Am. J. Hum. Genet. 50:347–359, 1992

A Multiple-Tubes Approach for Accurate Genotyping of Very Small DNA Samples by Using PCR: Statistical Considerations

W. Navidi,* N. Arnheim,† and M. S. Waterman*,†

Departments of *Mathematics and †Molecular Biology, University of Southern California, Los Angeles

Summary

A multiple-tubes procedure is described for using PCR to determine the genotype of a very small DNA sample. The procedure involves dividing the sample among several tubes, then amplifying and typing the contents of each tube separately. The results are analyzed by a statistical procedure which determines whether a genotype can be conclusively assigned to the DNA sample. Simulation studies show that this procedure usually gives correct results even when the number of double-stranded fragments in the sample is as small as 30. The procedure remains effective even in the presence of small amounts of laboratory contamination. We find that the multiple-tubes procedure is superior to the standard one-tube procedure, either when the sample is small or when laboratory contamination is a potential problem; and we recommend its use in these situations. Because the procedure is statistical, it allows the degree of certainty in the result to be quantified and may be useful in other PCR applications as well.

Introduction

PCR (Saiki et al. 1985, 1988; Mullis and Faloona 1987) is a highly sensitive method that can be used to determine the genotype of a sample of DNA at a given locus. In addition to the rapidity with which results can be obtained, PCR has important applications in cases where the amount of DNA in a sample may be limiting and too small to be analyzed by any other means. This includes prenatal genetic disease diagnosis and forensic and archaeological or evolutionary studies (reviewed in Erlich 1989; von Beroldingen et al. 1989; White et al. 1989; Arnheim et al. 1990*b;* Innis et al. 1990; Erlich et al. 1991). In these cases anywhere from a nanogram to a microgram of genomic DNA (300–300,000 copies of a unique sequence gene) is usually suitable for analysis.

PCR is not an error-free technique, and considerable attention has been paid to problems of contamination (Kwok and Higuchi 1989). Contamination appears to be the major source of error for most applications of PCR when nucleic acid samples are in the nanogram-to-microgram range. Under some circumstances important DNA samples may contain even less than 300 copies of a gene. In such cases other types of errors can become significant. For example, assume that a very small sample of 10 double-stranded gene fragments is available from a heterozygous individual. In about 10% of such cases, 8 or more of the 10 fragments will by chance contain the same allele. In this event, if the 10 fragments are amplified to a detectable level by PCR, and if the PCR product is typed, it is likely that the signal from the less common allele, if detected at all, will be comparatively quite weak and be attributed to background contamination. Thus an incorrect finding of homozygosity will be made. On the other hand, suppose that the 10 fragments consist of 5 containing each allele. In this case, it is likely that PCR would lead to a correct typing result. The conclusion is that some small samples are likely to give correct results but that others are not.

In addition to this "sampling error," the possibility of contamination is of particular concern when the sample is small, because the ratio of contaminating fragments to sample fragments is less likely to be negligible. Methods designed to minimize the possibility of

Received June 6, 1991; revision received August 15, 1991.

Address for correspondence and reprints: Dr. William Navidi, Department of Mathematics, University of Southern California, Los Angeles, CA 90089.

0002-9297/92/5002-0010$02.00

contamination have been discussed in the literature (e.g., see Higuchi and Blake 1989; Kwok and Higuchi 1989), but rules for interpreting the results of PCR reactions have generally assumed that the procedure is contamination free or that the presence of contamination will invariably be detectable from the typing results or from negative (no DNA added) controls.

There may be circumstances where the investigator wants to determine the genotype of DNA samples which contain very small numbers of molecules (i.e., fewer than 300). Forensic studies are one example. Because of the errors inherent in typing small samples, the commercially available PCR kit (AmpliType; Cetus) for DNA typing at the HLA DQA locus has a built-in control which warns the investigator that the observed results have come from a small amount of target and might thus be unreliable. In addition to forensic applications, DNA typing by PCR will be able to provide important information from archaeological, museum, or fossil materials. However, standard methods of analyzing these samples depend on there being large enough numbers of target typing molecules for reliable results to be obtained.

It should be noted that some PCR experiments involving few initial targets do not suffer from the sampling-error problem. Examples include single-sperm, oocyte, or polar-body typing in recombination and preimplantation genetic disease analysis (see the review by Arnheim et al. [1990*a*]). This is because a single cell cannot contain two alleles in unequal proportion. Of course single-cell analysis is subject to other errors, and statistical approaches to single-cell analysis by PCR have been published (Boehnke et al. 1989; Cui et al. 1989; Goradia et al. 1991; Navidi and Arnheim 1991).

