Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 449–451, January 1992
Plant Biology

An *rbcL* sequence from a Miocene *Taxodium* (bald cypress)

(molecular evolution/PCR/phylogeny/fossil DNA)

PAMELA S. SOLTIS*†, DOUGLAS E. SOLTIS*, AND CHARLES J. SMILEY‡

*Department of Botany, Washington State University, Pullman, WA 99164-4238; and ‡Department of Geology and Geological Engineering, University of Idaho, Moscow, ID 83843

Communicated by David Dilcher, October 1, 1991 (received for review August 20, 1991)

ABSTRACT **During the past decade, ancient DNAs from both animals and plants have been successfully extracted and analyzed. Recently, the age of DNA that can be recovered and sequenced was increased manyfold by the amplification and sequencing of a DNA fragment from a *Magnolia* fossil obtained from the Miocene Clarkia deposit (17–20 million yr old). However, the validity of this report has been questioned based on models predicting that DNA should be completely degraded after 4 million yr. We report here the successful amplification, sequencing, and analysis of a 1320-base-pair portion of the chloroplast gene *rbcL* from a Miocene *Taxodium* specimen, also from the Clarkia site. These data not only validate the earlier report of sequence data for a *Magnolia* species from the same site but also suggest that it may be possible to isolate and sequence DNAs routinely from the Clarkia deposit. The ability to recover and sequence DNAs of such age offers enormous research possibilities in the areas of molecular evolution, biogeography, and systematics.**

Many areas of evolutionary inquiry rely entirely on comparison among living organisms to infer past events. Evolutionary studies of divergence rates, biogeography, and phylogenetic relationships would benefit greatly from molecular data obtained from extinct species. Molecular evolutionists have, therefore, sought sources of ancient DNA to calibrate molecular studies of extant organisms.

During the past decade, the analysis of ancient DNAs has met with considerable success, and ancient DNAs have been extracted from museum specimens (e.g., ref. 1), Egyptian mummies (2, 3), and preserved animal tissues (4). A 13,000-yr-old ground sloth is the oldest preserved animal tissue that has yielded DNA. Many of these preserved animal DNAs have been so extensively damaged, however, that they have not yielded sequence data (but see ref. 4). Several sources of ancient plant material, some of which date from 44,600 yr ago, have yielded analyzable and, in some cases, high-molecular-weight DNA (5).

A major breakthrough in the study of ancient DNAs was the amplification and sequencing of an 820-base-pair (bp) portion (of ≈1431 bp total) of *rbcL*, a chloroplast gene that encodes the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase, from a Miocene species of *Magnolia* (6). DNA of the now-extinct *Magnolia latahensis* was obtained from a 17- to 20-million-yr-old fossil found at the Clarkia fossil site in northern Idaho, demonstrating that the age of DNA that can be successfully isolated and sequenced is considerably greater than previous attempts indicated.

The validity of the fossil *Magnolia* sequence (6) has been questioned, however, primarily on theoretical grounds (7). *In vitro* studies of DNA degradation (8, 9) formed the foundation of Pääbo and Wilson's calculations (7) that DNA should be completely degraded after 4 million yr. Pääbo and Wilson (7) further questioned the potential value of the Clarkia fossils by suggesting that even were DNA occasionally isolated from a Miocene fossil and analyzed, repeatability of this feat was so low as to render most attempts not worth the effort.

Here we report the successful amplification, sequencing, and analysis of a 1320-bp portion of *rbcL* from a fossil *Taxodium* (bald cypress) specimen from the Clarkia site. All available evidence indicates that this fossil deposit dates from the Miocene, 17–20 million yr ago (10). These data not only help validate the earlier report of *rbcL* sequence data for a fossil *Magnolia* from the Clarkia site but also suggest that Miocene fossils from this site hold real potential for future studies in molecular evolution and systematics.

