0014-4754/93/100906-04$1.50 + 0.20

PCR jumping in clones of 30-million-year-old DNA fragments from amber preserved termites (*Mastotermes electrodominicus*)

R. DeSalle[a], M. Barcia[a] and C. Wray[a,b]

[a]*Department of Entomology, American Museum of Natural History, New York (New York 10024, USA) and* [b]*Department of Geology and Geophysics, Yale University, New Haven (Connecticut 06511, USA)*
Received 13 April 1993, accepted 13 July 1993

Abstract. DNA from 30-million-year-old amber preserved termites (*Mastotermes electrodominicus*) was PCR amplified with nuclear ribosomal RNA small subunit primers and cloned into the TA vector (INVITROGEN). We obtained several classes of recombinant clones as a result. Authentic *Mastotermes electrodominicus* clones were identified. The source of other classes of clones was identified as contaminants of the ancient DNA template. Several of the clones appeared to be chimeric in structure with half of the clone identical to the termite sequence and the other half identical to contaminant sequences. The phenomenon of PCR jumping was identified as a possible source for the chimeric clones.
Key words. Polymerase chain reaction; ancient DNA; termite relationships; *Mastotermes electrodominicus.*

DNA has been isolated from several ancient sources including human remains[1-3], dried museum skins[4-7], fossilized plant material[8], vertebrate bone[9-12], Rancho La Brea fossils[13] and amber preserved insect material[14-17]. Each of these studies demonstrates the feasibility and utility of using ancient tissues as a source of DNA. In most cases the DNA isolated from these ancient tissues is PCR amplified and directly sequenced. This procedure requires highly purified ancient DNA template. Occasionally, DNA prepared from an ancient organism's tissue is contaminated with DNA from other organisms and in these cases direct sequencing gives ambiguous results.

The isolation and sequencing of parts of a nuclear gene and a mitochondrial gene from an amber preserved termite has recently been reported[17]. In this previous study the products obtained from PCR amplification of the ancient DNA were directly sequenced. Only one of three extinct *Mastotermes electrodominicus* was disturbed in the piece of Miocene-Oligocene amber during our inital attempt to obtain DNA sequences. We have isolated DNA from the remaining two specimens in this single piece of amber in order to further verify our previous results. Due to contamination from other insect specimens in the lab (such as *Drosophila*), we were unable to directly sequence the PCR products we obtained after amplification of the template DNA from these two extinct specimens. We report the rescue of ancient DNA sequences of amber preserved termites from contaminated template using a cloning and sequencing technique in the present study.

Two *Mastotermes electrodominicus* specimens preserved in Miocene-Oligocene amber were used as starting material for DNA isolation. The age and provenance of the amber has been described in detail[17]. The amber was sliced with a sterile razor blade to expose the surface of the insect. Once the insect material was reached, the preserved material was physically removed by taking a sterile pipette tip and scooping the brown, leathery material from the amber and placing it in a sterile eppendorf tube. DNA was extracted by resuspending the insect material in HOM buffer (80 mM EDTA, 100 mM Tris-HCl [pH = 7.5], 160 mM sucrose) and by addition of proteinase K (1 mg/ml final concentration) with incubation for 2 h at 60 °C. This solution was extracted with phenol and chloroform and ethanol precipitated twice. Positive displacement pipette tips were used at all times.

Template DNA was polymerase chain reaction (PCR) amplified using standard protocols with several control reactions to ensure that obvious contamination would be detected. The primers used in the present study were 18S primers (18sai and 18S b5.0; for exact sequences of these primers see reference 17). This primer pair generates a 230-bp-long fragment. Because the DNA isolated from the amber insects was highly degraded (less than 400 bp), we used this primer pair. Primer pairs expected to produce products longer than 400 bases will not give PCR products of this length using highly degraded DNA as a template. In fact, we routinely use a primer pair (18Sai and 18Sbi) that gives a 1.1 kb fragment as a negative control[17]. The length of the ancient DNA isolated in our study from amber material should be contrasted with the much longer fragments isolated from Miocene-Oligocene amber preserved bees[14].

PCR products from the 18Sai and 18Sb5.0 reactions were cloned in the TA cloning vector (INVITROGEN). Detection of clones with inserts of the proper size was done by restricting DNA from positive white colonies with *Eco*RI and *Hind*III. 215 colonies were screened

using this method. The PCR products were cleaned for sequencing with GENECLEAN (Bio 101) and double-strand sequencing was done using the SEQUENASE system (US Biochemicals). Clones obtained from the TA cloning experiments were sequenced by purifying plasmid DNA with a standard phenol prep[18] followed by double-strand sequencing using the SEQUENASE system (US Biochemicals) and universal forward and reverse primers.