We suggest an approach which will, in many cases, allow the accurate typing of DNA even when the available sample is very small. The approach is to divide the sample DNA among several tubes, then amplify and type the contents of each tube separately. The typing of the PCR product may be done by any conventional procedure—for example, by using allele-specific probes (Saiki et al. 1986), restriction-enzyme digestion (Kogan et al. 1987), allele-specific PCR (see Ugozzoli and Wallace 1991), or OLA (Landegren et al. 1988). The number of PCR cycles used must be large enough (e.g., 50) so that even single molecules can be amplified to detectable levels (for technical details, see Li et al. 1991). The feasibility of obtaining this degree of amplification has been proved with statistical rigor in sperm-typing experiments (Li et al. 1988; Cui et al. 1989; Goradia et al. 1991). The typing results from the various tubes are then analyzed by statistical methods to determine which genotype, if any, can be conclusively assigned to the sample.

It turns out that this multiple-tubes approach works as well or better than the conventional one-tube procedure, under large-sample, nearly contamination-free conditions, and that it is clearly superior when the sample size is not large or when the assumption of a contamination-free procedure is not desirable. We show that under certain assumptions, which are spelled out below, the multiple-tubes procedure has the following advantages: (1) A determination of genotype can be made, and the certainty of the conclusion can be quantified. Of course, when the certainty is low, the results will be declared inconclusive. (2) When 30 or more fragments are in the original sample, the results are almost always conclusive as to genotype. (3) The procedure remains effective even in the presence of small amounts of contamination.

Contamination

With regard to the multiple-tubes procedure, contamination events can be divided into two types. The first type consists of contamination events which are likely to affect all tubes simultaneously. Events of this type include reagent contamination and contamination of the sample from other sources which occurs before the sample is divided among the reaction tubes. An instance of the latter, which can occur in forensic investigations, is contamination of sperm DNA with DNA from a rape victim's vaginal epithelial cells. In other instances, this type of contamination can just as easily come from an unknown source.

The multiple-tubes procedure does not offer specific protection against contamination which affects all tubes equally. We assume that standard remedies will be applied and will be effective in preventing errors from this type of contamination. For example, blank tubes should be processed as controls during the amplification and typing process, to check for reagent contamination. When the sample DNA is contaminated with DNA from another source, and if the contaminating DNA exists in a smaller quantity than does the DNA from the target individual, the expected result is a comparatively weak background signal from the contamination occurring in conjunction with a stronger signal from the target. We will assume that

this sort of contamination can be dealt with effectively by ignoring alleles whose signals are considerably weaker than the strongest signal (see below).

Contamination events of the second type are those which affect individual reaction tubes, and estimates of its occurrence that are based on large samples range from 0% to 7%. For example, contamination from a human source in the laboratory may enter a tube, or a small amount of material may be inadvertently transferred from one sample to another or from a completed PCR to a sample. In these cases, especially if the contamination is from a completed PCR, the signal from the contamination may be stronger than the signal from the target. It will be shown that dividing the sample among several tubes, as in the multiple-tubes procedure, greatly reduces the risk of error from contamination events of this type.

Modeling PCR and Contamination

We describe a mathematical model of the PCR amplification process and of the contamination process. Our analysis of the multiple-tubes approach is based on this model; however, simulation studies, described below, indicate that our results remain valid under moderate deviations from the model assumptions.

The model takes into account several factors which limit the efficiency of PCR. First, a fragment may completely fail to amplify—for example, by adhering to the side of the reaction tube. In addition, replication errors and other factors result in the increase in the number of fragments per cycle being less than the doubling which would result from a perfect process.

Since we are assuming that alleles whose signals are comparatively weak will be ignored, it is possible that an allele actually present in the target DNA may escape detection even if fragments containing it are amplified to what would seem to be a sufficiently great extent. If another allele is present in much greater quantity, the signal from the less common allele may be considerably weaker. This can happen either if the two alleles were present in greatly differing numbers before amplification, which is often the case when the sample is small, or if the fragments containing one allele happen to undergo greater amplification than do the fragments containing the other. This latter condition also tends to arise when the sample is very small. For example, if a tube contains two double-stranded fragments, one containing allele A and the other containing allele a, and if, on the first PCR cycle, both A strands replicate while neither a strand does, then there will be twice as many A fragments as a fragments. Further cycles will, on average, tend to preserve this imbalance.

