Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 6296–6300, August 1990
Genetics

Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules

(single-molecule dilution/polymerase chain reaction/β-globin cluster)

GUALBERTO RUANO*, KENNETH K. KIDD*†, AND J. CLAIBORNE STEPHENS*†

*Department of Human Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510; and †Yale–Howard Hughes Medical Institute Human Gene Mapping Library, 25 Science Park, New Haven, CT 06511

Communicated by Wyatt W. Anderson, May 31, 1990 (received for review November 1, 1989)

ABSTRACT **We have developed a reliable method for the direct resolution of haplotypes or linkage phase from individuals who are multiply heterozygous in a given genomic region. The method is based on single-molecule dilution (SMD) of genomic template and amplification via biphasic polymerase chain reaction (booster PCR). We have verified the feasibility of the SMD method for a highly polymorphic region within the β-globin cluster by analysis of triply heterozygous individuals of known haplotype. This approach should be useful in many studies in population or evolutionary genetics and in a variety of clinical settings.**

Over the past half century, the discipline of population genetics has evolved from the theory-driven era of Fisher, Haldane, and Wright into an enterprise driven by molecular data. This transformation was punctuated by experimental insights and breakthroughs that allowed the phenotypes being monitored to more directly represent their underlying genetic basis. Critical phases in this progression are work on *Drosophila pseudoobscura* chromosomal inversions by Dobzhansky and his colleagues (1), protein electrophoretic studies (2), mitochondrial DNA studies (3), and now the potential for direct analysis of DNA variants.

The tremendous wealth of variability at the DNA level would seem to provide the potential to answer many long-standing questions in population genetics (4). However, this level of resolution also poses many new questions. In particular, a view of polymorphic sites as simple independent markers is inadequate in discussions of the evolutionary history of any given gene or genomic region. The number of polymorphic sites and the linkage disequilibrium among them requires such discussions to focus on the DNA "haplotypes" at a locus. A specific DNA haplotype is the specific combination of variants at each of the polymorphic sites being studied. If one thinks of haplotypes as being the alleles at a locus, there is now an added dimension compared to, say, electrophoretic alleles—the possibility of addressing evolutionary relationships among alleles.

The high levels of polymorphism and consequent heterozygosity pose a problem in this regard—ambiguity of which haplotypes are present in any given diploid organism. This is due to the frequent occurrence of multiple heterozygous sites in a given individual. With two heterozygous sites there are four possible haplotypes and the number of possible haplotypes doubles for each additional heterozygous site in the individual. Ideally, an individual's genotype across multiple heterozygous loci should include resolution into the two constituent haplotypes.

Classically, haplotypes in multiply heterozygous individuals or organisms have been resolved by pedigree analysis, breeding programs, or such special attributes as a haploid sex, depending on the organism. In humans, family studies have been the basis for determining haplotypes at loci such as phenylalanine hydroxylase (5) and the β-globin cluster (6). Recently, molecular cloning and restriction site mapping (7–9) or DNA sequencing (4, 10) of relevant clones has been used to define precise haplotypes of multiallelic chromosomal regions. These strategies, however, are too laborious to produce the large sample sizes typically required in a population study for a given species. For instance, resolution of haplotypes by such conventional molecular means would entail creating and screening a library for each individual so that clones containing either paternal or maternal alleles could be analyzed.

Another alternative is to take advantage of any association, physical or otherwise, among polymorphisms that might allow sequential resolution of alleles at each polymorphic site. For instance, sites of nucleotide substitution that are nearby sites of length variation (e.g., variable number of tandem repeats or dinucleotide repeats) can be coamplified with polymerase chain reaction (PCR). In this case, amplification would resolve allelic size variants, and alleles at the polymorphic nucleotide site could be resolved subsequently by dot blot or other means. This approach has been used successfully to resolve haplotypes of the human HLA-DQ region (11). One can also imagine separating allelic bands on the basis of a restriction fragment length polymorphism (RFLP) or on a denaturing gradient gel, with secondary typing of the separated bands. We have also used allele-specific amplification to obtain haplotypes directly (12). In this case, PCR primers are designed to anneal at their 3′ ends to one of the polymorphic sites: one primer is specific for each polymorphic base. Amplification is allele specific, as only the perfectly matched primer sustains amplification. The other polymorphic sites nested between primers are typed and haplotypes are determined from the allele-specific PCR product (12). Such techniques have the advantage of being applicable to genomic DNA but do not seem to be applicable to regions much larger than a few kilobases. The allele-specific amplification method suffers from the need for strict empirical optimization of amplification conditions to ensure allele specificity. To overcome these problems, we developed the single-molecule dilution (SMD) method described below as a general method of resolving ambiguous arrangements of polymorphic markers into definitive haplotypes for each chromosome.