MATERIALS AND METHODS

DNA was isolated from a fossil of *Taxodium* immediately after the specimen was exposed at the Clarkia site, following the procedure described by Golenberg *et al.* (6). The DNA sample was then diluted 1:50 with water. Chloroplast DNAs of extant *Taxodium distichum* and *Metasequoia glyptostroboides*, also in the Taxodiaceae, were isolated from fresh leaf tissue following a modification (11) of Palmer (12).

The double-stranded *rbcL* gene was amplified from all three DNA samples using the PCR and the protocol provided by the supplier of *Taq* DNA polymerase (Promega). PCR primers were constructed from the *rbcL* sequence of *Zea mays* (13). The 5′ primer (Z1) was composed of the first 30 bases of *rbcL* of *Z. mays* (5′-ATGTCACCACAAACAGAAACTAAAGCAAGT-3′); the 3′ primer (Z1351R), also a 30-mer, began at position 1351 on the reverse strand of the *rbcL* sequence of *Z. mays* (5′-CTTCACAAGCAGCAGCTAGTTCAGGACTCC-3′). Single-stranded PCR amplifications using the double-stranded product as template and the two primers individually (14) produced DNA for sequencing.

The single-stranded DNAs were sequenced using the dideoxynucleotide chain-termination method and the Sequenase system (United States Biochemical). Sequencing primers were derived from the *rbcL* sequence of *Z. mays* (13, 15) and provided by G. Zurawski (DNAX). Each strand was sequenced two or three times to confirm the sequences.

Relationships among the three sequences were analyzed phylogenetically, along with a published *rbcL* sequence for *Pseudotsuga menziesii* (Pinaceae; the only other conifer for which *rbcL* data have been published; ref. 16) and a published sequence for *Marchantia polymorpha* (17), a liverwort, which was used to root the tree. Phylogenetic analyses were performed using the program DNAPARS, a parsimony algorithm of the PHYLIP package, version 3.22 (J. Felsenstein, University of Washington), and DNABOOT, a parsimony algorithm employing the bootstrap (PHYLIP version 3.22), with 100 replicates. A distance matrix was computed for the five DNA sequences using Kimura's (18) two-parameter model and PHYLIP program DNADIST (version 3.2). These distances

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†To whom reprint requests should be addressed.

were then used to construct a tree following the method of Fitch and Margoliash (19), as performed by PHYLIP program FITCH (version 3.2). A maximum-likelihood model (20) using PHYLIP program DNAML (version 3.2) was also used to evaluate relationships among the five sequences.

RESULTS AND DISCUSSION

A DNA fragment of 1380 bp, representing nearly the entire 1431-bp *rbcL* gene, was amplified from the fossil *Taxodium* DNA using the *rbcL* primers; 1320 bp of this fragment were derived from the DNA of *Taxodium*, the remaining bp deriving from the primers. The size of this fragment, over 500 bp longer than the portion of *rbcL* amplified and sequenced by Golenberg *et al.* (6), illustrates the remarkable preservation of the Clarkia specimens.

The *rbcL* sequence from the fossil *Taxodium* sample was confirmed as a Miocene *Taxodium* sequence by using both molecular and phylogenetic analyses (after refs. 1, 6, and 21). The fossil *rbcL* sequence differs from the extant *Taxodium* sequence by 11 base substitutions; both *Taxodium* sequences differ from the *Metasequoia* sequence by 38 substitutions (Fig. 1).

Phylogenetic analysis of *rbcL* sequences from fossil *Taxodium*, extant *T. distichum*, *Metasequoia glyptostroboides*, *P. menziesii*, and *Marchantia polymorpha* demonstrated the similarity of the two *Taxodium* sequences and their similarity to the sequence of the confamilial *Metasequoia glyptostroboides*. All nodes on the tree were supported in all 100 of the bootstrap replicates (Fig. 2). The identical tree topology was obtained with the Fitch–Margoliash method and maximum-likelihood procedure. The uniqueness of the fossil sequence among all *rbcL* sequences obtained in our laboratory plus its similarity to the extant *Taxodium* sequence argue convincingly against it being a contaminant from another DNA source.

