

Theoretical Underpinning of the Single-Molecule-Dilution (SMD) Method of Direct Haplotype Resolution

J. Claiborne Stephens,*† Jeffrey Rogers,* and Gualberto Ruano*

*Department of Human Genetics, Yale University School of Medicine, and †Yale-Howard Hughes Medical Institute Human Gene Mapping Library, New Haven, CT

Summary

In a recent paper we have shown that DNA haplotypes of multiply heterozygous individuals can be resolved directly by polymerase-chain-reaction (PCR) amplification of a single molecule of genomic template. Our method (the single-molecule-dilution [SMD] method) relies on the stochastic separation of maternal and paternal alleles at high dilution. The stochasticity of separation and the potential for DNA shearing (which could separate the loci of interest) are two factors that can compromise the results of the experiment. This paper explores the consequences of these two factors and shows that the SMD method can be expected to work very reliably even in the presence of a moderate amount of DNA shearing.

Introduction

An important component of an organism's genotype is the linkage phase of alternative alleles at the variable sites. Traditionally, phase information has been obtained from pedigree analysis, although more recently cloning and sequencing of genomic DNA has produced haplotypes directly (e.g., see Maeda et al. 1983). In *Drosophila*, construction of isochromosomal lines has been a fruitful approach to the generation of restriction-map haplotypes (Kreitman and Aguadé 1986; Aquadro et al. 1986; Schaeffer et al. 1987). All of these options tend to be time-consuming and may not be available in general. Both a general inferential approach to resolving haplotypes from a sample of PCR-amplified DNA (Clark 1990) and a promising molecular approach dealing directly with hemizygous DNA (sperm) (Li et al. 1988) have been described recently.

We have recently described a simple three-phase approach to the direct determination of DNA haplotypes (Ruano et al., in press). The three phases in this single-molecule-dilution (SMD) method are (1) dilution of genomic DNA to an average of approximately one

haploid equivalent per PCR reaction vial, (2) amplification of replicate vials with "booster" PCR (Ruano et al. 1989a), and (3) typing of the alleles at each locus by standard techniques (e.g., RFLPs, allele-specific oligonucleotides (ASOs), and direct sequencing). Booster PCR has been described elsewhere (Ruano et al. 1989a), but briefly described, it is an efficient method for amplifying highly diluted DNA. It is a two-phase method that closely titrates PCR primer concentration to template concentration. Likewise, the haplotype-determination phase of the SMD method is well-understood technology and poses no special theoretical problems, so we may now turn to a consideration of the dilution phase of SMD.

Theoretical Framework without DNA Shearing

For concreteness, we will consider application of SMD to a study meeting the following criteria: (1) the study focuses on the high-resolution characterization of the genetic variation within a given genomic region within a population; (2) a certain amount of variation in that region is already known (e.g., restriction sites and ASOs); and (3) the known variation is characterized well enough that primers flanking the variable sites are available for PCR amplification. We will also make the important assumption that alleles at every heterozygous locus can be unambiguously and faithfully

Received December 5, 1989; revision received February 14, 1990.

Address for correspondence and reprints: J. Claiborne Stephens, Yale-Howard Hughes Medical Institute Human Gene Mapping Library, 25 Science Park, New Haven, CT 06511.

© 1990 by The American Society of Human Genetics. All rights reserved. 0002-9297/90/4606-0015\$02.00

resolved from each PCR vial. This assumption implies (1) that each PCR vial will be unambiguously null, hemizygous, or heterozygous for each locus and (2) that fidelity of amplification for all alleles at all loci is 100% whether the variable sites are contained within a single pair of PCR primers or within multiple pairs of PCR primers. We will discuss violation of both aspects of this assumption below.