We reasoned that alleles could be stochastically separated into maternal and paternal contributions by dilution of genomic DNA until only a single molecule of the desired region was expected in an aliquot. For humans, a millionfold dilution from 3 μg per aliquot would bring DNA concentration down to one haploid equivalent per aliquot. As suggested by

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.

Abbreviations: SMD, single-molecule dilution; PCR, polymerase chain reaction; ASO, allele-specific oligonucleotide; RFLP, restriction fragment length polymorphism.

our theoretical studies (13), we designed our studies around analysis of 10 aliquots of the final dilution to assay both alleles and provide replicates for each. Amplification from one molecule of template, possible with conventional PCR (14), would follow, using "booster" PCR (15), a more efficient biphasic approach that reduces artifactual formation of so-called primer dimers. If product amplified reliably from a single molecule of genomic template, it should be possible to extract directly the haplotype encompassed between primers. To verify our method, SMD was performed initially on individuals unequivocally haplotyped by pedigree analysis for relevant polymorphisms.

MATERIALS AND METHODS

Target for Amplification. The 770-base-pair (bp) inter-*Alu* region between human $\psi\beta$- and δ-globin genes, numbered in Fig. 1 according to the sequence of Maeda *et al.* (16), was amplified by using the PCR (17) with primers (amplimers) GR6 and GR5, which anneal 10 nucleotides away from flanking *Alu* regions. Allele-specific oligonucleotides (ASOs) GR1 and GR3 were designed with the polymorphic bases at the 3' end to enhance destabilization when annealed to the mismatched allele. GR1/GR3 and GR6 were synthesized to anneal to the strand extending 5' to 3' from δ to $\psi\beta$. GR2/GR4 and GR5 were synthesized to anneal to the complementary strand. Sequences of oligonucleotides GR1–GR6 are given in Fig. 1. Genomic clone CJ52 (kindly provided by B. G. Forget, Yale University), which overlaps most of the distal half of the segment, served to verify the specificity of the amplification process. CJ52 was used to probe the *Taq* I RFLP on genomic and on amplified DNA. For the inter-*Alu* segment, this clone recognizes distinctly only a 540/430-bp polymorphism since it does not overlap the constant band and encompasses only half the nucleotides in the 110-bp band of the + allele.

Synthesis of Oligonucleotide Primers and Probes. Oligonucleotides were prepared by the solid-phase phosphoramidite method in an automated DNA synthesizer (Applied Biosystems) and purified by electrophoresis in a 7 M urea/20% polyacrylamide gel.

Standard PCR. One microgram of genomic DNA from each individual was amplified enzymatically in a 100-μl reaction mixture, which included 10 pmol each of amplimers GR5 and GR6, 200 μM each deoxynucleotide (Pharmacia), 2 units of *Thermus aquaticus* DNA polymerase (*Taq* polymerase) (Cetus AmpliTaq), 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris·HCl (pH 8.4), 0.01% gelatin, overlaid with mineral oil. PCR was performed in a programmable thermal cycler (Perkin–Elmer) for 30 cycles, each consisting of 1 min of denaturation at 94°C, 2 min of annealing at 45°C, and 2 min of polymerization at 72°C.

Booster PCR. Template and amplimers GR5 and GR6 were diluted as specified below to maintain an initial 10^7-fold molar excess of each primer to template. This reaction mixture was subjected to 20 cycles of amplification during stage I of booster PCR. At the end of this initial stage, 10 pmol of each amplimer was added to each vial to bring primer concentration up to 0.1 μM. A further 50 cycles of amplification constituted stage II of booster PCR. Cycle profile and reaction buffer were as described above for standard PCR, except for longer annealing and polymerization during stage I (15). In general, it will likely be necessary to optimize booster PCR conditions separately for each region to be studied, but this is true even for standard PCR procedures starting with microgram amounts of genomic template. Any primer pair with high specificity that functions robustly in standard PCR (product yield of 100 ng or more) should be usable for SMD analysis.

