Rapid Surveying of DNA Sequence Variation in Natural Populations[1]

Enrique P. Lessa

Museum of Vertebrate Zoology, University of California, Berkeley

DNA sequencing can be costly and time consuming for population studies because of the relative rarity of variation along exons. These problems can be substantially reduced by the use of the polymerase chain reaction on introns using primers from the exon region. These problems can be further reduced by the use of denaturing gradient gel electrophoresis to identify those alleles in need of sequencing.

Introduction

The study of variation of DNA sequences promises to provide important insights into the genetic structure and dynamics of natural populations. However, sequencing is too costly and time consuming for the large samples and multiple loci often required for population-level studies. In addition, heterozygotes are difficult to score in sequencing gels; this complicates the assessment of genotypic variation in nuclear loci. Here I present an approach for overcoming these two problems, capitalizing on the known intron/exon structure of many eukaryotic genes and on available laboratory techniques.

General Approach

The coding segments (exons) of most eukaryotic genes are interrupted by noncoding introns. The latter are likely to vary within species and can therefore be targeted as markers of population variation and subdivision, by using conserved regions of flanking exons to design primers for amplifications via the polymerase chain reaction (PCR; Saiki et al. 1988). The analysis of these double-stranded PCR products usually proceeds by asymmetric amplifications (or strand separation) followed by direct sequencing of both strands. However, the gathering of allelic data may be simplified by scoring individual genotypes by using denaturing gradient gel electrophoresis (DGGE) and limiting the sequencing efforts to the distinct alleles identified by this process [fig. 1(a)].

Detailed descriptions of laboratory equipment and procedures for DGGE are readily available (Myers et al. 1987, 1989*a*, 1989*b*; Sheffield et al. 1989; Abrams et al. 1990). In brief, DGGE allows double-stranded PCR products that differ by one or more mutations to be separated on acrylamide gels cast with linear gradients of denaturants (urea and formamide). This is possible because mutations alter the DNA segments' resistance to denaturation, resulting in changes in mobility on gradient gels. As shown diagramatically in figure 1, there are two types of denaturing gradient gels:

1. Key words: polymerase chain reaction, denaturing gradient gel electrophoresis, intron, β-globin.

Address for correspondence and reprints: Enrique P. Lessa, Museum of Vertebrate Zoology, University of California, Berkeley, California 94720.

Mol. Biol. Evol. 9(2):323–330. 1992.

