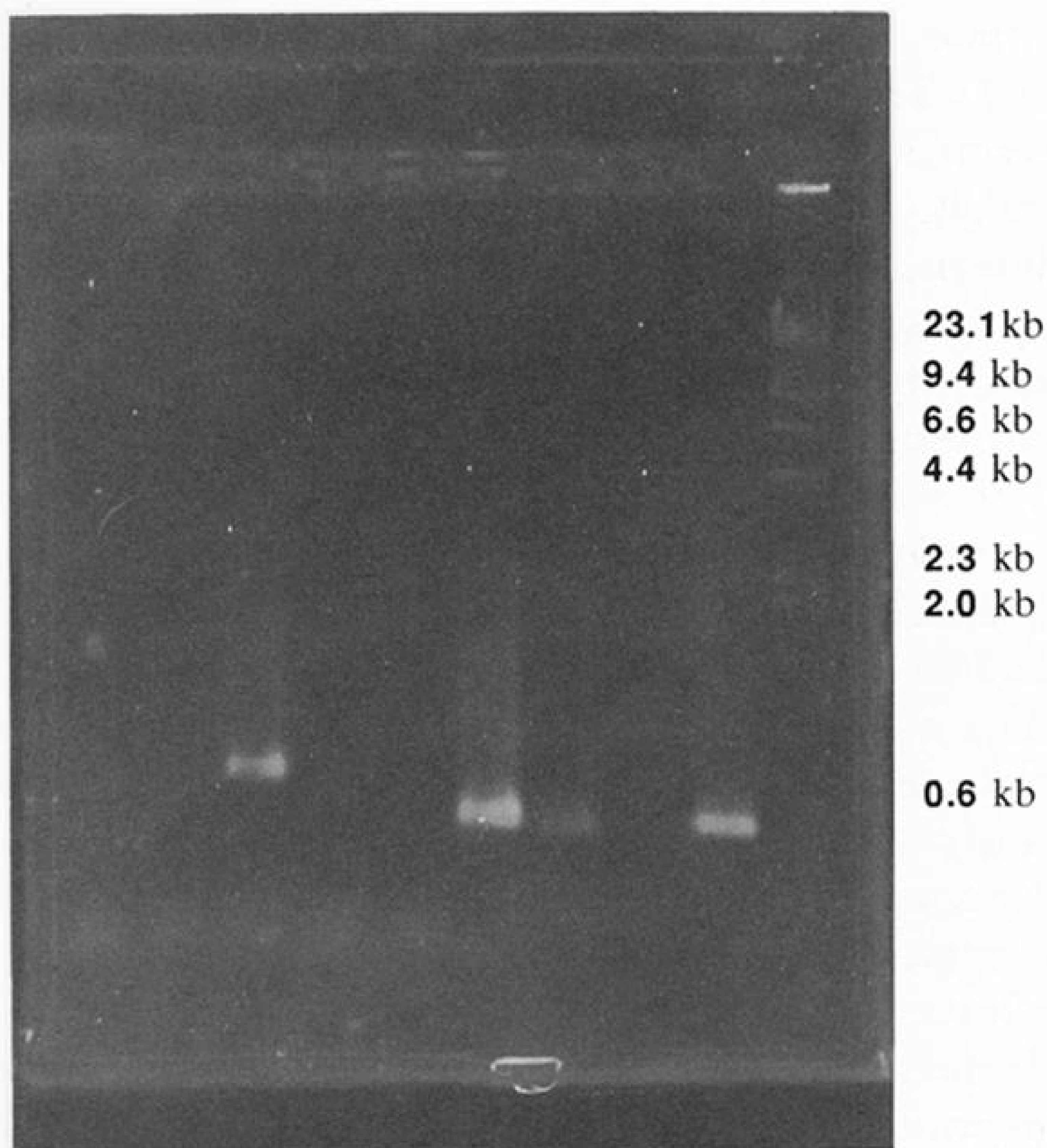
Downloaded from rstb.royalsocietypublishing.org