Dilutions. A stock solution of 3 μg of whole DNA (10^6 copies of genomic template) from heterozygote B (see Fig. 2) and 0.1 μM each primers GR5 and GR6 was diluted 1:1000 in reaction buffer to yield an expected 1000 copies (3 ng of DNA) and initial 0.1 nM primers (see Fig. 3, lane 1). Two further 1:10 serial dilutions in reaction buffer yielded, respectively, 100 copies of template (300 pg of DNA), 10 pM primers (lane 2) and 10 copies (30 pg of DNA), 1 pM primers (lane 3). Because of the stochasticity inherent in the final 1:10 dilution to achieve one molecule of genomic template per vial (on average), 10 replicates (lanes 4–13) were amplified. Under the Poisson distribution with an average concentration of one molecule of template per vial, 36.8% of the vials would not be expected to contain a molecule of desired template, another 36.8% would be expected to contain a single molecule, and the remainder would contain multiple molecules. Stephens *et al.* (13) give further theoretical considerations relevant to experimental design.

Electrophoresis. Ten microliters of PCR product was run at 35 V for 9 hr on a 2.5% agarose gel (1.5% SeaKem, 1% NuSieve; FMC). Lanes * in Figs. 3 and 4 contain molecular size markers of 770, 500, and 250 bp. The figures show photographs of ethidium bromide-stained gels.

Dot Blots. Eight microliters of PCR product was alkali denatured and spotted on nylon blotting membrane (Hybond) in a Mini-Fold dot blot apparatus (Schleicher & Schuell). Four replicate strips were prepared for every person, and each strip was hybridized to one of the four different ASOs. Prehybridizations and hybridizations were performed in 10 ml of solution (6× SSPE/5× Denhardt's solution/0.2% SDS; 1× SSPE = 0.18 M NaCl/10 mM phosphate, pH 7.4/1 mM EDTA; 1× Denhardt's solution = 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone) in a 50-ml plastic tube. After purification in a G25-50 (Sigma) spin column, 5 pmol of each ASO, end labeled at 2–4 μCi/pmol (1 Ci = 37 GBq), was hybridized overnight at 42°C to its respective strip. Dot blots were washed in 6× SSPE/0.2% SDS (15 min each at room temperature and at 54°C) and exposed to film (Kodak XAR-5) for 6–12 hr.

RFLP. Twenty microliters of PCR product was run on an 8% acrylamide gel. The 770-bp band was excised, eluted, and digested with 10 units of *Taq* I (New England Biolabs). Digested product was electrophoresed on a 2.5% agarose gel as described above. In Fig. 2, the gel was blotted on nylon membrane and hybridized for 12 hr to CJ52 [radiolabeled with random hexamers (18)]. In Figs. 3 and 4, the gel was stained with ethidium bromide and photographed.

Prevention of Contamination. Diluted DNA aliquots and reaction mixtures were prepared under semi-sterile conditions following published guidelines (19) to avoid contamination. Some specific precautions were as follows. Small volumes (50 μl or less) of the stock solutions for oligonucleotides, enzyme, deoxynucleotides, and buffer were aliquoted into several vials. Those vials opened for a given experiment were discarded after even partial consumption. A separate set of pipettes was devoted exclusively to preparation of reaction mixtures and was kept in a separate station of the laboratory isolated from electrophoresis and hybridization units. Several blank controls lacking template DNA were run in each experiment and yielded no product, demonstrating that contamination was not a problem. However, on a larger scale for the analysis of numerous samples, it is expected that SMD would place a great demand on technique in order to prevent contamination, a problem not unlike that seen in clinical diagnostic and forensic settings.

RESULTS

Amplification and Typing of the Genomic Target. SMD was tested in a highly polymorphic region between human $\psi\beta$- and δ-globin genes (Fig. 1*a*). In a multiethnic sample, Maeda *et al.* (10) have sequenced 3.1 kilobases (kb) of this region and found 18 polymorphic sites, but only as two haplotypes, *R* and *T*. We selected a 770-bp inter-*Alu* segment for typing of