Although we exercised the identical precautions against contamination in the present study as were exercised in our previous study of amber termite DNA[17], contamination in the PCR products generated for this study was obvious. Some regions of the sequence were readable while others showed multiple bands across positions. This pattern can best be explained by the following. Areas where sequence could be read are areas of high conservation across extreme phylogenetic distances. These areas are identical in both the termite sequence and any contaminant that might be present in the direct PCR product. Areas of ambiguity represent stretches of sequence that are variable in the termite sequence and the contaminant sequences. This interpretation is confirmed when we examined the aligned 18S sequences of several phylogenetically different organisms. Areas of ambiguity in our direct sequencing runs are indeed areas of variability in the phylogenetically diverse comparisons.

We eventually sequenced 49 clones that had the correct size insert of which 20 proved to contain 18S rDNA sequences. We subdivided these twenty clones into five arbitrary but distinct classes (A through E) as described below and in table 1. The origins of some of the classes of clones were identified by computing a pairwise distance matrix for several eukaryotes in the current 18S data base (table 2) using PAUP[19]. The analysis revealed that there were only two classes of the clones that were identical to arthropods that have already been sequenced. Eight of the clones (class A) were identical to *Drosophila* 18S sequences. An additional two clones (class B) were identical to the sequences reported for the amber termite *Mastotermes electrodominicus*[17]. Since these two clones were identical in sequence to the previously reported amber sequence and were generated from a new source of amber preserved termites, we take this as direct evidence in confirmation of the validity of the ancient DNA we have reported earlier[17].

Some of the clones (class C) were very similar to sequences reported in the 18S data set for other arthropods. For instance, clone C-23 is only one substitution different from the sequence for *Gryllus* (the Northeast beach cricket). The template that generated the sequences for clone C-23 could have originated from *Gryllus* or a close relative. Another class of clones (class E) that could not be identified from the current data

Table 1. Sequences of clones generated for this study. Only one sequence is reported for the class A clones (A-2) as all eight were identical in sequence. Only one sequence is reported for the class B clones (B-50) as both clones were identical in sequence. Dr = *Drosophila*; Tr = *Mastotermes electrodominicus* (amber preserved termite sequence[17]).

```
Dr      TTACCCACTCCCAGCTCGGGGAGGTAGTGACGAAAAATAA
A-2     TTACCCACTCCCAGCTCGGGGAGGTAGTGACGAAAAATAA
Tr      TTACCCACTCCCGGCACGGGGAGGTAGTGACGAAAAATAA
B-50    TTACCCACTCCCGGCACGGGGAGGTAGTGACGAAAAATAA
C-23    TTACCCACTCCCGGCACGGGGAGGTAGTGACGAAAAATAA
C-69    TTACCCACTCCCGGCACGGGGAGGTAGTGACGAAAAATAA
D-165   TTACCCACTCCCAGCTCGGGGAGGTAGTGACGAAAAATAA
D-16    TTACCCACTCCCGGCTCGGGGAGGTAGTGACGAAAAATAA
D-131   TTACCCACTCCCGGCACGGGGAGGTAGTGACGAAAAATAA
E-172   TTACCCAATCCCAACACGGGGAGGTAGTGACAATATGTAA
E-178   TTACCCAATCCCGGCACGGGGAGGTAGTGACAATAAATAC
E-156   TTACCCAATCCCGACACGGGGAGGTAGTGACAATAAATAC
Dr      CAATACAGGACTCATATCCGAGGCCCTGTAATTGGAATGA
A-2     CAATACAGGACTCATATCCGAGGCCGTGTAATTGGAATGA
Tr      CGATACGGGACTC--TTCCGAGGCCCCGTAATCGGAATGA
B-50    CGATACGGGACTC--TTCCGAGGCCCCGTAATCGGAATGA
C-23    CAATACGGGACTC--TTTTGAGACCCCGTAATTGGAATGA
C-69    CAATACGGGACTC--TTTTGAGACCCCGTAATTGGAATGA
D-165   CAATACGGGACTC--TATCGAGGCCCCGTAATTGGAATGA
D-16    CAATACGGGACTC--TATCGAGGCCCCGTAATTGGAATGA
D-131   CGATACGGGACTC--GTCCGAGGCCCCGTAATCGGAATGA
E-172   CAATACGGGCCTTATATCCGAGGCCCCGTAATTGGAATGA
E-178   CAATACAGGCCTC--TTTTGGGGTCTGGTAATTGGAATGA
E-156   CAATACGGGCCTC--TTTTGGGTCCTTGTAATTGGAATGA
Dr      GTACACTTTAAATCCTTTAACAAGGACCAATTGGAGGG
A-2     GTACACTTTAAATCCTTTAACAAGGACCAATTGGAGGG
Tr      GTACACTCTAAATCCTTTAACGAGGATCCATTGGAGGG
B-50    GTACACTCTAAATCCTTTAACGAGGATCTATTGGAGGG
C-23    GTACACTCTAAATCCTTTAACGAGGATCTATTGGAGGG
C-69    GTACACTCTAAATCCTTTAACGAGGATCTATTGGAGGG
D-165   GTACACTCTAAATCCTTTAACGAGGATCCATTGGAGGG
D-16    GTACACTCTAAATCCTTTAACGAGGATCTATTGGAGGG
D-131   GTACACTTTAAATCCTTTAACGAGGACCAATTGGAGGG
E-172   GTACATTTTAAATCCTTTAACGAGGACCAATTGGAGGG
E-178   GTACAATTTAAATCCCTTAACGAGGAACAATTGGAGGG
E-156   GTACAATTTAAATCCGTTAACGAGGAACAATTGGAGGG
```