We now describe in detail the assumptions which underlie our approach to the sampling-error and contamination problems:

1. Each gene fragment initially present in the tube has probability r of not being able ever to interact with the PCR reagents, independently of any other fragment. This may happen when, for example, a fragment adheres to the side of the reaction tube or is thermally degraded during an early cycle. The value of r need not be known to the experimenter.

Assumption 2 applies to fragments which do undergo some amplification.

2. During each PCR cycle, each fragment is completely replicated with probability p, independently of each other fragment. The value of p need not be known to the experimenter. In one experiment, starting with a single molecule, the value of p was estimated, over 50 cycles, to be .65 (Li et al. 1988). As will be discussed below, larger values of p have little effect on the accuracy of the procedure.

3. The frequency with which enzyme-misincorporation errors change one allele into another is negligible (see Boehnke et al. 1989).

4. If a tube contains two distinct alleles from the sample DNA, and if, after amplification, one allele yields more than m times as much product as does the other, then the signal from the less-amplified allele will be defined as comparatively so weak as to be attributed to background or contamination. Thus in such cases the less-amplified allele will not be detected. The number m is called the "limiting ratio." A plausible value for m is 3, but we also consider the value $m = 5$.

5. Each tube has probability c of being contaminated in an event of the second type described above, independently of each other tube. Analysis of 700 sperm suggests that in a carefully done procedure the value of c can range from 0% to 7% (Cui et al. 1989; Goradia et al. 1991).

In practice, c must be estimated—for example, by control studies. It is difficult to estimate c reliably. Fortunately, as the simulation studies below indicate, it

turns out that a rough estimate of c is adequate and that overestimating it has less effect than does underestimating it. In particular, our simulation studies show that an estimate of 10% will work well whenever the true c is between 0% and 10%.

6. All contaminating fragments contain an allele not found in the individual from which the sample DNA came.

Assumption 6 is conservative in that it tends to increase the difficulty caused by contamination.

7. Every contaminating allele yields a product amount which allows it to be detected, without interfering with the detection of other alleles.

Assumption 7 is an attempt to balance two possible contamination conditions. In one situation, there is much more sample than contaminant in the tube, so the contaminant may escape detection. This may happen, for example, when a tube is contaminated with a microdrop from another sample of roughly equal concentration. In the other situation, there is more contaminant than sample, so the contaminant may overwhelm the sample. This is likely to happen when a tube is contaminated by a microdrop of material from a completed PCR, where the concentration of product will be much greater than that in the sample.

Statistical Analysis of Multiple-Tube Data

The goal of the statistical analysis is to reduce the chance of error to an acceptably small level while producing a conclusive result as often as possible. To achieve this, we perform a series of hypothesis tests. For each possible genotype, we test the hypothesis that it is the genotype of the sample. To conclude that a particular genotype is the genotype of the sample, we must fail to reject it at some level α, and we must reject each other genotype at some level α'. If no genotype fits the above requirements, the data are regarded as inconclusive. In this setup, smaller values of α and α' decrease the probability of making an incorrect finding but increase the probability that the results will be dismissed as inconclusive. Simulation studies, to be discussed later, indicate that reasonable values for α and α' are .05 and .01, respectively.

The number of hypothesis tests to be performed depends on the number of distinct alleles detected in the sample. If only one allele is detected, we test two hypotheses: (1) that the sample is homozygous and (2) that the sample is heterozygous, with one allele escaping detection. If n distinct alleles are detected, where $n > 1$, then, based on these n alleles, there are n possible homozygous genotypes and $n(n-1)/2$ possible heterozygous genotypes, for a total of $n(n+1)/2$ genotypes to be tested in all.

We now describe the hypothesis tests. The test of the hypothesis that the DNA sample comes from a homozygous individual (for allele 1, say) is based on the fact that, if the hypothesis is true, then tubes in which alleles other than allele 1 are detected are contaminated tubes. The hypothesis will be rejected if the number of such tubes is significantly large. Let N be the total number of tubes used in the procedure. Since, by assumption 4, the number of contaminated tubes is binomially distributed with known parameters N and c, a one-sided binomial test is appropriate.