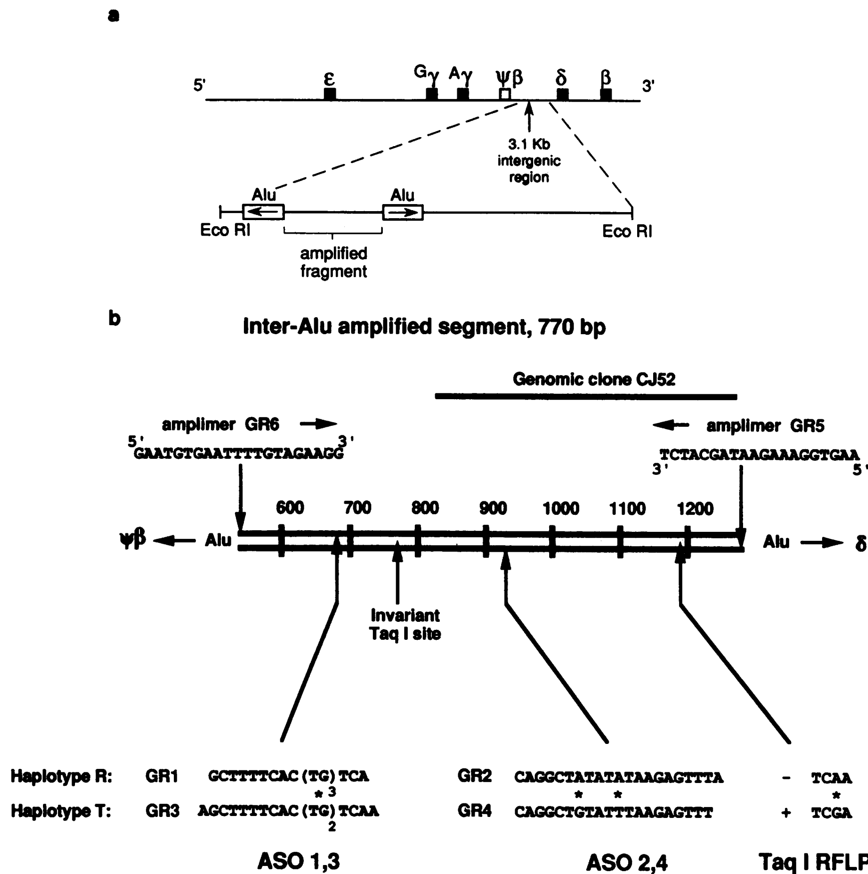

FIG. 1. Inter-*Alu* amplified segment in the β-globin cluster (*a*) and polymorphisms typed with ASOs and a RFLP (*b*). ASO systems 1,3 and 2,4 as well as a *Taq* I RFLP constitute the polymorphisms analyzed in this study (there is another known polymorphism at 795 not typed here). ASO system 1,3 types a TG deletion (*) within a TG repeat. ASO system 2,4 includes two separate single-nucleotide polymorphisms (*). The RFLP involves a single base substitution (*) at the *Taq* I restriction site, which yields two restriction fragment patterns: bands of 430 and 110 bp if the *Taq* I site is present (+ allele) or a 540-bp band if the restriction site is absent (− allele). The segment also contains an invariant *Taq* I site, which yields a 230-bp constant band. Maeda *et al.* (10) have sequenced 3.1 kb from the intergenic region between ψβ- and δ-globin genes, which includes the inter-*Alu* segment, in six individuals of different races and observed that 16 base substitutions and 2 deletions fall exclusively into either of two haplotypes, *R* or *T*. The *R* haplotype for the amplified segment consists of alleles 1,2,−; *T* consists of 3,4,+.

three polymorphic sites by using RFLPs (20) or ASOs (21), which remained annealed to their respective allelic sequence at the highest stringency of hybridization and wash (Fig. 1*b*).

The 770-bp inter-*Alu* piece, amplified enzymatically (17) from whole DNA of nine individuals belonging to a Caucasian pedigree (Fig. 2*a*), was typed for a dinucleotide deletion (ASO 1,3 in Fig. 2*b*); for 2 base substitutions 3 nucleotides apart (ASO 2,4 in Fig. 2*c*); and for a base substitution at the *Taq* I restriction site (RFLP +, − in Fig. 2*d*). On the basis of the segregation pattern of these three polymorphic markers, haplotypes could be deduced for each pedigree member (Fig. 2*e*). Notice that individuals B, D, G, and H are heterozygous at all three loci and hence their haplotypes cannot be resolved directly from typings of their PCR product amplified from whole genomic DNA. The four children (D, E, F, and G) establish the phase in B: one chromosome is 3,4,+ (transmitted to D and G) and the other is 1,2,− (transmitted to E and F). Phase analysis for multiple heterozygote G also bears out the haplotype 3,4,+ for one of its chromosomes since offspring I is homozygous 3,4,+. Furthermore, I confirms the 3,4,+ haplotype for B since G inherited alleles 3,4,+ from B. Hence, variation in this family occurs as only two of eight possible haplotypes, in agreement with the *R*,*T* hypothesis (10). These experiments also validate our ASO and RFLP methodology, since each probe recognized only its specific allele. Allele-specific amplification of this family provides additional partial confirmation of these haplotypes (12).

