Letter to the Editor

Recombinant DNA Sequences Generated by PCR Amplification

Robert D. Bradley and David M. Hillis†*

*Department of Biological Sciences, Texas Tech University; and †Department of Zoology, University of Texas at Austin

The polymerase chain reaction (PCR) has greatly enhanced the field of molecular biology by making numerous regions of the genome (coding and noncoding), in both extant and extinct taxa, accessible for detailed analysis. PCR is especially well suited for applications in systematic biology because conserved regions that flank variable portions of the genome can be used as primer sites for the amplification and sequencing of variable regions from a wide variety of species (see Palumbi 1996). In addition, PCR has been used successfully to examine polymorphisms within populations.

Although many advantages are offered by PCR, the technique is not error-free. In fact, the replication-like process that drives PCR can lead to erroneously amplified products. For example, recombinant amplification products (Saiki et al. 1988; Scharf 1990) can be produced in vitro when amplifying elements of multi-gene families. Thus, the amplified product may not represent a sequence that actually exists in a single continuous stretch of DNA within any organism. Although this phenomenon was reported shortly after the development of PCR, it appears to be poorly known by many evolutionary biologists who use PCR in studies of gene evolution. Here we show that in vitro recombination also can occur when amplifying alleles of single-gene loci in heterozygous individuals, and that recombination frequency can vary as a function of using different polymerases in the extension step of PCR.

In vitro recombination occurs when DNA polymerases incompletely extend a segment of DNA and the partially amplified sequence acts as a primer during subsequent amplification cycles (fig. 1). If the partially amplified segment hybridizes to an alternative form of a template strand, then subsequent PCR products will be recombinants of the two original template sequences. Partial amplification is most likely to occur when polymerases pause or prematurely disassociate from the template strand during the extension reaction, although the recombination may also be enhanced by short extension cycles or long amplifications.

Although this phenomenon was initially identified in amplification of multi-gene families (Scharf, Long, and Erlich 1988; Scharf et al. 1988), we identified products of in vitro recombination in pocket gophers (*Geomys*) that were heterozygous for the alcohol dehydrogenase−1 (*Adh*−1) locus (Bradley et al. 1993). This experiment was designed to determine the DNA sequences of three *Adh*−1 alleles (designated K, M, and N) that had been identified from allozyme studies. Initially, we determined the sequences of the K and M alleles from individuals homozygous for the two respective allozymes, and found that eight base substitutions distinguished these two alleles. To determine the sequence of the N allele, it was necessary to amplify the *Adh*−1 region from a heterozygous individual (KN), because the N allele is rare and has not been found in the homozygous state (Baker et al. 1989; Bradley et al. 1993). We cloned and sequenced seven isolates from this PCR reaction, and found that two (29%) of the clones were K-like, two (29%) were interpreted as representing the N allele, and three (43%) were recombinants between K and N. The presence of the parental sequences was verified in the original genomic DNA using restriction enzymes and by sequencing additional clones and individuals, and recombinant alleles were identified using five DNA base substitutions that differed between the parental sequences as markers (fig. 2). Of the three recombinant sequences isolated, one could have been produced from a single recombination event, a second required two recombinations, and the third required three recombinations.

We could not determine the reason for premature polymerase termination. No secondary structures or GC-rich regions appear to be correlated with areas in which the disassociation events must have occurred. The initial amplifications were conducted with *Taq* polymerase (from *Thermus aquaticus*) using the following conditions: one cycle—denaturation at 94°C for 3 min, annealing at 50°C for 1 min, extension at 73°C for 1.5 min; and 29 cycles—denaturation at 94°C for 1 min, annealing at 50°C for 1 min, extension at 73°C for 1.5 min. We then repeated this amplification using a thermostable polymerase from *Thermococcus litoralis* (Vent®) and the following conditions: one cycle—denaturation at 96°C for 2 min, annealing at 50°C for 1 min, extension at 73°C for 1.5 min; and 29 cycles—denaturation at 96°C for 1 min, annealing at 50°C for 1 min, extension at 73°C for 1.5 min. Recombinant sequences were still obtained, but the frequency of recombinants was lowered to 1 in 12 (8%). It appears that this polymerase does not prematurely disassociate from the template strand at the same frequency as the *Taq* polymerase, but we do not know if these results are specific to the sequences we amplified. The amplification conditions for the two polymerases differed only in denaturation temperature, 96°C for the Vent® and 94°C for the *Taq* reactions, a variable that was unlikely to have affected disassociation of the polymerase in the extension reaction. However, Vent polymerase has a higher sequence fidelity and 3′–5′ exonuclease activity that *Taq* does not possess. It may be that the increased fidelity and exonuclease (proofreading) activity of Vent® allows the polymerase to remain associated with the template DNA in troublesome regions, such as hairpins or pseudostops.

DNA sequences amplified by PCR are assumed by most evolutionary biologists to represent contiguous sequences from a single chromosome location. Our results

Key words: polymerase chain reaction, recombinant DNA sequences, DNA polymerase.