Table 2. Examples of nucleotide differences between cloned sequences and sequences from selected taxa in the 18S data base. Numbers across the top of the table represent the following taxa: 1 = *Mastotermes electrodominicus*; 2 = *Dysodius*; 3 = *Belastoma*; 4 = *Drosophila*; 5 = *Papilio*; 6 = *Tenebrio*; 7 = *Warramaba*; 8 = *Gryllus*; 9 = *Scutigera*; 10 = *Artemia*; 11 = *Nephila*; 12 = *Limulus*; 13 = *Homo*; 14 = *Strongylocentrotus*.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A2	24	21	43	0	13	21	24	19	34	22	27	23	33	32
B50	0	7	37	25	14	4	8	8	20	12	15	10	24	23
C23	6	12	37	21	14	9	13	1	22	15	14	12	24	23
C69	6	12	37	21	14	9	13	1	22	15	14	12	24	23
D165	6	12	38	22	19	10	15	9	24	14	14	10	23	23
D16	7	13	38	22	19	10	14	8	23	15	15	9	24	24
D131	4	7	35	22	12	5	9	12	22	12	18	14	25	23
E172	30	31	42	15	23	30	34	29	43	32	35	33	43	41
E178	19	21	33	28	22	21	25	19	31	25	26	26	31	29
E156	19	21	33	30	22	21	25	18	31	24	25	26	30	28

base of eukaryote sequences was also identified. These clones showed several positions that were radically different from the eukaryotes in the 18S data base. Inspection of table 2 reveals that these clones were from 15 to 43 substitutions different from the sequences in this data base. A further attempt to define the origin of these clones was made by conducting a phylogenetic analysis of these clones with a full eukaryote 18S rDNA sequence data base. The sequences for clones E-178 and E-156 show a high degree of similarity to higher fungi sequences. It is possible that these two sequences represent 18S fungal sequences also derived from the amber; however, without more sequence data it remains difficult to positively identify these two sequences. Phylogenetic analysis placed clone E-172 in a basal metazoan position; this result indicates that this clone is a contaminant or a result of PCR jumping.

The rest of the clones (class D) differed in a readily interpretable manner. Table 2 demonstrates that the difference in the remainder of the clones from *Drosophila* and *Mastotermes* sequences ranged from four substitutions to 21 substitutions. The number of differences between *Drosophila* and *Mastotermes* sequences is 25 substitutions. The sequences that we report in table 1 for clones D-165 and D-16 can best be explained as chimeras of *Drosophila* and amber termite sequences (fig. 1). The sequence that we report for clone D-131 can best be explained as a chimera of the *Gryllus*-like sequence and the amber termite sequence. The amber termite sequence can be reconstructed from these chimeric clones using the limits of similarity to both *Drosophila* and *Mastotermes electrodominicus* 18S sequences from figure 1. It is also possible that the class E clones discussed above are chimeric. If these class E clones are chimeric, the sequences are too scrambled for positive association or identification.