Consider replicate PCR reactions using template DNA diluted to an average concentration of about one haploid equivalent of DNA per PCR vial (approximately 3 pg for humans). With the extraordinary sensitivity of PCR (Li et al. 1988; Saiki et al. 1988), it is reasonable to expect the PCR reaction to be “successful”—produce large amounts of amplified product—in at least one vial that contains only one of the two haplotypes. For example, a conservative estimate is that 37% of PCR vials will contain one (and only one) of the two haplotypes, when the *average* concentration is one haploid equivalent per PCR vial. This elementary probability estimate is calculated directly from the Poisson distribution without any correction for DNA shearing or aggregation. However, it is overly conservative in that it is calculated as the probability of one *and only one* copy of the relevant genomic region. Multiple copies in a vial could still lead to successful haplotype determination if all copies are of the same haplotype. Thus, the Poisson probability that a vial has one and only one copy given x copies per vial on average, $x \cdot e^{-x}$, can be augmented by a term corresponding to multiple identical copies in a vial. From the Poisson distribution, the probability of exactly i copies in a vial is $e^{-x} \cdot x^i / i!$. The probability of i copies in a vial all being identical is $1/2^i + 1/2^i = 2/2^i$ (*two* since there are two haplotypes in a heterozygous organism, assumed to be in equal frequency in the stock solution). Thus we can add to $x \cdot e^{-x}$ the probability of having multiple identical copies $\sum 2e^{-x} x^i / i! 2^i$ ($i = 2, 3, \dots$) = $2e^{-x} [e^{x/2} - x/2 - 1]$, to get $P(S) = 2e^{-x} [e^{x/2} - 1]$ as the probability of a PCR vial containing only one of the two haplotypes.

For enumeration of the possible statistical outcomes, we will define the event of one haplotype in a vial as a “success” and will define zero or both haplotypes in a vial as “unresolved.” Below we will define the event of an incorrect haplotype in a vial as an “error.” In the present framework, an error would be due to shearing or recombination during the dilution phase, since we are assuming unambiguous determination of alleles at each locus. At the level of the experiment, success is simply defined as having one or more of ν vials having

one but not both haplotypes. If the error rate is ignored for now, the probability of success of the experiment is simply $P(S_x) = 1 - [1 - P(S)]^\nu$, that is, one minus the probability of all ν vials being unresolved. Table 1 shows that the experiment has a high probability of success for five or more vials with one haploid equivalent of DNA on average. These probabilities are not reduced greatly if the average DNA content is between 0.5 and 2.0 haploid equivalents. Even one vial with one or two haploid equivalents has a probability of success of almost 50%. Again, without DNA shearing, and with our assumption of unambiguous determination of alleles after amplification, there is no possibility of making an error: each vial is either null, still heterozygous, or resolved into one of the two haplotypes. We therefore expect that, if DNA shearing is not too severe, the prospects for success are quite favorable.

DNA Shearing

To expand the above theoretical framework to include the effect of shearing, we will consider two heterozygous loci A and B , and we will treat the target DNA as if it is composed of two fractions—one that is intact between loci A and B (the unsheared fraction) and one in which loci A and B have been separated (the sheared fraction). Let x_{AB} be the average per-vial concentration of the unsheared genomic region. As above, we calculate the probabilities for the unsheared fraction as $P(0) = e^{-x_{AB}}$, the probability of zero unsheared fragments; $P(1) = 2e^{-x_{AB}}(e^{-x_{AB}/2} - 1)$, the probability of exactly one haplotype represented as one or more unsheared fragments; and $P(2) = 1 - e^{-x_{AB}}(2e^{-x_{AB}/2} - 1)$, the probability of both haplo-

Table 1

Estimates of Probability of Success

ν	DNA CONCENTRATION IN HAPLOID EQUIVALENTS				
	5	2	1	.5	.1
108047	.9981	.9985	.9854	.6223
87293	.9933	.9944	.9659	.5411
55581	.9562	.9610	.8790	.3855
11507	.4651	.4773	.3445	.0928

NOTE.—Probability of success per vial ($P(S)$, bottom row) is calculated as the probability that a vial contains a single haplotype (either as a single copy of template or as multiple copies of the same haplotype). The experiment is successful if at least one PCR reaction vial has one (and only one) haplotype initially, with probability $P(S_x) = 1 - [1 - P(S)]^\nu$.

types being present. We have now completely defined the per-vial possibilities for the unshared fraction of DNA.

