Modeling of Heteroduplex Formation during PCR from Mixtures of DNA Templates

Gualberto Ruano and Kenneth K. Kidd

PCR Methods and Applications *2:112-116©1992 by Cold Spring Harbor Laboratory ISSN 1054-9803/92 $3.00*

Modeling of Heteroduplex Formation during PCR from Mixtures of DNA Templates

Gualberto Ruano and Kenneth K. Kidd

Department of Genetics, Yale University School of Medicine, New Haven Connecticut 06510

We have investigated the capability of PCR to amplify a specific locus from each template in a mixture of allelic DNAs. Modeling experiments employed a 300-bp HOX2B segment as target and utilized, as distinct template "alleles," genomic DNAs from 1 human and from 2 chimpanzees known to differ in sequence at the target. Two modes were used: (1) mixtures of PCR products that had been previously amplified individually from a single template and (2) PCR amplification en masse from composite human / chimpanzee genomic DNA templates. Products generated by either mode were separated by denaturing gradient gel electrophoresis (DGGE). Detection of a 'trace" allele mixed with a "dominant" one was possible by simple ethidium bromide staining of the gel up to a sensitivity of 1 part in 20. A balanced mixture was represented by 5:5 and 4:3:3 mixtures of allelic PCR products or genomic templates; an uneven mixture of dominant and trace alleles, by 9:1 and 8:1:1 mixtures. For 5:5, two homoduplex bands and two heteroduplex bands of equal intensity were generated. For 9:1, the trace homoduplex disappears whereas the two heteroduplexes are easily visible. For 8:1:1, four heteroduplexes and one homoduplex were observed; homoduplexes and heteroduplexes formed from the trace alleles were not visible. These experiments demonstrate that PCR can amplify mixed allelic templates in direct proportion to the stoichiometric fraction of each template. Trace species are captured as heteroduplexes with the most abundant species and are clearly displaced on denaturing gradient gels from the dominant homoduplex species. Our analytical studies can be applied to analysis of sequence variants in DNA collected from cancerous or infected tissues.

PCR amplification from complex mixtures of DNA templates is often required in medical diagnostics and in forensic science. For instance, because tumor cells are often surrounded by normal tissue, any DNA extracted from a dissected specimen is likely to contain mutant and wild-type forms.[1] Similarly, in studies of infectious processes, a viral pool may contain a predecessor virus as well as newly derived strains.[2] Finally, in many forensic scenarios, biological evidence is likely to contain cells from both victim and aggressor.[3] In all situations, DNA extracted from such composite samples will reflect the heterogeneity of the source. Sequencing the informative bases in mutated oncogenes, in variant viruses, or in heterozygous loci is complicated by the presence of more than one DNA species. These challenging studies require products of PCR amplification representative of each component template as well as analytical procedures capable of assaying each amplified species.

We sought to establish whether PCR amplifies trace templates in stoichiometric proportion while in the presence of more abundant templates. In addition, we investigated the capability of denaturing gradient gel electrophoresis (DGGE) to separate and to enrich each PCR product derived from mixed templates. Products for a single locus amplified from mixtures of templates evidencing sequence variation at the locus will contain both homoduplex and heteroduplex molecules. DGGE is optimally suited for the separation of these molecules, as demonstrated in many recent reports.[4-8] Heteroduplexes contain mismatched base pairs that render them inherently less stable under the denaturing conditions of DGGE.

For these analytical studies, we chose a 300-bp locus 4 kb upstream from HOX2B. This locus has been extensively studied for DNA sequence variation in the higher primates.[9] It is an ideal fragment for this study because of its electrophoretic behavior on DGGE: Individual products from the different species are clearly resolved on DGGE as very sharp bands. We used as "alleles" human and chimpanzee DNA known to be homozygous for the HOX2B fragment as assessed by both DGGE bands and actual sequence.[10] The products were previously amplified individually from a single genomic template and then mixed or were amplified en masse from mixtures of genomic DNA from different sources. Biallelic mixtures were prepared from PCR products or genomic DNA of 1 human and 1 chimpanzee. Triallelic mixtures were prepared from 1 human and 2 chimpanzees or from 1 human and 1 heterozygous chimpanzee.