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

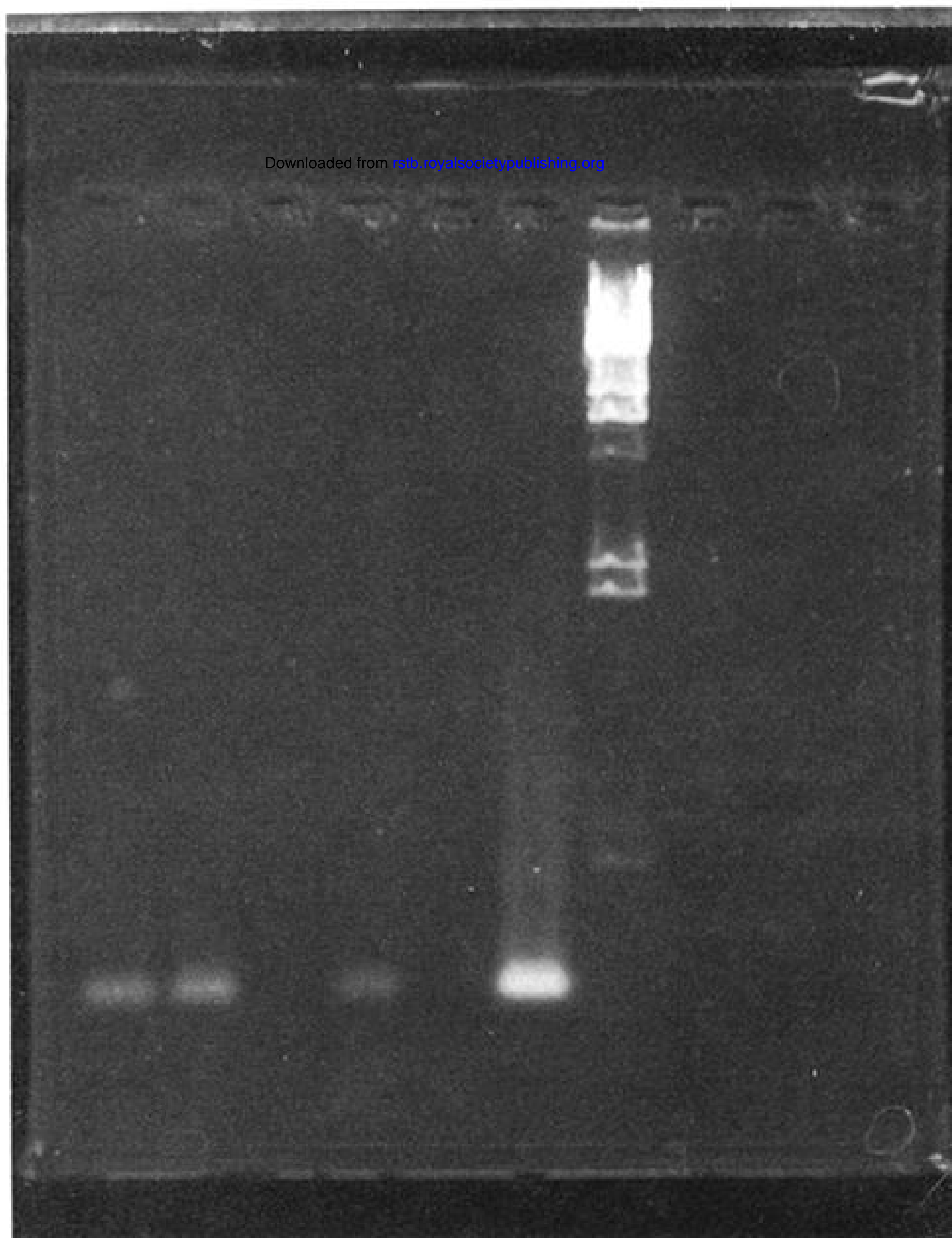
gure 1. Fossil DNA extractions separated on an agarose gel. ane 1 is a negative control extraction using the extraction uffer without sample. Lanes 2 and 3 are split samples from fossil *Persea* leaf. Lane 3 is a sample from a fossil *Magnolia* af. Lanes 5 and 6 are split samples from a *Cocculus* leaf. anes 7 and 8 are split samples from a separate *Cocculus* leaf id lane 9 is a sample from holding buffer of a *Cocculus* seed mple before grinding of the tissue. Lane 10 contains mbda phage DNA digested with *Hin*d III as size markers.