Now, with regard to the sheared fraction, there are six basic possibilities (fig. 1)— $P(0;0)$, $P(0;1)$, $P(0;2)$, $P(1;1)$, $P(1;2)$, and $P(2;2)$ —representing, respectively, the probabilities that a vial contains (1) no sheared DNA containing either locus, (2) none of one locus and one allele from the other, (3) none of one and both alleles from the other, (4) one allele from each locus, (5) one allele from one locus and both alleles from the other, and (6) both alleles from each locus. So far, treatment of the sheared fraction is analogous to that of the unshared fraction, with the exception that we are interested in joint probabilities corresponding to both loci. Let x_A be the average per-vial concentration of fragments containing only locus A, and let x_B be the average per-vial concentration of fragments containing only locus B. If it is assumed that $x_A = x_B$, then the joint probabilities may be calculated in straightforward fashion from the product of terms of the probability distribution given for the unshared fraction, replacing x_{AB} by x_A . The probabilities of each outcome are shown in figure 1.

Unlike the case for the unshared fraction, shearing allows the possibility that, when only one allele from each locus is present (corresponding to probability $P(1;1)$), alleles could be present in nonparental (n) as

Haplotypes	
Parental	++ and --
Nonparental	+• and •+
Unshared Fraction	
$P(0) = e^{-x_{AB}}$	No unshared template
$P(1) = 2e^{-x_{AB}}(e^{-x_{AB}/2} - 1)$	<+•> _n or <•+> _n
$P(2) = 1 - e^{-x_{AB}}(2e^{-x_{AB}/2} - 1)$	<+•> _n and <•+> _n
Sheared Fraction	
$P(0;0) = e^{-2x_A}$	No sheared template
$P(0;1) = 4e^{-2x_A}(e^{-x_A/2} - 1)$	<+•> _n , <•+> _n , <+•> _n , or <•+> _n
$P(0;2) = 2e^{-x_A}[1 - e^{-x_A}(2e^{-x_A/2} - 1)]$	<+•, •+> _n or <•+, +•> _n
$P(1;1;p) = 2e^{-2x_A}(e^{-x_A/2} - 1)^2$	<+•, •+> _n or <•+, +•> _n
$P(1;1;n) = 2e^{-2x_A}(e^{-x_A/2} - 1)^2$	<+•, •+> _n or <•+, +•> _n
$P(1;2) = 4e^{-x_A}(e^{-x_A/2} - 1)[1 - e^{-x_A}(2e^{-x_A/2} - 1)]$	<+•> _n , <•+, •+> _n or <•+> _n , <+•, •+> _n or <+•, •+> _n , <•+, +•> _n or <•+, +•> _n , <+•, •+> _n
$P(2;2) = [1 - e^{-x_A}(2e^{-x_A/2} - 1)]^2$	<+•, •+> _n and <•+, +•> _n

Figure 1 Considerations for two loci and DNA shearing. For unshared DNA, only haplotypes <+•> and <•+> are possible, for sheared DNA, separate alleles are possible, shown as <+•> and <•+> at locus A and as <•+> and <+•> at locus B.

well as the parental (p) combinations. In such cases, parental and nonparental combinations should be equally frequent ($P(1;1;p) = P(1;1;n)$). Under our assumptions, it is the nonparental combinations that are our only source of error.

The possibilities for the sheared fraction have now been enumerated independently of those for the unshared fraction. The probability of a vial containing any combination of sheared and unshared fragments is then simply the product of sheared-fraction probabilities multiplied by those for the unshared fraction. Schematically, the per-vial outcomes are classified in table 2 as S = success, U = unresolved, and E = error. Note that an error can only result if neither unshared correct haplotype is present (probability $P(0)$) and if the nonparental configuration occurs for the unshared pair of loci present (probability $P(1;1;n)$). On the other hand, any vial retaining heterozygosity at one or both loci is unresolved (e.g., see last row and column).

Note that a vial that has one but not both intact haplotypes (center column) is a success unless it is contaminated by sheared material from the other haplotype, in which case it becomes unresolved. An allele from a sheared haplotype has a 50% chance of matching the allele in an unshared haplotype, which gives a single, correct haplotype. Likewise, there is a 50% chance of not matching, which would make the vial unresolved. Such vials are labeled S;U in table 2.