MATERIALS AND METHODS

Denaturing Gradient Gel Electrophoresis

Linear gradients of denaturants on 8% acrylamide and 1 x TAE were poured ranging from 20% to 70% of a reference

solution prepared at 40% (vol/vol) formamide and 7 M urea. Meltmap predictions [11] and empirical observations showed this range of denaturants gave the best resolution of the fragments. Denaturants ranged approximately from 1.4 M urea and 8% formamide at the top of the gel to 4.9 M urea and 28% formamide at the bottom. During electrophoresis, the gels remained immersed in a tank with recirculated buffer held at a constant 60°C with a constant temperature circulator. Each percentage unit of the denaturant reference solution is roughly equivalent to 0.3°C. The equivalent "temperature was thus 66°C [60°C + (0.3°C/percentage denaturant unit) (20 percentage denaturant units)] at the beginning of the gradient and 81°C at the end. PCR products were boiled to enhance heteroduplex formation and ~200 ng of renatured DNA was loaded onto a gel. Electrophoresis was carried out at 150 V for 15 hr with sets of three gels sandwiched.

Gels were stained with ethidium bromide for visualization of bands under ultraviolet light and photographed. Occasional slanting or curving of the lanes in the gels results from disturbances in the gradient at the moment of pouring the gel or from localized dissipation of the gradient to the chamber buffer during electrophoresis. Though unfortunate for the purist, the slanting is of no consequence to the experimentalist dealing with the cumbersome apparatus as long as the pattern of bands in the gels is unequivocal.

Description and PCR Amplification of the HOX2B Locus

Primers *Pyg2* and *Pyg3* and amplification conditions for a noncoding 304-bp HOX2B fragment were as previously described.[9,12] When PCR was carried out from isolated genomic DNA of either human or chimpanzee, 1 µg of whole genomic DNA was used. For amplifications proceeding from composite templates, the total genomic DNA ranged from 100 ng to 2 µg. The allocation of human versus chimpanzee DNAs to the mixture of templates was variable and is described for each experiment. Differences in concentration of the DNA suitable for amplification among template preparations were unavoidable and led to minor fluctuations in PCR yield between experiments using purportedly identical template amounts.

The human DNA sample was from cell line jk920, a lymphoblastoid cell line established in our laboratory. The most commonly used chimpanzee sample, denoted chimp, was from animal YB81-109. Other chimpanzee DNA samples used in triallelic mixture experiment were from Cohn, denoted chimp', and from an animal heterozygous at the HOX2B locus, Carl. Sequences derived from human jk920 and from chimpanzees YB81-109 and Cohn for the HOX2B fragment have been published elsewhere.[9] In brief, human sequence differs from chimp sequence by two C/T transitions and by one C/A transversion; human from chimp', by the C/A transversion; and chimp from chimp', by the two C/T transitions.

Formation of Heteroduplexes

Formation of heteroduplexes is automatic during the later cycles of a 30cycle program when starting from genomic DNA amounts exceeding 100 ng.[13] Nevertheless, to assure full heteroduplex formation to the extent permitted by the components, the PCR products amplified en masse from composite templates were boiled for 5 mm in a water bath and then allowed to return gradually to room temperature.

PCR products amplified individually from human and chimpanzee genomes were combined in various ratios specific to the experiment. From each mixture, half was electrophoresed on denaturing gradient gels to resolve the component alleles. The other half of the mixture was denatured for 5 mm and allowed to reanneal at room temperature.

RESULTS

In Figures lA and 2A, PCR products amplified individually from human and chimpanzee genomes were utilized. The biallelic experiment (Fig. lA) tests 1:1 and 9:1 mixtures. In the 1:1 mixture, both homoduplexes are reduced in intensity by one-half after heteroduplex formation. When individual products are mixed and run on denaturing gradient gels, they are separated. When the mixture is first denatured and allowed to reanneal at room temperature (H lanes, on the left of each pair), heteroduplex formation is induced. In the 9: 1 mixture, the trace chimpanzee "allele" is almost fully consumed by heteroduplex formation. Assuming random reassociation of single strands, each of the two heteroduplexes from a 1 : 1 mixture should contain 25% of the DNA (0.5 x 0.5); from a 9:1 mixture, each heteroduplex should contain 9% (0.9 x