The test of the hypothesis that the sample DNA comes from a heterozygote (with alleles 1 and 2, say) is based in part on (a) the number n_1 of tubes in which allele 1 is detected but allele 2 is not and (b) the number n_2 of tubes in which allele 2 is detected but allele 1 is not. The following theorem gives the distribution of a test statistic.

THEOREM 1. Under assumptions 1–7, if the DNA sample comes from a heterozygote with alleles 1 and 2, then, conditional on the value of $n_1 + n_2$, the quantity n_1 is binomially distributed with parameters $n_1 + n_2$ and 1/2. (The proof of the theorem is in the Appendix).

Theorem 1 allows us to reject the hypothesis of heterozygosity if n_1 is significantly large or small, according to a two-sided binomial test. Another test of heterozygosity is also available, based on the fact that, when the hypothesis is true, the number of tubes containing alleles other than 1 and 2 is the number of contaminated tubes. The one-sided binomial test mentioned above can be used, rejecting the hypothesis when this number is significantly large. We compute p values for each of the above two tests and take the smaller one to be the overall p value for the hypothesis of heterozygosity. A more conservative approach would be to double the smaller p value, in accordance with the Bonferroni principle. However, simulations suggest that this approach is unnecessarily conservative.

We now present a hypothetical example to illustrate the method. In this example, 10 tubes are used. We estimate c to be 10%. We will take $\alpha = .05$ and $\alpha' = .01$. Of the 10 tubes, one is negative for all alleles, 3 are positive for allele 1 only, 2 are positive for allele

2 only, 3 are positive for alleles 1 and 2, and one is positive for alleles 1 and 3. It is fairly clear from a look at the data that the sample comes from an individual heterozygous for alleles 1 and 2. We will describe the hypothesis tests and show how they verify this conclusion. We have six hypotheses to test, three homozygous and three heterozygous.

To test the hypothesis that the individual is homozygous for allele 1, we note that there are six tubes containing an allele other than allele 1. We compute the probability that six or more tubes are contaminated. This is the probability that a binomial random variable with parameters 10 and .10 will have a value of 6 or more. This probability is about 1.47×10^{-4}, so we reject the hypothesis of homozygosity for 1. Similarly, we test the hypotheses of homozygosity for alleles 2 and 3 by determining the probability that the number of contaminated tubes is seven or more and nine or more, respectively. These probabilities are 9.12×10^{-6} and 9.10×10^{-9}, respectively, so these hypotheses are rejected as well.

To test the hypothesis that the sample is heterozygous for alleles 1 and 2, we first note that four tubes contain allele 1 but not allele 2, while two tubes contain allele 2 but not allele 1. We compute the two sided p value of an observation of four tubes, from a binomial distribution with parameters 6 (= 2 + 4) and 1/2. This p value is .688. Next we observe that one tube contains an allele other than 1 or 2. The probability that one or more of the tubes will be contaminated is .651. The p value for the hypothesis of heterozygosity for 1 and 2 is the smaller of .688 and .651, or .651. Similar calculations for the other two hypotheses of heterozygosity yield p values of 1.63×10^{-3} and 9.12×10^{-6}. Since the hypotheses of heterozygosity for alleles 1 and 2 are not rejected at the .05 level, while the other hypotheses are rejected at the .01 level, we find that the sample came from a heterozygous 1 and 2 individual.

We now present an example in which the conclusion is less obvious. Again we estimate c to be 10%. In this example, of the 10 tubes, 6 are positive for allele 1 only, and 4 are positive for alleles 1 and 2. When the hypothesis tests are performed as above, the p value for the hypothesis of homozygosity for allele 1 is .0128, the p value for the hypothesis of homozygosity for allele 2 is 1.00×10^{-10}, and the p value for the hypothesis of heterozygosity for alleles 1 and 2 is .0313. While the heterozygous hypothesis is preferred, the data do not strongly support any hypothesis. This is evidence that the procedure which produced the data did not conform to the model. The conclusion that one will reach will depend on one's degree of confidence that the procedure did, in fact, conform to the model. If one believes this strongly enough to use a value of $\alpha < .0313$, then one would make a finding that the data came from an individual heterozygous for alleles 1 and 2. Otherwise (say, if $\alpha = .05$), the data would be dismissed as inconclusive. In fact, these data were generated to be homozygous for allele 1, by a model which was correct except that c was .15 instead of .10. The low p value for the correct hypothesis is due to underestimating c.