Address for correspondence and reprints: Robert D. Bradley, Department of Biological Sciences, Texas Tech University, Lubbock, Texas 79409;. E-mail: izrdb@ttacs3.ttu.edu.

Mol. Biol. Evol. 14(5):592–593. 1997

FIG. 1.—The formation of a recombinant PCR product is depicted in steps *a–f*. *a,* A standard PCR reaction with Primer 1 annealing to template DNA, depicted by hypothetical Allele A, which possesses **T** and **A** at the indicated polymorphic sites. *b,* Premature termination occurs during the extension of Allele A, resulting in an incomplete or partial PCR product. *c,* In a subsequent PCR cycle, the partial PCR product from step *b* acts as a primer and anneals to Allele B, which has different nucleotides (**C** and **G**) at the two polymorphic sites. *d,* The annealed partial product from Allele A is extended to completion and incorporates the 5′ polymorphism from Allele B. *e,* Primer 2 anneals to the recombinant or shuffled PCR product generated in step *d*. *f,* Complete extension of the shuffled PCR product from step *e* results in polymorphic sites from Alleles A and B present together in a single amplification product, which can then be amplified in subsequent rounds of PCR without additional premature termination.

indicate that this may not be a safe assumption. Amplified sequences should not be assumed to be allelic sequences without verification such as restriction digests and/or the examination of multiple clones and individuals. In addition to recombination among alternative alleles in heterozygotes, the recombination phenomenon also may extend to recombination among repetitive sequences or duplicated loci, or between functional genes and pseudogenes. Thus, great care should be exercised in interpreting the evolutionary history of single-amplification products; such studies have the potential to confuse the history of alleles, multiple loci, and pseudogenes.

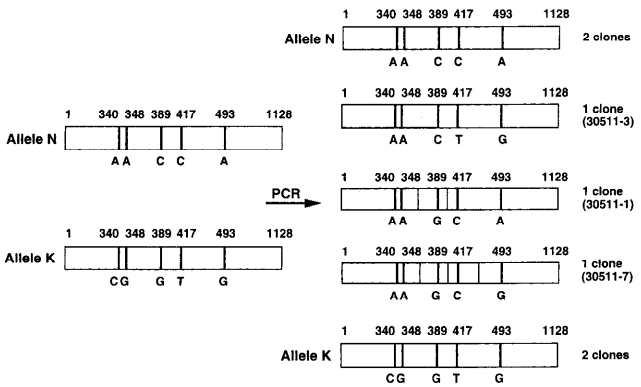

FIG. 2.—Recombinants produced from amplification of the *Adh*−1 locus from a heterozygous individual of *Geomys bursarius*. The heterozygous individual contained copies of alleles N and K (left); the amplification reaction contained the two original sequences as well as three different recombination products which were produced by a minimum of one (clone 30511-3), two (clone 30511-1), or three (clone 30511-7) recombination events.

LITERATURE CITED

BAKER, R. J., S. K. DAVIS, R. D. BRADLEY, M. J. HAMILTON, and R. A. VAN DEN BUSSCHE. 1989. Ribosomal-DNA, mitochondrial-DNA, chromosomal, and allozymic studies on a contact zone in the pocket gopher, *Geomys*. Evolution **43**:63–75.

BRADLEY, R. D., J. J. BULL, A. D. JOHNSON, and D. M. HILLIS. 1993. Origin of a novel allele in a mammalian hybrid zone. Proc. Natl. Acad. Sci. USA **90**:8939–8941.

PALUMBI, S. R. 1996. Nucleic acids II: the polymerase chain reaction. Pp. 205–247 *in* D. M. HILLIS, B. K. MABLE, and C. MORITZ, eds. Molecular systematics. 2nd edition. Sinauer, Sunderland, Mass.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. HIGUCHI, G. T. HORN, K. B. MULLIS, and H. A. ERLICH. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–491.

SCHARF, S. J. 1990. Cloning with PCR. Pp. 84–91 *in* M. A. INNIS, D. H. GELFAND, J. J. SNINSKY, and T. J. WHITE, eds. PCR protocols: a guide to methods and application. Academic Press, San Diego, Calif.

SCHARF, S. J., A. FRIEDMANN, C. BRAUTBAR, F. SZAFER, L. STEINMAN, G. HORN, U. GYLLENSTEN, and H. A. ERLICH. 1988. HLA class II allelic variation and susceptibility to *Pemphigus vulgaris*. Proc. Natl. Acad. Sci. USA **85**:3504–3508.

SCHARF, S. J., C. M. LONG, and H. A. ERLICH. 1988. Sequence analysis of the HLA-DRβ and HLA-DQβ loci from three *Pemphigus vulgaris* patients. Hum. Immunol. **22**:61–69.

RODNEY L. HONEYCUTT, reviewing editor

Accepted February 5, 1997