Molecular divergence between the fossil and extant *Taxodium rbcL* sequences has been relatively limited, compared with the amount of divergence between *rbcL* sequences of Miocene fossil *Magnolia latahensis* and extant *Magnolia macrophylla* (6). We detected only 11 substitutions (all transitions and silent third-position substitutions; Fig. 1) of 1320 positions (0.83% divergence) between the two *Taxodium* sequences, whereas the *rbcL* sequences of *Magnolia latahensis* and *Magnolia macrophylla* differed by 17 substitutions of 759 positions examined (2.2% divergence; ref. 6). Although the time since isolation of the populations from which the *Taxodium* samples were obtained is uncertain, it is at least 17–20 million yr and possibly longer. If the two Taxodiums actually shared a common ancestor 17–20 million yr ago, this amounts to a minimum sequence divergence rate of 0.55–0.65 substitution per million yr over nearly the entire length of the gene, or 4.2×10^{-4} to 4.9×10^{-4} substitution per site per million yr. During the early to mid-Miocene epoch, *Taxodium* had a distribution that extended from eastern Asia across Beringia and across much of North America; its range was not reduced until the early Quaternary period (22). Despite its extensive range in North America, populations from northern Idaho and eastern North America, from whence our sample of extant *Taxodium* was obtained, have probably been isolated longer than 17 to 20 million yr, suggesting an even slower rate of nucleotide substitution. The lesser divergence between *Taxodium* sequences relative to those of fossil and extant *Magnolia* species may reflect the apparent conspecific status of the *Taxodium* samples. Although several morphological features distinguish extant *T. distichum* from Miocene *Taxodium* from northern Idaho, the fossil specimens have not been recognized taxonomically as a distinct species (C.J.S., unpublished data). Alternatively, there may be different rates of sequence evolution in the *Magnolia* and *Taxodium* lineages.

```
     1
Tax  NTGTCACCAC AAACAGAAAC TAAAGCAAGT GTTGGATTCA AAGCTGGTGT TAAAGATTAT
Fos  .......... .......... .......... .......... .......... .........C
Met  .N........ .......... .......... ..C....... .......... .........C

     61
Tax  AAATTGACTT ACTACACCCC GGAATACGAA ACCAAAGATA CTGATATCTT GGCAGCATTC
Fos  .......... .T..T..T.. ......T... .......... .......... ..........
Met  .G...A.... .T..T..T.. ........C. .......... .......... ..........

     121
Tax  CGAGTAACTC CTCAGCCTGG AGTTCCGCCT GAAGAAGCAG GAGCAGCAGT AGCTGCCGAA
Fos  .......... ....A..... ......A... .......... .......... ..........
Met  .....C.... .G........ ...A..C..C ........G. .......... ..........

     181
Tax  TCTTCTACTG GTACATGGAC AACTGTTTGG ACCGATGGAC TTACCAGTCT TGATCGTTAC
Fos  .......... .......... .......... .......... .......... ..........
Met  .....G.... ....G..... C......... .......... .......... ..........

     241
Tax  AAAGGACGAT GCTATGATAT TGAACCCCTT CCTGGAGAGG AAAGTCAATT TATTGCCTAT
Fos  .......... .......... .......... .......... .......... ..........
Met  ..G..C.... ....C..... .......G.. .......... ...A...... ..........

     301
Tax  GTAGCTTACC CTTTAGATCT TTTTGAAGAA GGTTCTGTTA CTAACCTGTT CACTTCTATT
Fos  .......... .......... .......... .......... .......... ..........
Met  .......... .C........ .......... .......... .......... ..........

     361
Tax  GTAGGTAATG TATTTGGATT CAAAGCCTTA CGGGCTCTAC GTCTGGAAGA TCTACGAATT
Fos  .......... .......... .......... .......... .......... ..........