A Sequential Approach

In many cases, it is possible to reduce the number of reactions performed by adopting a three-step sequential approach. In this approach, only a few of the tubes are amplified and typed at first. The data from these tubes are then analyzed, by using the hypothesis tests described above. If the results are conclusive, then the procedure stops. If the results are inconclusive, then several additional tubes are amplified. The data from these reactions are combined with the data from the first group of reactions and are analyzed. If the results are conclusive, then the procedure stops. Otherwise, all the remaining tubes are amplified, and the data from these reactions are combined with all the previous data and are analyzed.

The value of the sequential procedure is that it generally reduces the number of reactions needed. If this is not important, then equally accurate results can be obtained by amplifying all the tubes and analyzing all the data at once.

How Many Tubes Should Be Used?

We can determine appropriate numbers of tubes to use in each of the three stages of the sequential procedure described above, by calculating the minimum number of tubes needed to make a conclusive finding under ideal conditions. First assume that the sample is from a heterozygous individual and that all tubes show both alleles 1 and 2, say, so that the data provide the strongest possible evidence for the correct conclusion. There are three genotypes to be considered: homozygosity for allele 1, homozygosity for allele 2, and heterozygosity for alleles 1 and 2. A conclusive finding of heterozygosity for allele 1 and 2 will be made if both homozygous hypotheses are rejected at some level which we have been calling α'. For reasons

to be explained shortly, we will take this level to be .01/3. If the number of tubes used is denoted N, then the level at which the homozygous hypotheses are rejected is c^N. The minimum number of tubes is thus the smallest value of N for which $c^N \leq .01/3$. For example, if $c = .10$, then the minimum number of tubes is three. Now assume that the sample comes from a homozygous individual and that all tubes show only allele 1, so that, again, the data provide the strongest possible evidence for the correct conclusion. A finding of homozygosity will be made if a heterozygous hypothesis can be rejected at level $\alpha' = .01/3$. If the number of tubes used is denoted N, then the level at which any heterozygous hypothesis is rejected is $(.5)^{N-1}$. The minimum number of tubes is thus 10, because that is the smallest value of N for which $(.5)^{N-1} \leq .01/3$.

When the number of fragments in the initial sample is fairly large, and when c is not too high, it will often be the case that the data will approximate the ideal data described above. This suggests that in the first stage we use a number of tubes equal to or slightly larger than the minimum number of tubes necessary to make a conclusive finding of heterozygosity under ideal conditions; that is, it should be three or more when c is estimated to be 10%. The number of tubes used in the second stage should be enough to bring the total up to a number equal to or slightly larger than the minimum number of tubes necessary to make a conclusive finding of homozygosity under ideal conditions; that is, it should be 10 or more. Finally the total number of aliquots into which the sample is divided must be chosen. This number should be a few more than the total number of tubes used in the first two stages. In the sequential procedure described above, the hypothesis tests may be performed as many as three times. To ensure that the frequency with which an incorrect finding is made is no greater than α', each test should be conducted at level $\alpha'/3$. This is the reason for using the level .01/3 in the computations above.

Simulation Results

We now give the results of some simulations. In the simulations, we assumed that all the contaminating fragments contain the same allele. This is the most unfavorable contamination process consistent with the assumptions above. We divided our hypothetical samples into 15 aliquots, and we amplified 4 in the first stage, 6 in the second stage, and the remaining 5 in the third stage. We took $r = .1$ and $p = .6$ as values of the parameters describing the PCR amplification process. For the limiting ratio we took $m = 3$. This value indicates that whenever the PCR product in a tube contains two or more distinct alleles, and if the amount of product from one allele is more than three times as great as the amount of product from some other, the signal from the allele with the lesser amount of product will be disregarded. We also considered the situation in which $m = 5$. The true c is taken to be 5%, but we base our hypothesis tests on an estimated rate of 10%. Table 1 shows the results for various sample sizes when the sample is from a homozygous individual. For each choice of sample size, 10,000 data sets were generated and analyzed. With fewer than about 20 fragments, the data are almost always inconclusive. With 30 or more fragments in the sample, the procedure is usually conclusive. It is impossible to make a conclusive finding of homozygosity with four tubes. However, with 75 or more fragments in the sample, a conclusive finding is often made using 10 tubes. Of the 90,000 data sets generated in table 1, only 3 led to incorrect findings.