Downloaded from rstb.royalsocietypublishing.org

Figure 3. PCR amplification products separated on an agarose el. Lanes 1, 4, and 7 are products of reactions using fossil *'occulus* DNA as a template. Lanes 2, 5, and 8 are products rom a negative control using extractions from extraction uffer without sample as a template. Lanes 3, 6, and 9 are roducts from positive control reactions using modern *Iordeum spontaneum* DNA as template. All primers used are ased on the *Liriodendron tulipifera rbc*L sequence. Lanes 1 to are from reactions using primers LtrbcL 1 (ATG TCA CA CAA ACA GAG ACT AAA GC) and LtrbcL 724R CTT CGC ATG TAC CTG CAG TAG C); lanes 4 to 6 re from reactions using primers LtrbcL 209 (GGA CCG TG GAC TTA CCA GCC TTG ATC G) and LtrbcL 24R; lanes 7 to 9 are from reactions using primers LtrbcL 78 (GAT AAA TTG AAC AAG TAT GGT CGT CC) and trbcL 955R (GAG TGA ATA TGA TCT CCA CCA GAC TA CG). The reactions were run in 100 µl volumes using 0 pmol of each primer with the standard (Promega Taq olymerase) reaction buffer. Before the addition of the emplate DNA and enzyme, the solutions were exposed to hort-wave UV light for 3–5 min. Two µl (approximately 2 to ng) of template DNA were used. Before addition of enzyme, he solution was incubated at 94 °C for 2 min. The enzyme vas added while maintaining the solutions at 80 °C or igher. The heat settings were as follows: four cycles of 94 °C 5 min), 38 °C (2 min), 74 °C 3 (min); followed by 11 cycles f 94 °C (1.5 min), 42 °C (2 min), 74 °C (3 min); followed by 0 cycles of 94 °C (1.5 min), 48 °C (2 min), 74 °C (3 min) nd then 15 min at 74 °C.