Met  .......... .......... .......... .......... .......... .T........

     421
Tax  CCTCCTGCTT ATTCAAAAAC TTTCCAAGGC CCACCACATG GTATTCAAGT AGAAAGAGAT
Fos  .......... .......... .......... .......... .......... ..........
Met  .......... .......... .........A .......... .......... G.........

     481
Tax  AAATTAAACA AGTATGGTCG TCCTTTGTTG GGATGTACTA TAAAACCAAA ATTGGGTCTA
Fos  .......... .......... ......A... .......... .......... ..........
Met  .......... .A........ .......... .......... .C........ ..........

     541
Tax  TCTGCCAAGA ATTACGGTAG AGCGGTTTAT GAATGTCTCC GTGGTGGACT TGATTTTACC
Fos  .......... ....T..... .......... .......... .......... ..........
Met  .......... ....T..... .......... .......... .......... ..........

     601
Tax  AAGGATGATG AAAACGTGAA CTCCCAACCA TTTATGCGCT GGAGAGATCG TTTCTGCTTT
Fos  .......... .......... .......... ........T. .......... ..........
Met  .......... .......... T......... .......... .......... ..........

     661
Tax  TGTGCAGAAG CAATTTATAA AGCTCAGGCT GAGACGGGTG AGATTAAGGG ACATTACCTG
Fos  .......... .......... .......... .......... .......... ..........
Met  .......... ..C....... .......... .......... .......... ..........

     721
Tax  AATGCTACTG CAGGTACATG TGAAGAAATG ATGAAAAGAG CAATATTCGC CAGAGAATTG
Fos  .......... .......... .......... .......... .......... ..........
Met  .......... .......G.. .......... .......... ..G....... ..........

     781
Tax  GGAGTTCCTA TAGTCATGCA TGACTATCTG ACAGGAGGTT TTACGGCAAA TACTTCGTTG
Fos  .......... .......... .......... .......... .......... ..........
Met  .......... .C........ .......... ..T....... .......... ..........

     841
Tax  GCTCATTATT GCCGAGATAA CGGCCTACTT CTTCACATTC ACCGCGCAAT GCATGCAGTT
Fos  .......... .......... .......... .......... .......... ..........
Met  .......... .......... .......... .......... .......... ...C......

     901
Tax  ATTGATAGAC AAAGAATTCA TGGTATGCAC TTCCGTGTAC TGGCTAAAGC ACTACGTATG
Fos  .......... .......... .......... .......... .A........ ..........
Met  .....C.... ......A... .......... .......... .......... ..........

     961
Tax  TCTGGTGGAG ATCATATTCA CGCTGGTACT GTAGTAGGTA AACTTGAAGG AGAACGAGAA
Fos  .......... .......... .......... .......... .......... ..........
Met  .......... .......... .......... .......... .......... ..........

     1021
Tax  GTCACTTTGG GTTTTGTTGA TCTATTGCGT GATGATTTTA TTGAAAAAGA CCGAAGTCGT
Fos  .......... .......... .......... .......... .......... ..........
Met  .......... .......... ......A... .......... .......... ..........

     1081
Tax  GGTATTTATT TCACTCAAGA TTGGGTCTCT ATGCCGGGTG TCCTGCCTGT AGCTTCAGGA
Fos  .......... .......... .......... .......... .......... ..........
Met  .......... .......... .......... .......... .TT....... ..........

     1141
Tax  GGTATTCACG TTTGGCATAT GCCTGCTCTG ACCGAGATCT TTGGGGATGA TTCCGTATTA
Fos  .......... .......... .......... .......... .......... ..........
Met  .......... .......... .......... .......... .......... ..........

     1201
Tax  CAGTTTGGTG GAGGGACTTT GGGGCACCCT TGGGGAAATG CACCTGGTGC AGTGGCTAAC
Fos  .......... .......... .......... .......... .......... ..........
Met  .......... ....A..... .......... .......... .......... ..........

     1261
Tax  CGGGTCGCTT TAGAAGCTTG TGTACAAGCT CGTAATGAAG GACGTGATCT TGCGCGTGAA
Fos  .......... .......... .......... .......... .......... ..........
Met  ..A....... .......... .........C .......... .......... ..........

     1321
Tax  GGTAATGAAG TGATCCGCGA AGCTACTAAA TGGAGTCCTG AACTA
Fos  .......... .......... .......... .......... .....
Met  .......... .......... .......... .......... .....
```