When the sample comes from a heterozygote, conclusive results can be obtained from even smaller samples, and, on average, fewer tubes are needed. Table 2 shows that conclusive results can be obtained from a sample of 20 fragments about 50% of the time, and from a sample of 30 fragments almost 90% of the time. Of the samples containing 200 fragments, more than half can be successfully analyzed by amplifying only four tubes. When 500 or more fragments are in the sample, four tubes are almost always enough. Of the 90,000 data sets generated in table 2, only 4 led to incorrect findings. The reason that fewer tubes are needed, on average, when the true genotype is heterozygous is that, roughly speaking, heterozygosity is easier to establish than homozygosity. If several alleles of each of two types are observed, then it is clear that the sample came from a heterozygous individual. It requires a greater number of observations of a single allele before one can be sure that no other allele will be seen.

Tables 3 and 4 show the effect of a higher value of c. In tables 3 and 4, the true c is 10%, and the hypothesis tests are based on an estimated c of 10% as well. Comparing tables 3 and 4 with tables 1 and 2, respectively, shows that, even though the true c is twice as much as before, the procedure gives conclusive results only slightly less often. In the homozygous case (table 3), the probability of an incorrect finding increases to

Table 1

Results of Multiple-Tubes Procedure: True Genotype Homozygous, True c = .05

No. of Fragments	Proportion of Samples Yielding Given Result						Average No. of Tubes
	Correct 4 Tubes	Correct 10 Tubes	Correct 15 Tubes	Total Correct	Incorrect	Inconclusive	
10	.0000	.0000	.0022	.0022	.0002	.9976	15.00
20	.0000	.0074	.5483	.5557	.0000	.4443	14.96
30	.0000	.0833	.7999	.8832	.0000	.1168	14.58
50	.0000	.3593	.6274	.9867	.0000	.0133	13.20
75	.0000	.5295	.4688	.9983	.0001	.0016	12.35
100	.0000	.5860	.4131	.9991	.0000	.0009	12.07
200	.0000	.5942	.4056	.9998	.0000	.0002	12.03
500	.0000	.5940	.4051	.9991	.0000	.0009	12.03

NOTE.—Results are of simulated typing experiments for various fragment sample sizes. For each size, 10,000 samples were generated. The values of the parameters are r = .1, p = .6, and m = 3. The estimated c is .10.

Table 2

Results of Multiple-Tubes Procedure: True Genotype Heterozygous, True c = .05

No. of Fragments	Proportion of Samples Yielding Given Result						Average No. of Tubes
	Correct 4 Tubes	Correct 10 Tubes	Correct 15 Tubes	Total Correct	Incorrect	Inconclusive	
10	.0000	.0002	.0000	.0002	.0000	.9998	15.00
20	.0002	.1371	.3363	.4736	.0002	.5262	14.31
30	.0036	.5350	.3578	.8964	.0002	.1034	12.29
50	.0345	.8740	.0800	.9885	.0000	.0115	10.25
75	.1128	.8700	.0132	.9960	.0000	.0040	9.41
100	.2136	.7829	.0027	.9992	.0000	.0008	8.74
200	.5437	.4557	.0005	.9999	.0000	.0001	6.74
500	.9395	.0601	.0002	.9998	.0000	.0002	4.37

NOTE.—Results are of simulated typing experiments for various fragment sample sizes. For each size, 10,000 samples were generated. The values of the parameters are r = .1, p = .6, and m = 3. The estimated c is .10.

Table 3

Results of Multiple-Tubes Procedure: True Genotype Homozygous, True c = .10

No. of Fragments	Proportion of Samples Yielding Given Result						Average No. of Tubes
	Correct 4 Tubes	Correct 10 Tubes	Correct 15 Tubes	Total Correct	Incorrect	Inconclusive	
10	.0000	.0000	.0011	.0011	.0024	.9965	15.00
30	.0000	.0459	.6896	.7355	.0026	.2619	14.76
50	.0000	.2109	.7347	.9456	.0023	.0521	13.94
100	.0000	.3353	.6513	.9866	.0017	.0117	13.32
500	.0000	.3531	.6344	.9875	.0019	.0106	13.23

NOTE.—Results are of simulated typing experiments for various fragment sample sizes. For each size, 10,000 samples were generated. The values of the parameters are r = .1, p = .6, and m = 3. The estimated c is .10.