SMD of Genomic Template. Following pilot experiments, SMD was performed systematically on heterozygotes B and G, and the first such experiment from B is shown in Fig. 3*a*. In this experiment, 4 SMD samples (lanes 5–7, and 10) of 10 (lanes 4–13) amplified, in reasonable agreement with expectations from our template concentration of one haploid equivalent per reaction vial. Even if our template concentration were achieved precisely and amplification from one molecule were 100% efficient, only 63% of all vials from this stock solution would be expected to amplify, assuming a Poisson distribution of target molecules per vial. From a given 10-vial experiment with this expectation, anywhere from 4 to 9 vials might amplify. The 40-bp band is a dimer of primers, a common artifact when amplifying very dilute DNA samples (15). The band of ≈250 bp in lane 13 is a nonhomologous spurious product. In other experiments, different random artifacts were observed, but probing the gel with CJ52 invariably revealed the presence of a single homologous product in those vials that amplified target.

All diluted samples were amplified by using booster PCR (15), a biphasic method that closely titrates initial primer and template concentration. All products are clearly visualized in Fig. 3*a*; probing a blot of this gel with CJ52 revealed no trace of product in the empty lanes nor any homology to the 250-bp band in lane 13 (data not shown). However, amplification yield is variable due, presumably, to primers "catching" template at different cycles and to the inverse relationship between yield and primer dimer synthesis. Total 770-bp product yield in lanes 5–7 and 10 ranges from ≈100 ng to 1.5 μg, with a mean of 1 μg (≈2 pmol), which corresponds approximately to a 10^{12}-fold amplification from a single molecule of genomic template.

To ascertain the success of the dilution process in separating alleles, products from lane 1 (1000 copies of genomic template) and from lanes 5–7 and 10 (potentially from single molecules of template) were typed for the *Taq* I RFLP (Fig. 3*b*). The sample is heterozygous in lane 1, as expected, but hemizygous for the 540-bp band (− allele) in lanes 5 and 7, and hemizygous for the 430-bp band (+ allele) in lanes 6 and 10. Thus, a heterozygous sample was rendered hemizygous by dilution.

Typing and Direct Haplotype Determination of SMD Products. Seven additional SMD experiments were done on DNA from individuals B (five experiments) and G (two experiments) to resolve their two constituent haplotypes; each 10-aliquot experiment was performed following the scheme presented in

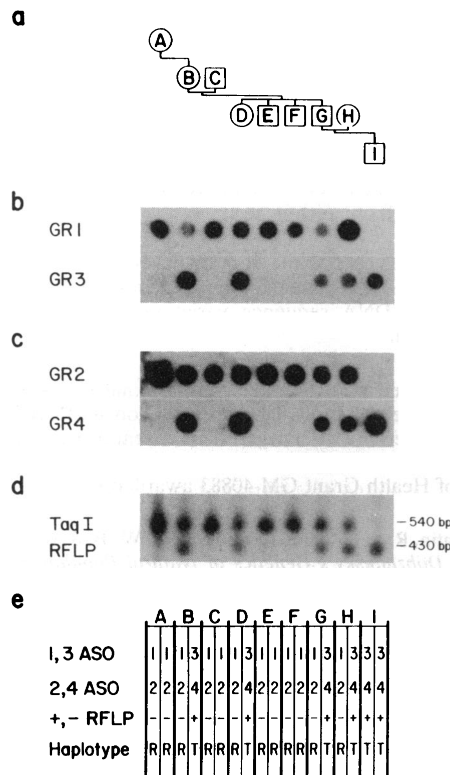

FIG. 2. Typing of polymorphic loci in a Caucasian pedigree and establishment of haplotypes by classical means. The inter-*Alu* segment was amplified by PCR (14, 17) for each member of a four-generation Caucasian family from our Tourette syndrome/Oregon reference pedigree (*a*). Each reaction started from 1 μg of whole DNA (≈3 × 10^5 haploid equivalents). PCR products were typed for ASO system 1,3 (oligonucleotides GR1 and GR3; *b*), for ASO system 2,4 (oligonucleotides GR2 and GR4; *c*), as well as for the *Taq* I RFLP (540 bp is −, 430 bp is +; *d*). (*e*) From the segregation of markers, haplotypes 1,2,− and 3,4,+ were deduced (*e*).