Higuchi and Wilson[20] first suggested the possibility that highly degraded template DNA in PCR reactions might generate PCR products where the degraded template DNA would serve as primer for part of the reaction in the early phases of amplification. These small stretches of template would then allow the PCR amplification to proceed by 'jumping' from template to template, producing chimeric PCR products (fig. 2). Paabo et al.[21] showed that this phenomenon would readily occur using an experimental system, and Bradley et al.[22] have demonstrated the shuffling of cloned Adh-1 sequences in pocket gopher DNA amplifications. Lawlor et al.[3] recognized chimeric PCR products from ancient human HLA genes. It is not at all surprising given the degraded nature and age of our termite DNA preparations that this phenomenon occurs frequently in our amplifications. Unlike the phenomena reported by Lawlor et al.[3], it is apparent from the number of clones in this study that PCR 'jumping' (fig. 2) can be common and is efficient at producing full length chimeric PCR products from extremely ancient DNA templates.

Is is possible to recover full-length information from highly degraded template DNA by piecing together sequences from chimeric clones that have incorporated a known DNA sequence (such as *Drosophila*) and a target sequence. Figure 1 shows that clone D-165 can best be explained by a jump between the *Drosophila* 18S sequences on the left to termite sequences on the right at position 46. Positions 46 thru 117 are therefore termite-

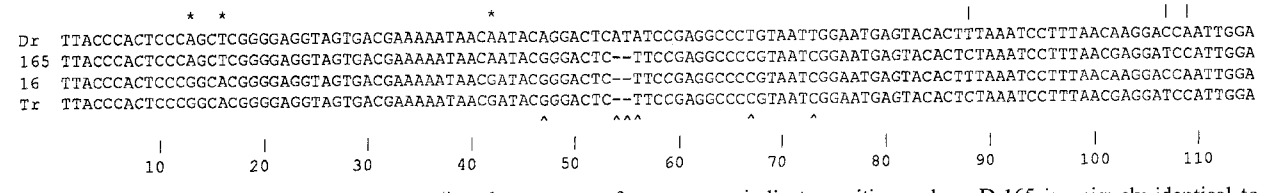

Figure 1. Analysis of chimeric clones. The aligned sequences of the 18S region in *Drosophila melanogaster* (Dr), *Mastotermes electrodominicus* (Tr), clone B-16 (16) and B-165 (165) are shown. Asterisks above the sequence indicate positions where clone D-16 is uniquely identical to the termite sequence. Lines above the sequence indicate positions where D-165 is uniquely identical to the termite sequence. Arrows below the sequences indicate positions where both clones are identical to the termite sequence. Numbers below the sequence are for reference to the text.

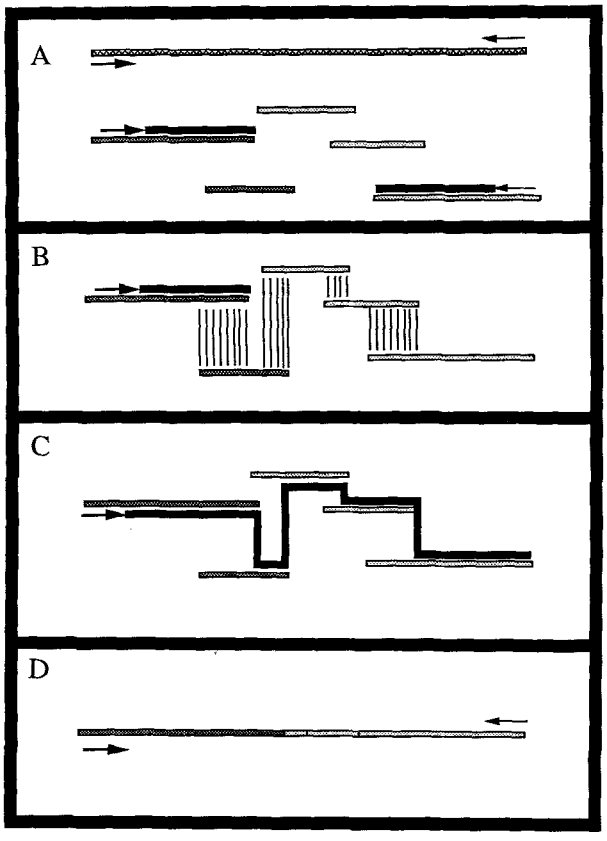

Figure 2. Schematic representation of PCR jumping. *A* Full-length template and several shorter degraded template molecules. Dark shading and light shading represent two taxa respectively. Small arrows indicate primers. Heavy black lines represent first cycle PCR extension products for degraded molecules. *B* In subsequent cycles the short PCR extended fragments act as primers to secondary, tertiary etc. degraded templates. The vertical lines indicate regions on the degraded templates where priming would occur. *C* Tracing of hypothetical path of PCR jumping for the fragments pictured in *B*. *D* Final chimeric PCR product equivalent in length to the original full-length target diagramed in *A*.