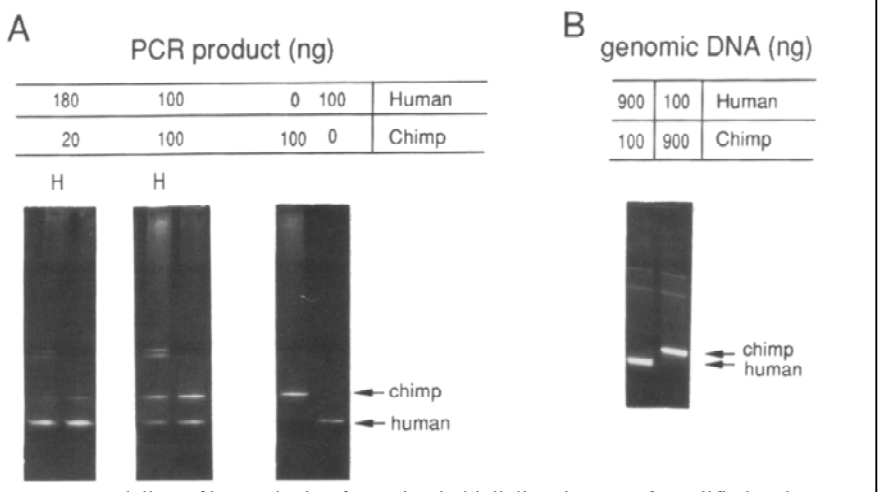

FIGURE 1 Modeling of heteroduplex formation in biallelic mixtures of amplified and genomic DNA. *(A)* PCR products for the segment at the HOX2B locus were amplified individually from genomic DNA of 1 human and 1 chimpanzee. Individual products are clearly resolved on DGGE. When individual products are mixed and run on denaturing gradient gels, they are separated. When the mixture is first denatured and allowed to reanneal at room temperature (H lanes, on the left of each pair), heteroduplex formation is induced. *(B)* Mixtures of chimpanzee and human genomic DNA are amplified en masse in the same reaction. Products are then electrophoresed on denaturing gradient gels. Amount of total product loaded was ~200 ng.

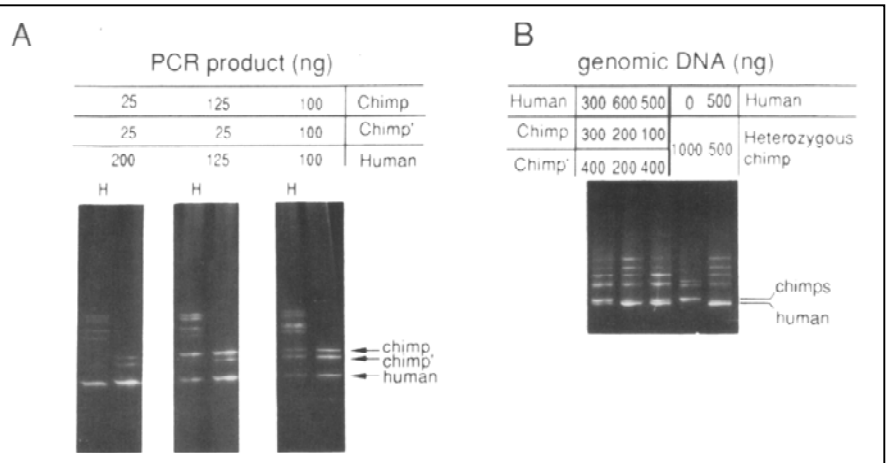

FIGURE 2 Modeling of heteroduplex formation in triallelic mixtures of amplified and genomic DNA. *(A)* HOX2B PCR products were amplified individually from genomic DNA of 1 human and 2 chimpanzees homozygous for either allele of a polymorphism at the locus. All three individual products are resolved after DGGE. When the mixture is first denatured and allowed to reanneal at room temperature *(H* lanes), heteroduplexes are evident. *(B)* Mixtures of chimpanzee and human genomic DNA were amplified en masse in the same reaction. The first three lanes depict PCR product obtained from amplification of genomic DNA mixtures prepared from 1 human and from 2 homozygous chimpanzees. Products in the last two lanes resulted from en masse amplification of genomic DNA pools composed of 1 human and 1 heterozygous chimpanzee. Products were electrophoresed on denaturing gradient gels and loaded as in Fig. lB.