The probabilities of a vial being successful, erroneous, or unresolved are obtained by collecting the relevant cells of table 2 and are, respectively,

$$P(S) = 2e^{-x_{AB}}e^{-2x_A} [1 - 2e^{x_A/2} + e^{x_{AB}/2}e^{x_A}] ;$$

$$P(E) = 2e^{-x_{AB}}e^{-2x_A} [1 - e^{x_A/2}]^2 ;$$

$$P(U) = 1 - 2e^{-x_{AB}}e^{-2x_A} [2 + e^{x_A} - 4e^{x_A/2} + e^{x_{AB}/2}e^{x_A}] .$$

It is simple to show that the worst possible case (analogous to loci on different chromosomes, $x_{AB} = 0$) has

Table 2

Schematic Per-Vial Results When Shearing of DNA Is a Factor

SHEARED FRACTION	UNSHARED FRACTION		
	P(0)	P(1)	P(2)
$P(0;0)$	U	S	U
$P(0;1)$	U	S;U	U
$P(0;2)$	U	U	U
$P(1;1;p)$	S	S;U	U
$P(1;1;n)$	E	U	U
$P(1;2)$	U	U	U
$P(2;2)$	U	U	U

Table 3
Per-Vial Probability of Success When Shearing of DNA Is a Factor

x_A	x_{AB}						
	.0	.1	.5	1.0	1.4	2.0	2.5
.0	.0000	.0928	.3445	.4773	.5000	.4651	.4088
.1	.0043	.0878	.3144	.4335	.4535	.4214	.3703
.2	.0148	.0894	.2911	.3962	.4130	.3828	.3359
.5	.0594	.1100	.2450	.3113	.3179	.2901	.2528
1.0	.1139	.1372	.1958	.2175	.2120	.1865	.1598
1.4	.1250	.1360	.1608	.1637	.1541	.1316	.1111
1.5	.1242	.1331	.1522	.1522	.1422	.1206	.1014
2.0	.1082	.1104	.1122	.1044	.0943	.0776	.0642
2.5	.0836	.0832	.0790	.0699	.0616	.0495	.0404

a maximum per-vial error rate of .125 at $x_A = x_B = 2\ln 2 \approx 1.386$. On the other hand, the best possible success rate is .5 and occurs when there is no shearing ($x_A = x_B = 0$) and $x_{AB} = 2\ln 2$. Sample per-vial success rates for a range of concentrations are shown in table 3.

Experimental Considerations

For a routine DNA extraction protocol, we expect shearing to be primarily a function of the distance between loci, so that the farther apart two loci are, the more likely the occurrence of a break between them. The outcome of a given experiment is dependent on (1) the number of PCR vials used, (2) the concentration of DNA, and (3) the proportion of chromosomes that are broken between the two loci. The first two factors are reasonably easy to control, although the third is not. Recasting our equations in terms of these three factors, we have

$$P(S) = 2e^{-C(1+b)}[1 - 2e^{bC/2} + e^{C(1+b)/2}] ;$$

$$P(E) = 2e^{-C(1+b)}[e^{bC/2} - 1]^2 ;$$

$$P(U) = 1 - 2e^{-C(1+b)}[2 + e^{bC} - 4e^{bC/2} + e^{C(1+b)/2}]$$

as the probabilities of each outcome per vial, where C is the DNA concentration in haploid equivalents and b is the proportion of target chromosomes that are broken between the loci of interest. These equations follow from the identities $x_{AB} = (1-b)C$ and $x_A = x_B = bC$.

The corresponding probabilities that an experiment is successful or erroneous are $P(S_x) = [1 - P(E)]^\nu - [1 - P(U)]^\nu$ and $P(E_x) = [1 - P(S)]^\nu - [1 - P(U)]^\nu$, respectively, where ν is the number of PCR vials used in each experiment. Success is defined conservatively here, as

an experiment in which at least one parental haplotype is observed and in which no nonparental haplotypes are observed. An erroneous experiment is exactly the opposite, i.e., one in which at least one nonparental haplotype is observed and in which no parental haplotypes are observed. Tentatively, experiments in which both parental and nonparental haplotypes are observed are classified as unresolved. Although a researcher does not know, a priori, which haplotypes are parental and which are not, experiments will be unresolved if they produce pairs of haplotypes that together do not yield the original multilocus phenotype of the individual being studied.