0737-4038/92/0902-0011$02.00

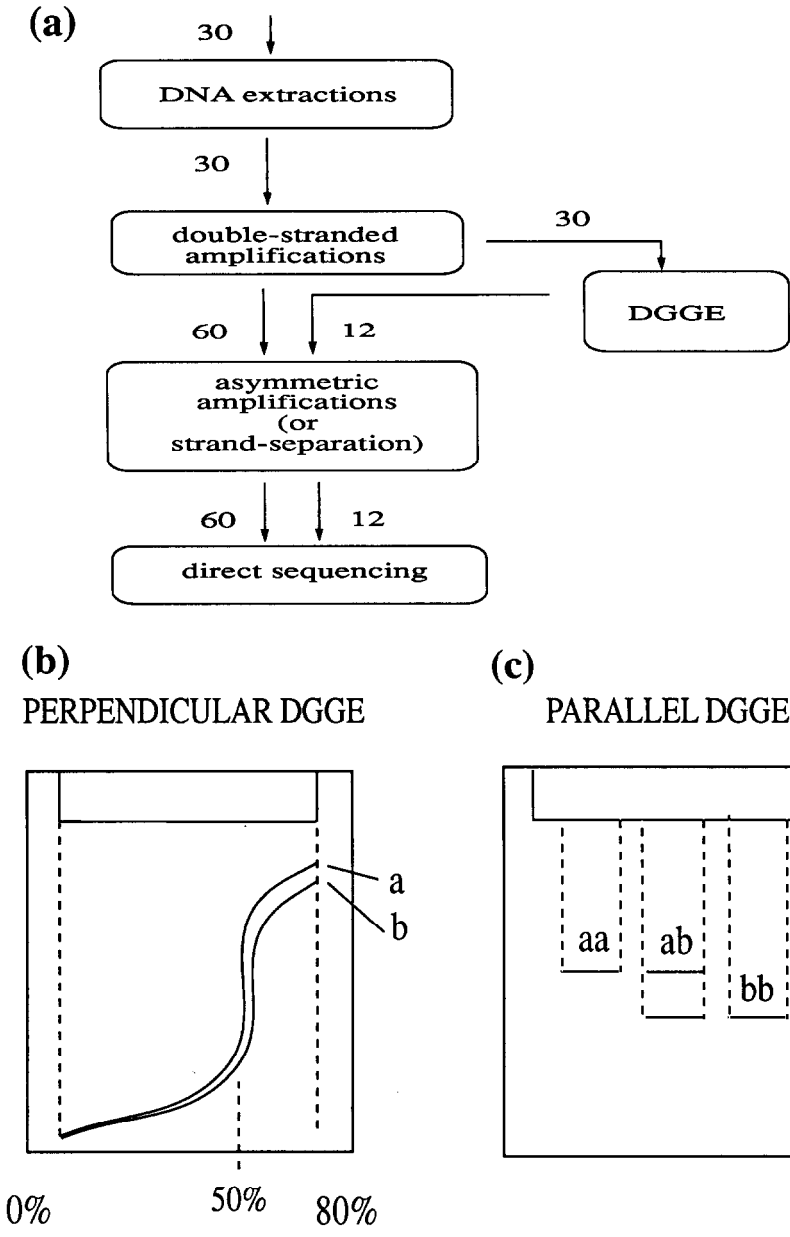

FIG. 1.—(a), Flow chart depicting use of DGGE in combination with PCR and direct sequencing. DGGE represents a detour from the usual procedures that allows one to score individual genotypes and to limit sequencing efforts. In the example presented in the present paper, 30 specimens would have required 60 asymmetric amplification/sequencing reactions. DGGE showed that only three alleles were present in the samples; sequencing replicates of each required a total of 12 asymmetric amplification sequencing reactions. (b), Diagrammatic representation of perpendicular DGGE of single heterozygous specimen. The point of partial denaturation of the two alleles is shown at ~50% denaturant. (c), Same two alleles separated and scored on parallel gels with ranges of denaturant centered at ~50%.

1. Perpendicular gradient gels are used to evaluate the response of a single PCR product to the gradient. The gel has a full range (0%–80%) of denaturants from side to side, i.e., perpendicular to the direction of electrophoresis. Perpendicular DGGE is used to determine the point of partial denaturation of the DNA segment under study [fig. 1(b)].

2. Parallel gradient gels are cast with a gradient of denaturants from top to bottom, i.e., parallel to the direction of electrophoresis. The denaturants on parallel gels cover a range of concentrations of 20%–30% points centered about the point of denaturation previously determined by perpendicular DGGE. On this narrower range and under constant temperature, parallel DGGE allows the separation of PCR products differing by one or more mutations, because such mutations change the DNA's resistance to denaturation. In a heterozygous sample, for instance, each of the alleles will partially denature (and consequently slow down its rate of migration) at a different point along the gradient. Multiple samples can be loaded and run simultaneously on parallel denaturing gradient gels and can be scored after ethidium bromide staining [fig. 1(c)].

In this context, DGGE may be conceived of as a detour from the usual PCR/sequencing procedures [fig. 1(a)], one that both simplifies the scoring of genotypes and limits costly and labor-intensive asymmetric amplifications (or strand separation) and sequencing to the minimum. Also, alleles are effectively gel purified in parallel DGGE. If the sequences of alleles only found in heterozygotes are of interest (e.g., rare alleles in hybrid zones), they can be cut out of the parallel denaturing gradient gels for subsequent reamplification and sequencing.