0.1). Thus, the expected intensity difference between heteroduplexes in 1 : 1 versus 9: 1 mixtures is only half that expected from the relative concentrations of alleles alone, namely, 2.75/1 (25%/9%) rather than 5/1 (50%/ 10%). By eye, this is what is observed. The triallelic experiment (Fig. 2A) tests 1:1:1, 5:5:1, and 8:1:1 mixtures. Six heteroduplexes are seen in the equimolar mix but only four are seen in the 8:1:1 mix. Homoduplexes for the two trace alleles are not seen in the 8:1:1 mix; DNA from both trace elements is captured as heteroduplexes with the dominant species.

In Figures lB and 2B, mixtures of chimpanzee and human genomic DNA were amplified en masse in the same reaction. The biallelic experiment (Fig. lB) tests both 9:1 and 1:9 combinations of human and chimpanzee genomic DNA. The genome represented at 10% total DNA is not detected in homoduplex form in either combination. Heteroduplex formation, nevertheless, is the same in either reciprocal mixture. The triallelic experiments (Fig. 2B) include a heterozygous chimpanzee and 2 homozygous chimpanzees as well as a human DNA sample. The pattern obtained from 500 ng of heterozygous chimpanzee genomic DNA mixed with 500 ng of human DNA is identical to that obtained with 200 ng of DNA from both of 2 homozygous chimpanzees and 600 ng of human DNA.

We sought to establish the sensitivity of heteroduplexes for detecting infrequent "trace" alleles. The problem investigated here had two components. First, we had to model situations where the total DNA loaded in the gel is the same regardless of the heterogeneity among components of the mixture. The detection level for a trace allele when the total DNA loaded on the gel is approximately 200 ng was explored. Second, we studied the ability of PCR to amplify trace alleles in the presence of a dominant one. Of particular concern was competition by the dominant allele for the reagents and/or enzyme, especially in the first few cycles, and the possibility that the trace allele might not be amplified because of this effect.

In Figure 3A, PCR product mixtures were prepared with a dominant human product held constant at 200 ng and with decreasing amounts of trace chimpanzee product. Mixtures are compared on DGGE before and after heteroduplex formation (H lanes). The ratio of dominant to trace (d/t) product extended from 4:1 to 40:1. The level of detection sensitivity approximates the usual limit on ethidium bromide gels, 5-10 ng of DNA.

Genomic DNA mixtures were amplified en masse in two modes (Fig. 3B,C). In one mode, the trace DNA is constant while the dominant DNA progressively increases; in the other, the dominant DNA is constant while the trace DNA progressively decreases. In Figure 3B, the dominant allelic DNA (chimpanzee) is increased while the trace genomic DNA (human) is maintained at a constant amount. Beginning at d/t of 4:1, the trace homoduplex disappears. The amount of heteroduplex formed remains virtually unchanged even at a d/t of 19:1. The amount of chimpanzee homoduplex is virtually a linear function of the starting template amount. In contrast, Figure 3C shows PCR products obtained from a dilution series of human genomic DNA amplified in isolation (left) or in the presence of a constant amount of dominant chimpanzee genomic DNA. In the DNA mixtures of Figure 3C, dominant allelic DNA from chimpanzee remains constant while trace DNA from human is gradually decreased; the d/t ratio ranged from 1:1 to 1:0.037. The PCR yield reflects the dilution series of human genomic template in both cases. The limits of PCR product detection are also the same despite the trend, evident in the more balanced mixtures, for a slightly higher yield from the chimpanzee DNA preparation utilized in this series. The amount of dominant homoduplex increases as the trace DNA is reduced.

DISCUSSION

The detection limits of trace components in a mixture of DNA templates amplified en masse will be dependent on at least two factors. The first factor is the extent to which the amplification products from pooled templates represent each component genome. Of particular concern is the possibility of differential amplification of alleles or even allele dropout. [14] Second, the detection limits of DGGE itself are of concern. It is known that DGGE will not reveal polymorphisms in the upper or more stable melting domain which remains doublestranded when more labile domains become singlestranded. Let us consider these factors in turn.