The optimal number of vials for a given DNA concentration (C) and a given proportion of breakage (b) can be calculated. Figure 2 shows the number of vials needed to maximize the probability of a successful experiment $P(S_x)$ for a reasonable range of DNA concentrations. Clearly, $P(S_x)$ is maximized at about 10 vials for DNA concentrations in the range 0.6–2.6, and breakage less than 15%. The initial tendency of the optimum to decrease as breakage increases is due to our labeling as unresolved any experiment in which both parental and nonparental haplotypes are found. Note that, without breakage ($b = 0$), the optimal vial number is infinite, as there is no probability of an erroneous vial.

The values of $P(S_x)$ that correspond to the optimal vials numbers are shown in figure 3, along with values of $P(S_x)$ that correspond to 10-vial experiments. For a reasonable range of DNA concentrations, 10 vials is sufficient for a probability of success that is indistinguishable from its theoretical maximum value. Note that success is expected in nearly 95% of all experiments, even when DNA breakage occurs in 15% of the chromosomes.

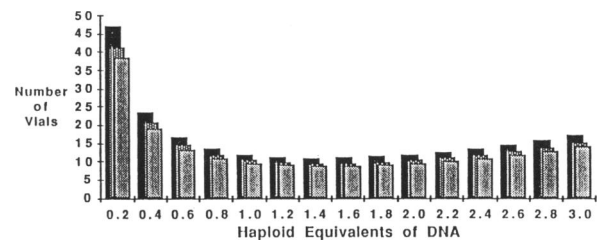


Figure 2 Number of vials needed to maximize $P(S_x)$ for DNA concentration in the range of 0.2–3.0 haploid equivalents and for breakage in the range of 5%–15% of target chromosomes. Black bar (■) = 5% broken; beaded bar (▨) = 10% broken; halftone bar (▩) = 15% broken.

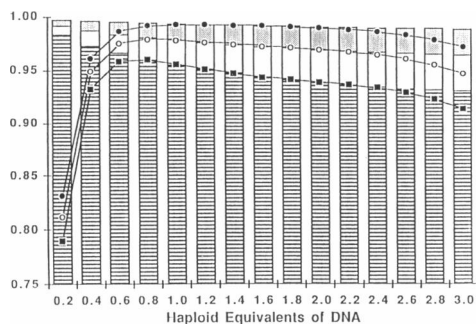


Figure 3 $P(S_x)$ for the optimal number of PCR vials (bars) and for 10 PCR vials (lines). DNA concentration and breakage are as in fig. 2. Half-tone bar (■) and —●— = 5% broken; open bar (□) and —○— = 10% broken; ladder bar (▨) and —■— = 15% broken.

Fortunately, the probability of an erroneous experiment is maximized for small numbers of PCR vials and drops off substantially for 10-vial experiments (fig. 4). For DNA concentrations up to two haploid equivalents and DNA breakage on the order of 15%, an erroneous inference would be made in less than 0.1% of experiments when 10 or more vials are used. In table 4, values of $P(E)$, $P(E_x)$, $P(S)$, and $P(S_x)$ are compiled for 10-vial experiments with $C = 1$ haploid equivalent of DNA. Clearly, even 40% breakage of chromosomes is not a major hindrance to the successful resolution of haplotypes by the SMD method. One could even improve on the proportion of successful experiments by using a rule such as that used by Boehnke et al. (1989) to stop the experiment when parentals outnumber non-parentals by a fixed number. The latter approach might be especially useful in situations where breakage reduces $P(S_x)$ below 50%, since $P(E_x)$ is appreciable in these situations.

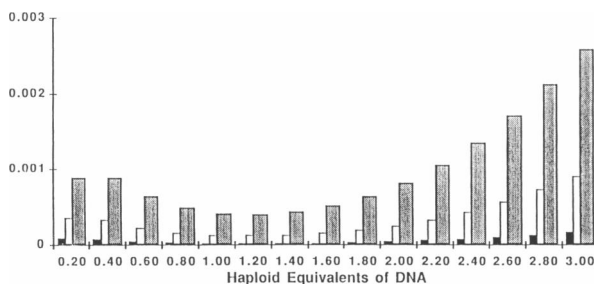


Figure 4 $P(E_x)$ for 10 PCR vials. DNA concentration and breakage are as in fig. 2. Black bar (■) = 5% broken; open bar (□) = 10% broken; half-tone bar (▨) = 15% broken.