If no sequences are available for the loci and species to be studied, the approach outlined above for the study of introns can be best implemented in several steps, each using a specific combination of primers (fig. 2). First, a pilot PCR/sequencing project directed at obtaining a few preliminary sequences may be carried out with primers designed on the basis of conserved exon regions identified in alignments of sequences of other taxa. The preliminary sequences should include data from at least one of the exons flanking the targeted intron, in order to verify the identity of the amplified gene and to design one species-specific primer for future experiments. This primer is synthesized in two versions. One version should include a 40-mer "GC-clamp" on the 5′ end, to allow maximal resolution in DGGE (Myers et al. 1989*a*). It is used both for double-stranded PCR amplifications of all specimens and for the screening of amplified products for allelic variation in denaturing gradient gels. Another version of the primer,

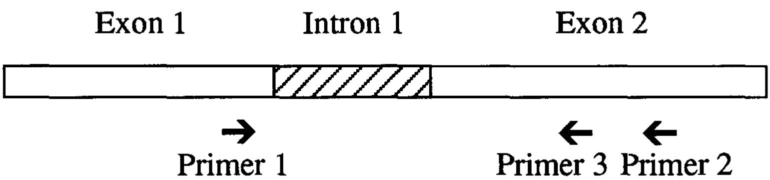

FIG. 2.—Diagrammatic representation of intron and flanking exons and of primers used for uncovering variation in intron by combination of amplifications by PCR, DGGE, and direct sequencing (see text for details). Primers 1 and 2 are based on conserved regions identified in alignments of available sequences of the target gene in other species and are used to obtain initial sequences. Primer 3 is species specific and is designed in two versions. One has a GC-clamp on the 5′ end and is used in conjunction with primer 1 for double-stranded PCR amplifications of products to be screened by DGGE. A version of primer 3 without the GC-clamp is used with primer 1 to reamplify and directly sequence the alleles identified by DGGE.

one that lacks the GC-clamp, is used both for asymmetric amplifications and for direct sequencing of the amplified products.

An Example: β-Globin Intron 1 in Pocket Gophers

This approach will be illustrated with an analysis of variation in intron 1 of the adult β-globin gene of pocket gophers (*Thomomys bottae*), surveyed in 30 individuals from eight localities in the lower Colorado River area. DNA was obtained from frozen liver samples by sodium dodecyl sulfate/proteinase K/RNAse lysis followed by phenol or sodium chloride extraction and alcohol precipitation (Maniatis et al. 1982, pp. 458–462; Miller et al. 1988).

The two initial primers were designed on the basis of alignments of known sequences of genes of the β-globin family in a variety of mammals. The 3′ ends of the three exons showed greater conservation than did other regions. Intron 1 of adult β-globin is consistently small (~120 bp long) and was targeted for initial amplifications with two primers—β-1 (5′-GTTGGTGGTGAGGCCCTGGGCAG) and β-2 (5′-CCTGAAGTTCTCAGGATCCACATGCA)—complementary to segments of exons 1 and 2, respectively.

The initial sequences of four pocket gopher samples showed greatest similarity to adult β-globins of various mammals, followed by embryonic β-globins and by pseudogenes in the family. The pocket gopher sequences of exon 2 were used to design primer β-6 (5′-TTCGCTCAGACTGGTAAAGGTGCCCTTCA). The GC-clamp attached to the 5′ end of one version of β-6 (β-6+GC) had the following sequence: 5′-CGCCCGCCGCGCCCCGCGCCCGGCCCGCCGCCCCCGCCCC. Myers et al.'s (1989*a*) suggestion not to include bases other than G and C was followed, causing the GC-clamp to differ from theirs by only one position.

The PCR products amplified with β-1 and β-6+GC were used for DGGE. First, the point of partial denaturation of a single amplified product was determined to be 55% on a perpendicular denaturing gradient gel (Myers et al. 1989*a*). Subsequent parallel denaturing gradient gels, used to screen allelic variation in the sample, were cast with linear gradients of 45%–65% urea. All gels were run for 3–5 h at 150 V while immersed in an aquarium at 60°C and were flushed with water and stained with ethidium bromide.