Our analytical mixing experiments with both PCR products and with genomic DNAs demonstrate that PCR products amplified from pooled templates can be representative for every template present even when a trace element

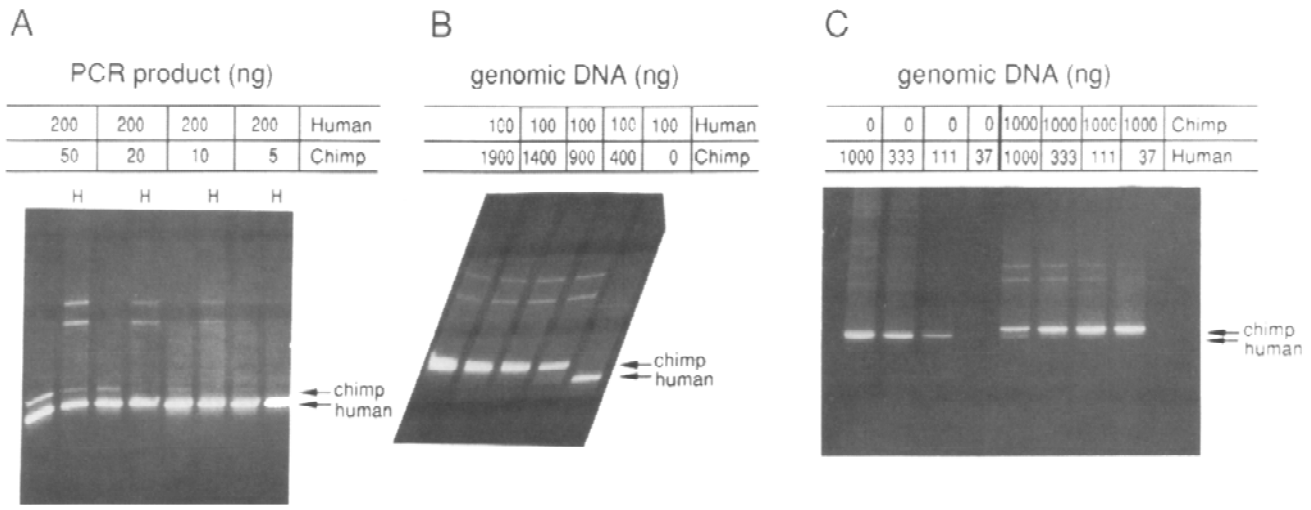

FIGURE 3 Sensitivity in detection of a trace allele in a mixture containing increasing amounts of dominant allele. *(A)* PCR product mixtures of "dominant" human product held constant at 200 ng and of decreasing amounts of "trace" chimpanzee product. Mixtures are depicted on DGGE before and after heteroduplex formation (H). Ratios of dominant to trace (d/t) product are 4:1, 10:1, 20:1, and 40:1. *(B)* Genomic DNA mixtures amplified en masse. The trace genomic DNA was maintained at a constant amount (100 ng of human DNA) while the dominant allelic DNA (chimpanzee) ranges from 0 to 1900 ng. Products were electrophoresed on denaturing gradient gels and loaded as in Fig. lB. *(C)* Genomic DNA mixtures amplified en masse. The panel shows PCR products obtained from a dilution series of human genomic DNA amplified in isolation *(left)* or in the presence of constant 1000 ng of dominant chimpanzee genomic DNA. Products were electrophoresed on denaturing gradient gels and loaded as in Fig. lB.

accounts for only 1 in 20 elements of the pool. As long as ~200 ng PCR product is loaded on a gel, the trace allele in such a mixture can be detected as heteroduplexes by simple ethidium bromide staining of the gel (Fig. 3A). If the products are labeled radioactively or fluorescently, PCR products from templates at even higher dilution in a mixture could be detected by autoradiography or fluorescent light emission.