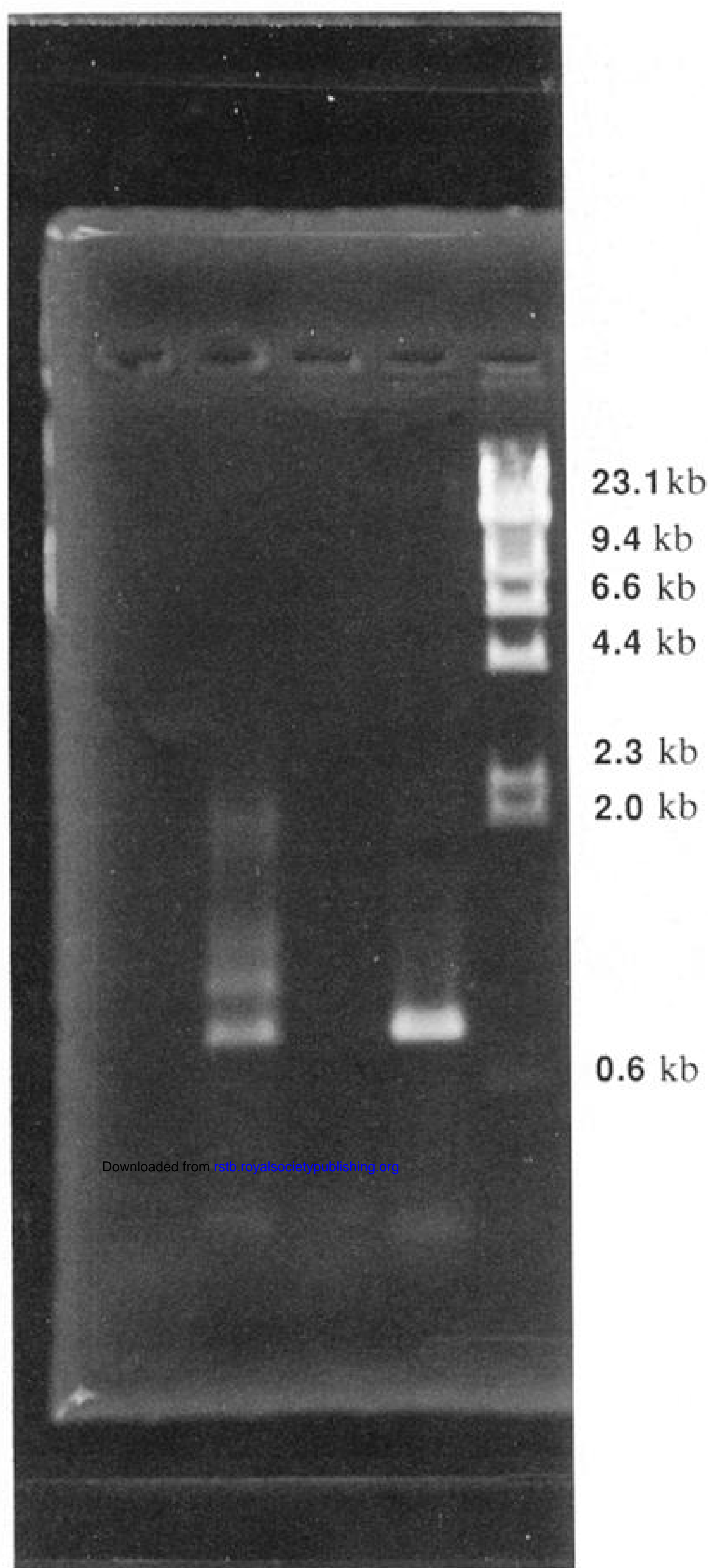
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES



Downloaded from rstb.royalsocietypublishing.org

gure 4. PCR amplification products from a dilution series
periment separated on an agarose gel. A sample of fossil
cculus DNA was estimated to contain approximately 60 ng
NA per µl. Lane 1 used 0.4 µl template, lane 2 used 0.2 µl
mplate, lane 3 used 1.0 µl template, lane 4 used 0.02 µl
mplate, lane 5 used a negative control extract as a template
d lane 6 used a positive control modern *Hordeum spontaneum*
NA as a template. All reactions used primers LtrbcL 478
d LtrbcL 724R. Other reaction conditions are identified to
ose described in the legend to figure 3.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES



Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

gure 5. PCR amplification products testing pretreatment of ›sil DNA. Lane 1 contains the products of the negative ntrol, lane 2 the products of a positive control using odern *Hordeum spontaneum* DNA for a template, lane 3 the oducts of a reaction using untreated fossil *Cocculus* DNA as template, lane 4 the products of a reaction using treated *cculus* DNA as a template. The reactions were initially run ing primers LtrbcL 1 and LtrbcL 1201R (CCT AAA GTT CT CCA CCG AAC TG). A portion (2 μl) of the reactions as then used in a second amplification using primers LtrbcL 9 and LtrbcL 955R. Reaction conditions are as above. etreatment of the fossil DNA includes a 20 min incubation DNA with Klenow fragment of DNA polymerase I in the esence of dNTPs and buffer. The solution is heated to °C to denature the enzyme followed by incubation at °C with T_4 polynucleotide kinase and the addition of iffer for 1 h. The treatment is completed by addition of ligase incubated at room temperature for 2 h. The DNA cleaned by a phenol–chloroform extraction and excess NTPs are removed by passing through a G-50 Sephadex lumn.