Three alleles were identified by DGGE (fig. 3). Two of them were restricted to one of the two "genetic groups" (defined by allozyme analyses; Smith and Patton 1980, 1984) represented by these populations, but the third was found in both (fig. 4). The distribution of β-globin alleles is consistent with the pattern uncovered by protein electrophoresis in that (1) variation in allele frequencies may be marked across short geographic distances and (2) whereas the two genetic groups differ by substantial

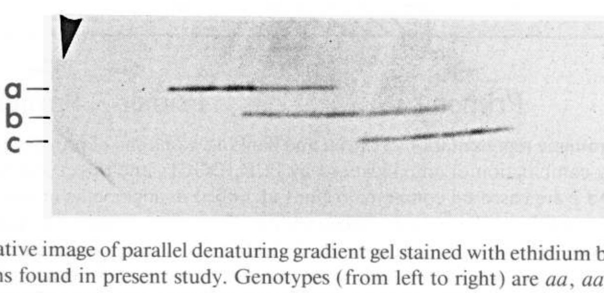

FIG. 3.—Negative image of parallel denaturing gradient gel stained with ethidium bromide and showing allelic combinations found in present study. Genotypes (from left to right) are *aa*, *aa*, *ab*, *ab*, *bb*, *bc*, *bc*, *cc*, and *cc*. The arrow indicates the direction of migration during electrophoresis.

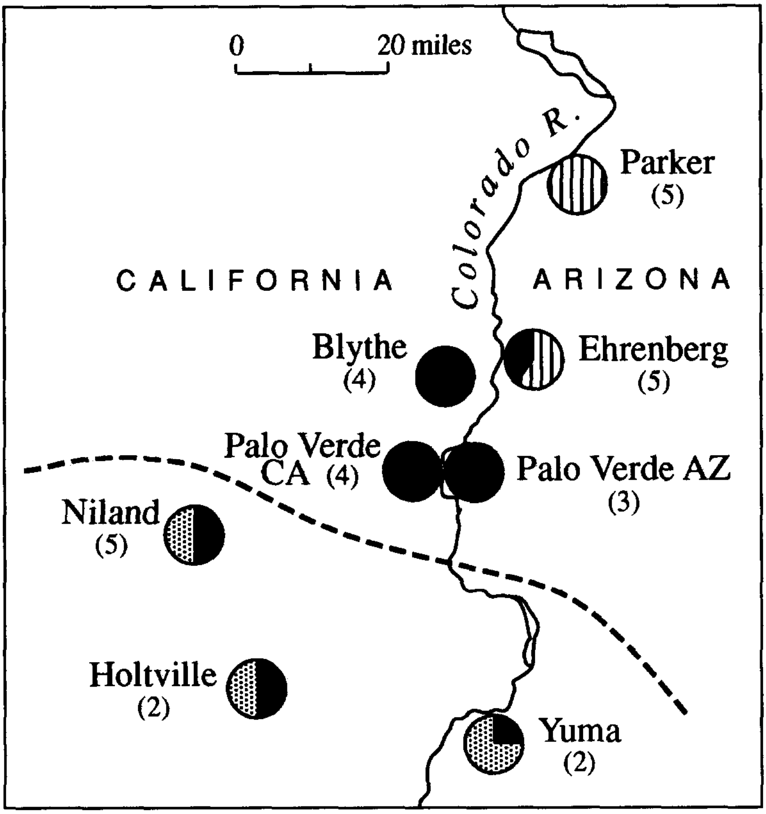

FIG. 4.—Map depicting distribution of β-globin alleles in study area. The dashed line separates the two "genetic groups" identified by protein electrophoresis (Smith and Patton 1980) and by sequences of the control region of mitochondrial DNA (author's unpublished data). ○ = Allele c; ● = allele b; and ◍ = allele a.

genetic distances, many alleles are common to both (Patton and Smith 1990). In contrast, mitochondrial haplotypes have more limited distributions and are clearly distinct in the two genetic groups (author's unpublished data). Although more data are needed to understand the population dynamics of pocket gophers in this area, the information at hand suggests that there is substantial population subdivision and that gene flow is male biased in these pocket gophers.

Asymmetric amplifications (Gyllensten and Erlich 1988) were carried out using alleles sampled from either agarose gels or acrylamide denaturing gradient gels. Direct sequencing of the reamplified alleles showed that intron 1 of the adult β-globin of pocket gophers is 126 bp long and that two point substitutions account for the differences between the three alleles (fig. 5). No additional variation was detected, either by sequencing additional samples or by the introduction of "heteroduplex analysis" to the DGGE protocol. In general, however, heteroduplex analysis is advisable (see Abrams et al. 1990, and references therein).