Table 4

Expectations for 10-Vial Experiments with One Haploid Equivalent of DNA, under Differing Degrees of Chromosome Shearing

% Sheared $x_A (= x_B)$	% Unsheared x_{AB}	$P(E)$	10-Vial $P(E_x)$	$P(S)$	10-Vial $P(S_x)$
0	100	.0000	.0000	.4773	.9985
5	95	.0004	.0000	.4478	.9929
10	90	.0018	.0001	.4199	.9784
15	85	.0038	.0004	.3935	.9559
20	80	.0067	.0010	.3685	.9263
25	75	.0102	.0021	.3449	.8905
30	70	.0143	.0039	.3226	.8496
35	65	.0190	.0067	.3015	.8048
40	60	.0242	.0106	.2816	.7569
45	55	.0299	.0161	.2628	.7071
50	50	.0360	.0233	.2450	.6561
60	40	.0494	.0438	.2123	.5543
70	30	.0642	.0738	.1832	.4569
80	20	.0800	.1139	.1573	.3679
90	10	.0966	.1639	.1343	.2897
100	0	.1139	.2230	.1139	.2230

Discussion

We have applied the SMD method to a highly polymorphic region of the β -globin cluster and have shown that it reliably determined the haplotypes for a 770-bp region (Ruano et al., in press). This region was short enough to be amplified by a single pair of PCR primers. It will be of considerable interest to apply SMD to a region at least 10–100-fold larger, by using multiple pairs of PCR primers simultaneously. For such regions, we can expect DNA shearing to be more of a factor, although the considerations above allow us to be optimistic about the outcome of such experiments.

Other than DNA shearing, the major experimental difficulty would seem to be variability among the PCR primer pairs in their ability to amplify. This could happen for a wide variety of reasons. For instance, one pair of PCR primers may tend to “catch” the DNA template at an amplification cycle different from that for the other pairs of primers. Alternatively, experimental conditions may favor a particular pair of primers. Any such sensitivity difference would lead to variable yield and, in the extreme case, would appear as a partial haplotype, which could presumably be scored as unresolved and be attributed to DNA shearing. This potential difficulty would be expected to hurt the efficiency, but not the accuracy, of the SMD method. Contamination is another serious concern when amplifying highly dilute target sequences (Kwok and Higuchi 1989) and could

affect both the efficiency and the accuracy of experiments. Note that, in the theoretical treatment above, we have assumed that PCR faithfully replicates whatever template is in the PCR reaction vial.

Fortunately, recent work by Arnheim and his colleagues (Li et al. 1988; Arnheim 1989; Boehnke et al. 1989; Cui et al. 1989) on simultaneous amplification of multiple loci of single sperm indicates that differential amplification is only a minor problem and is primarily due to problems with sperm lysis. They now estimate that each locus has a probability of amplification of about 95% and that multiple amplifications are correlated (Arnheim 1989; Boehnke et al. 1989; Cui et al. 1989). The latter correlation means that a vial containing the two true haplotypes is unlikely to produce an incorrect haplotype by sporadic amplification of the alleles present. It should also be noted that our use of "booster" PCR (Ruano et al. 1989a, and in press) seems to enhance yield (2.2×10^{12} compared with 7.7×10^{10}) over that seen in the amplification of single sperm (Li et al. 1988).

Another potentially compromising factor is the fidelity of PCR amplification. Use of paired ASOs for each polymorphic site is the strongest safeguard against allele misidentification. However, unless there is a bias toward a specific error in the PCR amplification, replicate PCR vials would also be able to suggest that an error has occurred during PCR amplification.

Even with these reservations, it would appear that the ease and generality of the SMD method makes it a potentially powerful approach toward large-scale haplotype determination for relatively long tracts of DNA. In human genetics, there are already many regions of the genome where multiple polymorphic sites are known: HLA, β -globin cluster, retinoblastoma, phenylalanine hydroxylase, the apolipoprotein genes, and the cystic fibrosis locus, to name only a very few. Correct determination of the haplotypes that exist in a population is valuable as a way to increase the PIC (Botstein et al. 1980) at a locus, which in our case would now be the entire target region.