The results obtained with the test fragment have been successfully applied to two other segments. The level of sensitivity has been validated in an intronic 440bp fragment of the glucose-6-phosphate dehydrogenase gene where a trace DNA constituting 10% of a mixture was routinely detected. [15] In a mixture of human genomic templates, the proportionality between the stoichiometric fraction of a given template and its PCR product yield has been observed for a 300-bp fragment in the tissue plasminogen activator gene evidencing a polymorphic 300-bp *Alu* insertion. [16] We expect the current results to be generally applicable to any PCR fragment devoid of differential amplification artifacts. Complications are likely to occur during PCR from dilute or contaminated DNA extracted from pathological or forensic specimens but those can be ameliorated by extensive preamplification heating and by assembly of the reaction mixture at high temperature. [17]

Heteroduplex formation, in principle, could overcome limitations of detection inherent to DGGE. The destabilization afforded by a mismatch in a heteroduplex may be as high as 6°C. [18] Yet the difference between melting domains in fragments approximately 400bp rarely exceeds 4°C. Under these typical conditions, a mismatch in the heteroduplex at the region constituting the upper or anchor domain of the homoduplex may render this region 2°C more unstable than the rest of the heteroduplex molecule. Sequence differences in the anchor domain may thus be detectable. [8]

One further advantage of the large destabilization conferred by mismatches is that heteroduplexes will be clearly detectable even under suboptimal technical conditions. The presence of heteroduplexes in deliberately mixed samples indicates heterogeneity in the pool. Optimization of electrophoretic conditions to detect melting variants representative of multiple domains in a fragment may be less critical for the heteroduplex-based approach than for one dependent on separating homoduplexes. In complex mixtures, however, resolution of an individual heteroduplex band may be difficult among the many heteroduplexes formed. (The number of hetero duplexes arising from *n* distinct allelic sequences is $n(n - 1)$.) Denaturing gradient gels with narrow gradient ranges may be required for separation of multiple heteroduplexes.

The relative abundance of a given homoduplex or heteroduplex can be calculated theoretically from the product of the stoichiometric fraction of each component species. Duplexes formed from the least abundant DNA strands will be scarce. For example, in a 9:1 mixture, 1% of the product will be trace homoduplex compared to 9% for each of the two heteroduplexes and to 81% for the dominant homoduplex. The two heteroduplexes contain 90% of the trace product. We have shown here that most of the trace PCR product is indeed heteroduplexed with the dominant sequence.

This feature raises important caveats for the analysis of uneven allelic mixtures based on homoduplex DNA alone. The importance is compounded by the

inevitable generation of heteroduplexes during the later inefficient cycles of PCR that merely rearrange previously synthesized single DNA strands into double-stranded products. [13] For example, successful detection of variant sequences forming new restriction enzyme sites (RFLPs) will depend on whether the heteroduplex is amenable to digestion by the specified enzyme. If the heteroduplex cannot be cut by the enzyme, the diagnostic power of RFLP is lost because the trace species has been effectively removed. Similarly, measuring the length of minisatellite VNTR variants in nondenaturing acrylamide or agarose gels will be compromised. The trace species will be complexed into heteroduplexes, which, because of length differences in the mismatched region, are likely to have aberrant electrophoretic properties. In both instances, analysis of single-stranded DNA is preferable. In the RFLP case, allele-specific annealing of one strand to an oligonucleotide probe in hybridization assays [19] or to a PCR primer in amplification assays [20] are valid options. In the VNTR case, high-resolution electrophoresis on denaturing acrylamide gels is optimal.[21]

In analytical separations of trace DNAs from mixtures with abundant species, heteroduplex formation could be used to capture the trace element. Because of their separation from homoduplexes, heteroduplexes usually can be easily excised and recovered from the gel. This feature conveys a preparative advantage as well: In heteroduplex DNA, the formerly trace species is as abundant as the previously dominant species. Reamplification using heteroduplex DNA as template followed by a second DGGE with a narrow gradient, should allow recovery of pure trace homoduplex. The trace homoduplexes could then be sequenced for full characterization. Thus, this heteroduplex-based approach to separation involving two reamplifications after two rounds of DGGE would first enrich and then isolate trace DNA.

ACKNOWLEDGMENTS

We are grateful to Dr. S.T. Reeders (Yale University) for useful discussions. This research was supported by National Institutes of Health grant HG00365 and National Science Foundation grant BNS 8813234 awarded to K.K. Kidd. G. Ruano was the recipient of a Dissertation Fellowship from the Ford Foundation.

REFERENCES

1 Brash, D.E., J.A. Rudolph, J.A. Simon, A. Lin, GJ. McKenna, H.P. Baden, AJ. Halperin, and J. Ponten. 1991. A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. *Proc. Natl Acad. Sci.* **88:** 10124-10128.