Fig. 3. In these seven experiments, product was detected (by Southern blotting and hybridization to probe CJ52) in 0 of 10, 1 of 10, 2 of 10, 4 of 10 (three experiments), and 5 of 10 vials. The overall success rate for amplification across all eight experiments is 24/80, which is lower than the 63% expected under a Poisson distribution. This could be due to a concentration of genomic template slightly lower than one molecule per reaction vial. Alternatively, the amplification reaction may fail in some aliquots, in spite of the presence of target.

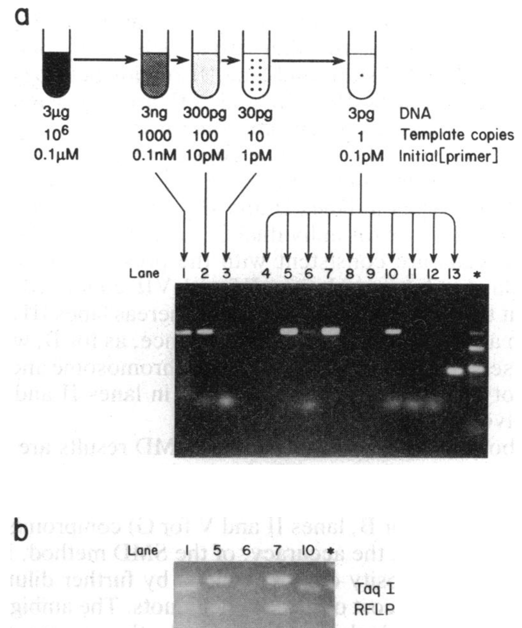

FIG. 3. Scheme for SMD of genomic DNA (from individual B in Fig. 2) followed by booster PCR of 770-bp target (*a*) and typing of amplified product for *Taq* I RFLP (*b*). At the single-molecule range (lanes 4–13), where a stochastic distribution describes the allocation of a specified DNA template among a series of aliquots, 4 of 10 samples amplified. RFLP analysis (*b*) demonstrates that while product amplified from 1000 template copies contains both alleles (lane 1), SMD products from this series contain one or the other. Photographs of ethidium bromide-stained gels with markers on lanes * (770, 500, and 250 bp) are shown.

Alleles at each of the three polymorphic loci were identified by typing DNA from SMD vials that contained product yields sufficient for visualization on ethidium bromide-stained gels (Fig. 4). Lane I is a control containing SMD product amplified from ≈1000 copies of genomic template. For both individuals B and G, lane I remains heterozygous at each of the three loci, as expected. Lanes II–VII contain DNA from six vials selected from two experiments for each individual (for B, these vials were from additional experiments distinct from that depicted in Fig. 3).

Individual B. Typing of ASO 1,3 showed that lanes IV, VI, and VII contain the 3 allele, lane V contains the 1 allele, and lane III retains heterozygosity. Lane II was ambiguous for this locus at this gel loading. Typing at the other two loci showed the same consistent pattern—lanes IV, VI, and VII