One type of study that can benefit immediately from the SMD approach is analysis of genetic differences among human populations. For example, Wainscoat et al. (1986) showed in a study of the β -globin region that the frequencies of various haplotypes differ substantially among populations. Such studies of human variation often depend on analysis of DNA obtained from blood samples collected in the course of fieldwork in remote areas. In such circumstances it is often not possible to collect entire pedigrees that would allow

reconstruction of haplotypes. Analyses of the geographic distribution of DNA polymorphism (e.g., see Bowcock et al. 1987) will be able to use the SMD approach to determine multisite haplotypes even when pedigrees are not available. Furthermore, molecular genetic studies of within- and between-species variation in nonhuman primates are now underway (Rogers 1989; Rogers and Kidd 1989; Ruano et al. 1989b). In such studies, determination of haplotypes for population genetic comparisons will usually be impossible without the SMD technique. This method will be useful in any study in which analysis of haplotypes is important but in which traditional approaches to haplotype reconstruction are not suitable.

Acknowledgments

We thank Dr. Ken Kidd for many useful suggestions and for his thorough reading of the manuscript.

References

- Aquadro CF, Dese SF, Bland MM, Langley CH, Laurie-Ahlberg CC (1986) Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. *Genetics* 114:1165-1190
- Arnheim N (1989) A new approach to constructing genetic maps: PCR analysis of DNA sequences in individual gametes. In: Erlich HA (ed) PCR technology. Stockton, New York, pp 119-135
- Boehnke M, Arnheim N, Li H, Collins FS (1989) Fine structure genetic mapping of human chromosomes using the polymerase chain reaction on single sperm: experimental design considerations. *Am J Hum Genet* 45:21-32
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331
- Bowcock AM, Bucci C, Hebert JM, Kidd JR, Kidd KK, Friedlaender JS, Cavalli-Sforza LL (1987) Study of 47 DNA markers in five populations from four continents. *Gene Geogr* 1:47-64
- Clark AG (1990) Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol Biol Evol* 7:111-122
- Cui X, Li H, Goradia TM, Lange K, Kazazian HH Jr, Galas D, Arnheim N (1989) Single sperm typing: determination of genetic distance between the γ -globin and parathyroid hormone loci by using the polymerase chain reaction and allele-specific oligomers. *Proc Natl Acad Sci USA* 86:9389-9393
- Kreitman M, Aguadé M (1986) Genetic uniformity in two populations of *Drosophila melanogaster* as revealed by filter hybridization of four-nucleotide-recognizing restriction enzyme digests. *Proc Natl Acad Sci USA* 83:3562-3566

- Kwok S, Higuchi R (1989) Avoiding false positives with PCR. *Nature* 339:237–238
- Li H, Gyllenstein UB, Cui X, Saiki RK, Erlich HA, Arnheim N (1988) Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 335:414–417
- Maeda N, Bliska JB, Smithies O (1983) Recombination and balanced chromosome polymorphism suggested by DNA sequences 5' to the human δ -globin gene. *Proc Natl Acad Sci USA* 80:5012–5016
- Rogers J (1989) Genetic structure and microevolution in a Tanzanian population of yellow baboons. Unpublished Ph.D. thesis, Yale University
- Rogers J, Kidd KK (1989) Population genetic structure in Tanzanian yellow baboons as measured by RFLPs in nuclear DNA. *Am J Phys Anthropol* 78:291–292
- Ruano G, Fenton W, Kidd KK (1989a) Biphasic amplification of very dilute DNA samples via “booster” PCR. *Nucleic Acids Res* 17:5407
- Ruano G, Gray MR, Ferguson-Smith A, Ruddle F, Kidd KK (1989b) Denaturing gradient electrophoresis (DGE) for detection of sequence differences in defined HOX2 segments amplified from human, gorilla and chimpanzee via Cross-species PCR (CS-PCR). *Am J Hum Genet* 45:A160
- Ruano G, Kidd KK, Stephens JC. Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules. *Proc Natl Acad Sci USA* (in press)
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487–491
- Schaeffer SW, Aquadro CF, Anderson WW (1987) Restriction-map variation in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. *Mol Biol Evol* 4:254–265
- Wainscoat JS, Hill AVS, Boyce AL, Flint J, Hernandez M, Thein SL, Old JM, et al (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. *Nature* 319:491–493