FIG. 1. *rbcL* sequences of extant *T. distichum* (Tax), Miocene *Taxodium* (Fos), and extant *Metasequoia glyptostroboides* (Met). A "." in the lines corresponding to Fos and Met indicates nucleotide identity with the sequence of extant *T. distichum*. N, undetermined nucleotide.

The *rbcL* sequence of *Metasequoia* differs from the shared sequence of both *Taxodium* samples by 24 transitions and 14

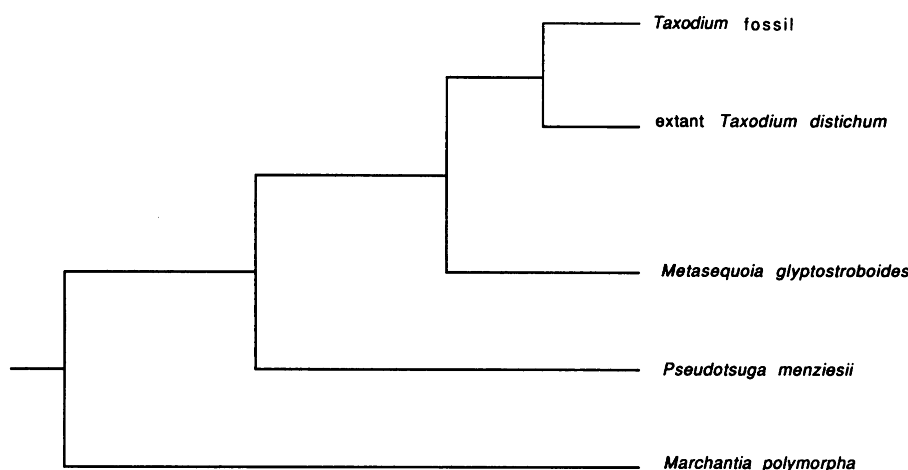

FIG. 2. Phylogenetic tree of *rbcL* sequences of fossil *Taxodium*, extant *T. distichum*, *Metasequoia glyptostroboides*, *P. menziesii*, and *Marchantia polymorpha*. Identical tree topologies were obtained using parsimony, similarity, and maximum-likelihood procedures.

transversions across all codon positions (Fig. 1), for a transition/transversion ratio of 1.71, compared with the average transition/transversion ratio of 1.35 observed for *rbcL* sequences by Ritland and Clegg (23). The rates of substitution by codon position deviate slightly from those expected for chloroplast DNA gene sequences (23). We detected five first-position substitutions (two synonymous, three nonsynonymous), four nonsynonymous substitutions in the second position, and 29 synonymous substitutions in the third position. This contrasts with the approximate 2:1:13 ratio predicted for first-, second-, and third-position substitutions from both *rbcL* and *atpB* sequences (23), suggesting a slightly increased rate of second-position substitutions in the Taxodiaceae relative to the sequences evaluated by Ritland and Clegg (23).

The successful amplification and sequencing of *rbcL* from a Miocene *Taxodium* confirms the possibility of analyzing DNA dating from 17–20 million yr ago (6) and also demonstrates the repeatability of such an analysis. Furthermore, the Clarkia flora, represented by three ecological communities (24), is highly diverse taxonomically, providing opportunities for research on a wide variety of conifers and angiosperms. We have also amplified double-stranded DNA products corresponding in size to *rbcL* from specimens of *Platanus* and *Pseudofagus*, both from the Clarkia site, whereas efforts to amplify *rbcL* from *Metasequoia, Hydrangea*, and *Lindera* have thus far failed. Although the products of the former have not yet been sequenced, their successful amplification suggests that, in contrast to the skepticism raised by Pääbo and Wilson (7), the Miocene fossils from northern Idaho may hold great potential for studies of molecular evolution, biogeography, and phylogenetic relationships, not only of plants, but perhaps also of the vertebrate and insect faunas represented at the Clarkia site.