The results attest to the resolving power of DGGE. Furthermore, because the observed differences between the three alleles are simple, a direct relationship can be drawn between such differences and the behavior of the alleles on the gradient gels. Thus, increasing resistance to denaturation, from allele a to allele b and allele c, is

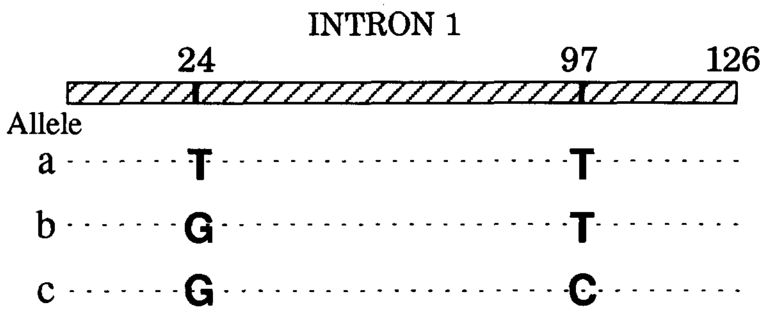

FIG. 5.—Diagrammatic representation of intron 1 of adult β-globin of pocket gophers. The intron is 126 bases long in the three alleles, which, as noted, differ only at positions 24 and 97.

accompanied by replacements of AT by GC base pairs (figs. 3 and 5). Also, the data show that even small introns may be variable at the population level.

Whether the level of variation found in this gene and taxon is representative of other introns remains to be determined. In principle, it seems desirable to target larger introns; DNA segments ≤ 500 bp are within the range of maximal resolution of DGGE (Myers et al. 1989*a*, and references therein).

Extensions, Limitations, and Alternatives

This approach is applicable to a broad spectrum of species and loci because PCR can withstand mismatches in the primers (Kwok et al. 1990); if no sequences are known for the species of interest, available sequences of other taxa will often suffice for designing primers (Kocher et al. 1989). Like allozyme electrophoresis, the combination of PCR, DGGE, and direct sequencing provides genotype-frequency data to population geneticists, conservation biologists, and other students of genetic diversity in the wild. Working with these methods at the DNA level, however, has distinct advantages. First, it provides greater sensitivity because of the ability to target a broader range of loci and to detect variation in noncoding regions and in silent mutations in exons. Second, information on sequence variation in nuclear genes allows the construction of "gene trees" and, more generally, the use of phylogenetic approaches to population genetics (Slatkin 1989; Slatkin and Maddison 1989), which have thus far been virtually limited to mitochondrial DNA data (Avise 1989). Third, the methods can be applied to DNA sources other than frozen tissues, including museum specimens (Thomas et al. 1990). Finally, the protocol can easily be extended to other cases in which large samples and/or sorting of variation within individuals are desired, such as population-level surveys of mitochondrial DNA (including the detection of heteroplasmy) and studies of sequence variation within gene families. The main limitations of this approach are (*a*) the costs involved in primer synthesis, PCR, and sequencing; (*b*) the need to target relatively small (≤500 bp) segments, in order to achieve full resolution in DGGE; and (*c*) the need to have sequence data for the target genes of related taxa, in order to design primers.

Obviously, choosing between this and other approaches depends on the research project. Among the possible alternatives, no method can match allozyme electrophoresis in terms of cost efficiency and the ability to study multiple loci. If working at the DNA level is preferred, alternatives include at least the following: (*a*) temperature gradient gel electrophoresis (Wartell et al. 1990), conceptually very similar to DGGE, but using gradients of temperature rather than denaturants; (*b*) single-stranded con-

formational polymorphisms (Orita et al. 1989), primarily for short (<200 bases) PCR products, especially if full resolution is not required; and (*c*) restriction-fragment analysis, which can be combined with DGGE to achieve increased levels of resolution (Myers et al. 1989*a*, and references therein).

Sequence Availability

Sequences have been deposited in GenBank under accession number M62863 and are also available from the author.