2. Wolinsky, S.M., C.M. Wike, B.T.M. Korber, C. Hutto, W.P. Parks, L.L. Rosenblum, KJ. Kunstman, M.R. Furtado, and J.L. Munoz. 1992. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *Science* **255:** 1134-1137.

3 Brinkman, B. 1991. Population studies on selected AMPFLPs and their use in the investigation of mixtures of body fluids. *Crime Lab Digest* **18:** 153-155.

4. 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.* **86:** 232236.

5. Cai, S.P. and Y.W. Kan. 1990. Identification of the multiple β-thalassemia mutations by denaturing gradient gel electrophoresis. *J. Clin. Invest.* **85:** 550-553.

6. Kogan, S. and J. Gitschier. 1990. Mutations and a polymorphism in the factor VIII gene discovered by denaturing gradient gel electrophoresis. *Proc. Natl. Acad. Sci.* **87:** 2092-2096.

7. Traystman, M.D., M. Higuchi, C.K. Kasper, S.E. Antonarakis, and H.H. Kazazian. 1990. Use of denaturing gradient gel electrophoresis to detect point mutations in the factor VIII gene. *Genomics* **6:** 293-301.

8. Sung, C.H., C.M. Davenport, J.C. Hennessey, K.H. Maumenee, S.G. Jacobson, J.R. Heckenlively, R. Nowakowski, G. Fishman, P. Gouras, and J. Nathans. 1991. Rhodopsin mutations in autosomal dominant retinitis pigmentosa. *Proc. Natl. Acad. Sci.* **88:** 6481-6485.

9. Ruano, G., J. Rogers, A.C. Ferguson-Smith, and K.K. Kidd. 1992. DNA sequence polymorphism within hominoid species exceeds the number of phylogenetically informative characters for a HOX2 locus. *Mol. Biol. Evol.* **9:** 575-586.

10. Ruano, G. and K.K. Kidd. 1991. Genotyping and haplotyping of polymorphisms directly from genomic DNA via coupled amplification and sequencing (CAS). *Nucleic Acids Res.* **19:** 6877-6882.

11. Lerman, L.S., S.G. Fischer, I. Hurley, K. Silverstein, and N. Lumelsky. 1984. Sequence-determined DNA separations. *Annu. Rev. Biophys. Bioeng.* **13:** 399-423.

12. Ruano, G. and K.K. Kidd. 1991. Coupled amplification and sequencing of genomic DNA. *Proc. Natl. Acad. Sci.* **88:** 2815-2819.

13. Ruano G., D.E. Brash, and K.K. Kidd. 1991. PCR: The first few cycles. *Amplifications* **7:** 1-4.

14. Walsh, P.S., H.A. Erlich, and R. Higuchi. 1992. Preferential PCR amplification of alleles: Mechanisms and solutions. *PCR Methods Applic.* **1:** 241-250.

15. Ruano, G. 1992. "A PCR-based paradigm for the analysis of DNA sequence variation." Ph.D. thesis, Yale University.

16. Tishkoff, S.A., G. Ruano, and K.K. Kidd, in preparation.

17. Ruano, G., E.M. Pagliaro, T.R. Schwartz, K. Lamy, D. Messina, R.E. Gaensslen, and H.C. Lee. 1992. Heat-soaked PCR: An efficient method for DNA amplification with applications to forensic analysis. *BioTechniques* **13:** 266-274.

18. Myers, R.M., V.C. Sheffield, and D.R. Cox. 1988. Detection of single base changes in DNA: ribonuclease cleavage and denaturing gradient gel electrophoresis. In: *Genome analysis* (ed. K.E. Davies) pp. 95-139. IRL Press, Oxford.

19. Saiki, R.K., T.L. Bugawan, G.T. Horn, K.B. Mullis and H.A. Erlich. 1986. Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. *Nature* **324:** 163-166.

20. Sommer, S.S., A.R. Groszbach, and C.D.K. Bottema. 1992. PCR amplification of specific alleles is a general method for rapidly detecting known single-base changes. *BioTechniques* **12:** 82-87.

21. Ludwig, E.H., W. Friedl and BJ. McCarthy. 1989. High-resolution analysis of a hypervariable region in the human Apolipoprotein B gene. *Am. J. Hum. Genet.* **45:** 458-464.

Received August 3, 1992; accepted in revised form September 16, 1992