We thank M. Clegg, D. Crawford, E. Golenberg, G. Learn, and an anonymous reviewer for valuable comments on the manuscript, M. Durbin for technical advice, and S. Novak, A. Grable, T. Tucker, and M. Edgerton for technical assistance. S. Brunsfeld and T. Ranker provided chloroplast DNAs of extant *T. distichum* and *Metasequoia glyptostroboides*, and G. Zurawski (DNAX) provided primers for PCR amplification and sequencing. We thank F. and V. Kienbaum, Clarkia, ID, for permission to work on their property and for their enthusiasm over the fossils. The contributions of the Tertiary Research Center, University of Idaho, are also acknowledged. This research was supported, in part, by National Science Foundation Grants BSR-8918247 and BSR-9007614.

1. Higuchi, R., Bowman, B., Friedberger, M., Ryder, O. A. & Wilson, A. C. (1984) *Nature (London)* **312**, 282–284.
2. Pääbo, S. (1985) *Nature (London)* **314**, 644–645.
3. Pääbo, S. (1985) *J. Arch. Sci.* **12**, 411–417.
4. Pääbo, S., Gifford, J. A. & Wilson, A. C. (1988) *Nucleic Acids Res.* **16**, 9775–9787.
5. Rogers, S. O. & Bendich, A. J. (1985) *Plant Mol. Biol.* **5**, 69–76.
6. Golenberg, E. M., Giannasi, D. E., Clegg, M. T., Smiley, C. J., Durbin, M., Henderson, D. & Zurawski, G. (1990) *Nature (London)* **344**, 656–658.
7. Pääbo, S. & Wilson, A. C. (1991) *Curr. Biol.* **1**, 45–46.
8. Lindahl, T. & Andersson, A. (1972) *Biochemistry* **11**, 3618–3622.
9. Lindahl, T. & Nyberg, B. (1972) *Biochemistry* **11**, 3610–3618.
10. Smiley, C. J. (1985) *Late Cenozoic History of the Pacific Northwest* (Am. Assoc. Adv. Sci., San Francisco).
11. Doyle, J. J. & Doyle, J. L. (1987) *Phytochem. Bull.* **19**, 11–15.
12. Palmer, J. D. (1986) *Methods Enzymol.* **118**, 167–186.
13. McIntosh, L., Paulsen, C. & Bogorad, L. (1980) *Nature (London)* **288**, 556–560.
14. Kreitman, M. & Landweber, L. F. (1989) *Gene Anal. Tech.* **6**, 84–88.
15. Zurawski, G., Clegg, M. T. & Brown, A. H. D. (1984) *Genetics* **106**, 735–749.
16. Hipkins, V. D., Tsai, C.-H. & Strauss, S. H. (1990) *Plant Mol. Biol.* **15**, 505–507.
17. Ohyama, K., Fukuzawa, H., Kohchi, R., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. & Ozeki, H. (1986) *Nature (London)* **322**, 572–574.
18. Kimura, M. (1980) *J. Mol. Evol.* **16**, 111–120.
19. Fitch, W. M. & Margoliash, E. (1967) *Science* **155**, 279–284.
20. Felsenstein, J. (1981) *J. Mol. Evol.* **17**, 368–376.
21. Golenberg, E. M. (1991) *Philos. Trans. R. Soc. Ser. B* **333**, 419–427.
22. Thomas, B. A. & Spicer, R. A. (1987) *The Evolution and Paleobiology of Land Plants* (Croom Helm, London), p. 189.
23. Ritland, K. & Clegg, M. T. (1987) *Am. Nat.* **130**, S74–S100.
24. Smiley, C. J. & Rember, W. C. (1979) *Guidebook and Road Log to the St. Maries River (Clarkia) Fossil Area of Northern Idaho* (Idaho Bureau of Mines and Geol. and Idaho Dept. of Lands, Moscow, ID).