Acknowledgments

I am grateful to James L. Patton, Kelley Thomas, Maria P. Echeverriarza, and Cristian Orrego for helpful discussions; to Karen Klitz for drawing the figures; and to Brian Charlesworth, Craig Moritz, Barbara R. Stein, and two anonymous reviewers for comments on the manuscript. This work was supported by NSF grant BSR-88-22345 (to J. L. Patton) and by an Alfred P. Sloan Postdoctoral Fellowship in Molecular Studies of Evolution.

LITERATURE CITED

ABRAMS, E. S., S. E. MURDAUGH, and L. S. LERMAN. 1990. Comprehensive detection of single base changes in human genomic DNA using denaturing gradient gel electrophoresis and a GC clamp. Genomics **7**:463–475.

AVISE, J. C. 1989. Gene trees and organismal histories: a phylogenetic approach to population genetics. Evolution **43**:1192–1208.

GYLLENSTEN, U. B., and H. A. ERLICH. 1988. Generation of single-stranded DNA by the polymerase chain reaction and its implications to direct sequencing of the HLA DQ-alpha locus. Proc. Natl. Acad. Sci. USA **85**:7652–7656.

KOCHER, T. D., W. K. THOMAS, A. MEYER, S. V. EDWARDS, S. PAABO, F. X. VILLABLANCA, and A. C. WILSON. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA **86**:6196–6200.

KWOK, S., D. E. KELLOG, N. MCKINNEY, D. SPASIC, L. GODA, C. LEVENSON, and J. J. SNINKSY. 1990. Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. Nucleic Acids Res. **18**:999–1005.

MANIATIS, T., E. F. FRITSCH, and J. SAMBROOK. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MILLER, S. A., D. D. DYKES, and H. F. POLESKY. 1988. A simple salting procedure for extracting DNA for human nucleated cells. Nucleic Acids Res. **16**:215.

MYERS, R. M., T. MANIATIS, and L. S. LERMAN. 1987. Detection and localization of single base changes by denaturing gradient gel electrophoresis. Methods Enzymol. **155**:501–527.

MYERS, R. M., V. C. SHEFFIELD, and D. R. COX. 1989*a*. Mutation detection, GC-clamps, and denaturing gradient gel electrophoresis. Pp. 71–88 *in* H. A. ERLICH, ed. PCR technology: principles and applications. Stockton, New York.

———. 1989*b*. Polymerase chain reaction and denaturing gradient gel electrophoresis. Pp. 177–181 *in* H. A. ERLICH, R. GIBBS, and H. H. KAZAZIAN, eds. Polymerase chain reaction. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

ORITA, M., Y. SUZUKI, T. SEKIKA, and K. HAYASHI. 1989. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics **5**:874–879.

PATTON, J. L., and M. F. SMITH. 1990. The evolutionary dynamics of the pocket gopher, *Thomomys bottae,* with emphasis on California populations. Univ. Calif. Publ. Zool. **123**: 1–161.

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 amplifications of DNA with thermostable DNA polymerase. Science **239**:487–991.

SHEFFIELD, V. C., D. R. COX, L. S. LERMAN, and R. M. MYERS. 1989. Attachment of a 40-base-pair G+C rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. Proc. Natl. Acad. Sci. USA **86**:232–236.

SLATKIN, M. 1989. Detecting small amounts of gene flow from phylogenies of alleles. Genetics **121**:609–612.

SLATKIN, M., and W. P. MADDISON. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics **123**:603–613.

SMITH, M. F., and J. L. PATTON. 1980. Relationships of pocket gopher (*Thomomys bottae*) populations of the lower Colorado River. J. Mammal. **61**:681–696.

———. 1984. Dynamics of morphological differentiation: temporal impact of gene flow in pocket gopher populations. Evolution **38**:1079–1087.

THOMAS, W. K., S. PAABO, F. X. VILLABLANCA, and A. C. WILSON. 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. J. Mol. Evol. **31**:101–112.

WARTELL, R. M., S. H. HOSSEINI, and C. P. MORAN, Jr. 1990. Detecting base pair substitutions in DNA fragments by temperature gradient gel electrophoresis. Nucleic Acid Res. **18**:2699–2705.

BRIAN CHARLESWORTH, reviewing editor

Received February 26, 1991; revision received September 4, 1991

Accepted September 9, 1991