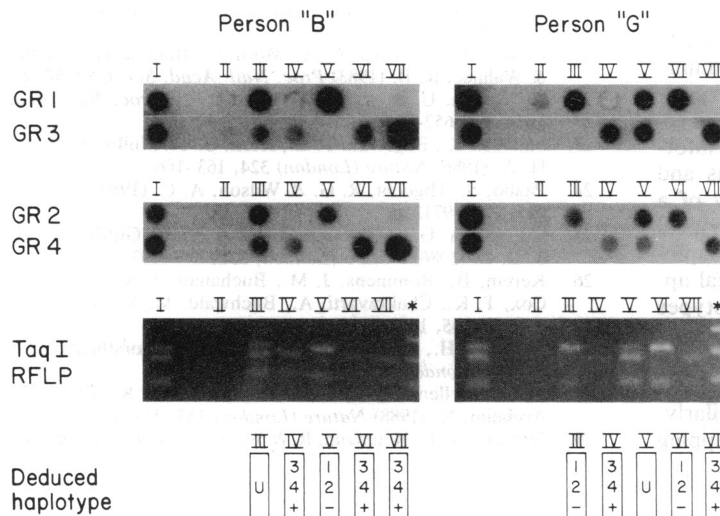

FIG. 4. Determination of haplotypes for heterozygotes B and G by SMD followed by booster PCR amplification and typing for three polymorphic loci. Persons B and G are heterozygous at all three loci (see Fig. 2). At 1000 copies of template (lanes I), products are heterozygous at each locus; for the SMD products (lanes II–VII), most samples show a single allele at each polymorphic locus, as would be expected theoretically. Since hemizygous lanes were amplified from a single DNA molecule, the haplotype can be derived directly. One chromosome of both B and G is 1,2,−, and the other is 3,4,+, exactly as deduced by classical analysis. (For B and G, lane II was ignored, as a lesser yield precluded unequivocal typing of product at photographic exposures and gel loadings appropriate to the other lanes.) Dot blots and RFLPs of amplified products are presented, respectively, as in Figs. 2 and 3.

contain the same allele (4 at ASO 2,4; + at the *Taq* I RFLP site), lane V contains the alternative allele (2 at ASO 2,4; − at the *Taq* I RFLP site), and lane III retains heterozygosity at both loci. Lanes IV, VI, and VII are consistent with individual B having one chromosome with 3,4,+ haplotype (the *T* haplotype), and lane V is consistent with individual B having one chromosome with 1,2,− haplotype (the *R* haplotype). Haplotypes in lanes II and III are unresolved.

Individual G. As for individual B, the typing results across loci and vials are consistent with the presence of just two particular haplotypes. Lanes IV and VII contained alleles 3,4,+ at the three loci, respectively, whereas lanes III and VI contain alleles 1,2,−, respectively. Hence, as for B, we infer the presence of the *T* haplotype in one chromosome and of the *R* haplotype in the other. Haplotypes in lanes II and V are unresolved.

For both individuals B and G, the SMD results are unambiguous and consistent with the haplotypes inferred from pedigree analysis (Fig. 2). The occurrence of unresolved vials (lanes II and III for B, lanes II and V for G) compromises the efficiency, but not the accuracy, of the SMD method. Retention of heterozygosity can be avoided by further dilution of template, but at a cost of more null aliquots. The ambiguity in lane II (both individuals) is experimental—the amount of DNA loaded was not sufficient to resolve the allele(s) at all three loci simultaneously. The latter could be overcome by retyping with an increased gel loading, although this is unnecessary since haplotypes are already resolved for both individuals.

DISCUSSION

SMD determines haplotypes nested between amplification primers in highly polymorphic regions as demonstrated here for the β-globin segment. A potentially important application would be direct sequencing (22) or typing (23) of HLA haplotypes from SMD products. For haplotype determination of sites separated by a distance exceeding the current limits of routine PCR ($\approx$3 kb), separate SMD amplifications of overlapping regions could allow logical reconstruction of the extended haplotype based on one allele shared at a heterozygous site within the overlap. However, overlap might not be required. If diluted DNA exists as relatively large fragments, distant segments in an intact template molecule are amenable to PCR with multiple primer pairs for direct haplotype determination. Even with DNA shearing, errors in haplotype determination should be held in check by analyzing replicate vials (13). Formation of chimeric products after homologous molecules primed from different chromosomes anneal to themselves in their central regions, the "jumping" artifact (24), would be minimized if not abolished by SMD, since a molecule will be isolated from its homologous jumping target by the extreme DNA dilution of these experiments.

Genome analysis utilizing SMD has particular relevance to evolution and medicine. Evolutionary studies of wild animal populations or anthropological isolates often depend on samples lacking organized families; SMD allows direct haplotyping of such field samples. Haplotypes have far more information content than classical protein polymorphisms and allow the reconstruction of the evolutionary history of a locus. In other biological contexts that require individual identification or knowledge of familial relationships, haplotypes serve as informative genetic markers. In medical applications, the increased information content of haplotypes makes them attractive as potential markers for improving the diagnosis of any linked disease genes or in epidemiological population screens for detecting specific haplotypes known to be associated with disease mutations (25, 26). Similarly, SMD haplotype determination applied to genetic mapping could increase the resolution of linkage analysis and facilitate crossover detection. While other approaches applicable to haplotype determination have amplified DNA from microdissected chromosomes (27) or spermatocytes (28), ours offers distinctive advantages. Direct haplotype determination with SMD and booster PCR not only obviates the need for pedigrees but is also technically simple and applies to DNA samples from either sex.