like. Position 46 is the first position in this chimeric clone sequence that is unambiguously termite-like. Clone D-16, on the other hand, can best be explained by a jump from termite sequences on the left to *Drosophila* sequences on the right at position 87. This interpretation means that positions 1 thru 87 are termite-like. This interpretation is supported, in part, by the exact match of sequences for clone D-16 and D-165 between positions 46 and 87 with the termite sequence. This is a 40-base-long region with eight positions different between termite and *Drosophila*. The exact match of class B clones with the previously reported amber termite sequence and the exact match of the reconstructed amber termite sequence (fig. 1) confirms our previous report of a 30-million-year-old DNA sequence from amber termites.

Acknowledgements. We thank W. Wheeler, A. Williams and J. Gatesy for advice and comments on the manuscript. D. Grimaldi provided the amber preserved termites and commented on an earlier draft of the manuscript.

1 Paabo, S., Nature *314* (1985) 644.
2 Doran, H. G., Dickel, D. N., Ballinger, W. E., Agee, P. J., and Laipis, W. W., Nature *323* (1986) 803.
3 Lawlor, D., Dickel, C., Hauswirth, W., and Parham, P., Nature *349* (1991) 785
4 Higuchi, R., Dowman, B., Freiberger, M., Ryder, O., and Wilson, A. C., Nature *312* (1984) 282.
5 Thomas, R. H., Schaffner, W., Wilson, A. C., and Paabo, S., Nature *340* (1989) 465.
6 Thomas, W. K., Paabo, S., Villablanca, F. X., and Wilson, A. C., J. molec. Evol. *31* (1990) 101.
7 Patton, J. L., and Smith, M. F., Evolution *46* (1992) 148.
8 Golenberg, E. M., Giannasi, D. E., Clegg, M. T., Smiley, C. J., Durbin, M., Henderson, D., and Zurawski, G., Nature *344* (1990) 656.
9 Hagelberg, E., Sykes, B., and Hedges, R., Nature *342* (1989) 485.
10 Yokoi, T., Aoki, Y., and Sagisaki, K., Zs. Rechtsmedizin *102* (1989) 323.
11 Williams, S. R., Longmire, J. L., and Beck, L. A., Am. J. Anthropol. *812* (1990) 318.
12 Cooper, A., Mourer-Chauvire, C., Chambers, G. K., von Haeseler, A., Wilson, A. C., and Paabo, S., Proc. natl Acad. Sci. USA *89* (1992) 8741.
13 Janeczewski, D. N., Yuhki, N., Gilbert, D. A., Jefferson, G. T., and O'Brien, S. J., Proc. natl Acad. Sci. USA *89* (1992) 9769.
14 Cano, R. J., Poinar, H., and Poinar, G. O., Med. Sci. Res. *20* (1992) 249.
15 Cano, R. J., Poinar, H., and Poinar, G. O., Med. Sci. Res. *20* (1992) 619.
16 Cano, R. J., Poinar, H., Pieniazek, N., Aftrim, A., and Poinar, G. Jr, Nature *363* (1993) 536.
17 DeSalle, R., Gatesy, J., Wheeler, W., and Grimaldi, D., Science *257* (1992) 1933.
18 Maniatis, T., Fritsch, E. F., and Sambrook, J., Molecular Cloning: A Laboratory Manual. Cold Spring Harbour Press, Cold Spring Harbour, New York 1982.
19 Swofford, D. L., Phylogenetic Analysis Using Parsimony (PAUP). Illinois Natural History Survey, 1990.
20 Higuchi, R., and Wilson, A. C., Fedn Proc. *43* (1984) 1557.
21 Paabo, S., Irwin, D. M., and Wilson, A. C., J. biol. Chem. *265* (1990) 4718.
22 Bradley, R. D., Bull, J. J., and Hillis, D. M., personal communication, 1992.