Note Added in Proof. While this manuscript was under review, Jeffreys *et al.* (29) independently applied single molecule dilution to the analysis of DNA sequence variation among repeat units at minisatellite loci.

We are indebted to Jeffrey Rogers for seminal discussions on SMD and to Wayne Fenton for collaboration on booster PCR. Probe CJ52 was a gift of Bernard G. Forget. This research was supported by National Science Foundation Grant BNS-8813234 and National Institutes of Health Grant GM-40883 awarded to K.K.K.

1. Lewontin, R. C., Moore, J. A., Provine, W. B. & Wallace, B., eds. (1981) *Dobzhansky's Genetics of Natural Populations* (Columbia Univ. Press, New York).
2. Lewontin, R. C. (1974) *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, New York).
3. Avise, J. C. & Lansman, R. A. (1983) in *Evolution of Genes and Proteins*, eds. Nei, M. & Koehn, R. K. (Sinauer, Sunderland, MA), pp. 147–164.
4. Kreitman, M. (1983) *Nature (London)* **304**, 412–417.
5. Woo, S. L. C. (1988) *Am. J. Hum. Genet.* **43**, 781–783.
6. Orkin, S. H. & Kazazian, H. H., Jr. (1984) *Annu. Rev. Genet.* **18**, 131–171.
7. Nozari, G., Rahbar, S. & Wallace, R. B. (1988) *Anal. Biochem.* **172**, 180–184.
8. Aquadro, C. F., Desse, S. F., Bland, M. M., Langley, C. H. & Laurie-Ahlberg, C. C. (1986) *Genetics* **114**, 1165–1190.
9. Schaeffer, S. W., Aquadro, C. F. & Anderson, W. W. (1987) *Mol. Biol. Evol.* **4**, 254–265.
10. Maeda, N., Bliska, J. B. & Smithies, O. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5012–5016.
11. Trucco, G., Fritsch, R., Giorda, R. & Trucco, M. (1989) *Diabetes* **38**, 1617–1622.
12. Ruano, G. & Kidd, K. K. (1989) *Nucleic Acids Res.* **17**, 8392.
13. Stephens, J. C., Rogers, J. & Ruano, G. (1990) *Am. J. Hum. Genet.* **46**, 1149–1155.
14. Saiki, R. K., Gelfand, D. H, Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
15. Ruano, G., Fenton, W. & Kidd, K. K. (1989) *Nucleic Acids Res.* **17**, 5407.
16. Maeda, N., Wu, C., Bliska, J. & Reneke, J. (1988) *Mol. Biol. Evol.* **5**, 1–20.
17. Mullis, K. B. & Faloona, F. A. (1987) *Methods Enzymol.* **155**, 335–350.
18. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
19. Kwok, S. & Higuchi, R. (1989 *Nature (London)* **339**, 237–238.
20. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980) *Am. J. Hum. Genet.* **32**, 314–331.
21. Conner, B. J., Reyes, A. A., Morn, C., Itakura, K., Teplitz, R. L. & Wallace, R. B. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 278–282.
22. Gyllensten, U. B. & Erlich, H. A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7652–7656.
23. Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1986) *Nature (London)* **324**, 163–166.
24. Pääbo, S., Higuchi, R. G. & Wilson, A. C. (1989) *J. Biol. Chem.* **264**, 9709–9712.
25. DiLella, A. G., Marvit, J., Lidsky, A. S., Güttler, F. & Woo, S. L. C. (1986) *Nature (London)* **322**, 799–803.
26. Kerem, B., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., Buchwald, M. & Tsui, L. C. (1989) *Science* **245**, 1073–1080.
27. Lüdecke, H., Senger, G., Claussen, U. & Horsthemke, B. (1989) *Nature (London)* **338**, 348–350.
28. Li, H., Gyllensten, U. B., Cui, X., Saiki, R. K., Erlich, H. A. & Arnheim, N. (1988) *Nature (London)* **335**, 414–417.
29. Jeffreys, A. J., Neumann, R. & Wilson, V. (1990) *Cell* **60**, 473–485.