Table 4

Results of Multiple-Tubes Procedure: True Genotype Heterozygous, True c = .10

No. of Fragments	Proportion of Samples Yielding Given Result						Average No. of Tubes
	Correct 4 Tubes	Correct 10 Tubes	Correct 15 Tubes	Total Correct	Incorrect	Inconclusive	
10	.0000	.0002	.0000	.0002	.0000	.9998	15.00
30	.0031	.5287	.3554	.8872	.0003	.1125	12.32
50	.0365	.8620	.0850	.9835	.0000	.0165	10.28
100	.2125	.7713	.0086	.9924	.0000	.0076	8.81
500	.9308	.0669	.0010	.9987	.0000	.0013	4.43

Note.—Results are of simulated typing experiments for various fragment sample sizes. For each size, 10,000 samples were generated. The values of the parameters are r = .1, p = .6, and m = 3. The estimated c is .10.

about 1 in 400 when the sample is very small and increases slightly less otherwise. In the heterozygous case (table 4), the probability of an incorrect finding is about the same as with a true c of 5%. Thus, although the multiple-tubes procedure, because of its added complexity, could increase c, this should be more than offset by the procedure's relative lack of sensitivity to contamination, as indicated in the above results.

We also did some simulations in which the true c was 15%, while the estimated c was 10%. In the homozygous case, the frequency of incorrect typings was about 1% to 2%. In the heterozygous case, the probability of an incorrect finding was less than 0.5%. To examine the effect of overestimating c, we did some simulations where the true c was 1% but the estimated c was 10%. The accuracy of the procedure was as good as or better than when the true c was 5%, so overestimating c seems to have little effect. We conclude that estimating c to be 10% will produce good results for true c between 0 and 10% but that it will be somewhat less reliable when the true value of c is higher. In practice, if it is suspected that the true c may be much greater than 10%, then a higher estimate should be used.

If the amount of contamination in a tube is much greater than the amount of sample DNA, then only the contamination will be detected during the typing procedure. This is particularly likely to happen when the contamination is from a completed PCR. Tables 5 and 6 give the results of a simulation where c is 5%, and only the contaminating allele is detected in those tubes which are contaminated. Comparing table 5 with table 1 shows that, in the homozygous case, the frequency with which the results are conclusive has declined somewhat, and the error rate has increased slightly. Comparing table 6 with table 2 shows that there is almost no difference in the heterozygous case.

Table 5

Results of Multiple-Tubes Procedure: True Genotype Homozygous, True c = .05

No. of Fragments	Proportion of Samples Yielding Given Result						Average No. of Tubes
	Correct 4 Tubes	Correct 10 Tubes	Correct 15 Tubes	Total Correct	Incorrect	Inconclusive	
10	.0000	.0000	.0027	.0027	.0000	.9973	15.00
30	.0000	.0803	.4722	.5525	.0002	.4473	14.60
50	.0000	.3631	.4234	.7865	.0000	.2135	13.18
100	.0000	.5896	.2588	.8484	.0002	.1514	12.05
500	.0000	.6057	.2422	.8479	.0003	.1518	11.97

Note.—Results are of simulated typing experiments for various fragment sample sizes. For each size, 10,000 samples were generated. The values of the parameters are r = .1, p = .6, and m = 3. The estimated c is .10. When a tube is contaminated, only the contamination is detected.

Table 6

Results of Multiple-Tubes Procedure: True Genotype Heterozygous, True c = .05

No. of Fragments	Proportion of Samples Yielding Given Result						Average No. of Tubes
	Correct 4 Tubes	Correct 10 Tubes	Correct 15 Tubes	Total Correct	Incorrect	Inconclusive	
10	.0000	.0001	.0000	.0001	.0000	.9999	15.00
30	.0026	.4495	.3836	.8357	.0000	.1643	12.72
50	.0257	.8180	.1344	.9781	.0001	.0218	10.63
100	.1755	.8092	.0123	.9970	.0000	.0030	9.02
500	.7664	.2327	.0004	.9995	.0000	.0005	5.40

Note.—Results are of simulated typing experiments for various fragment sample sizes. For each size, 10,000 samples were generated. The values of the parameters are r = .1, p = .6, and m = 3. The estimated c is .10. When a tube is contaminated, only the contamination is detected.

The procedure also remains reasonably effective under moderate violations of the assumption that the contamination process is independent across tubes, although the error rate does increase somewhat in the homozygous case. Tables 7 and 8 give the results of a simulation in which dependence was introduced into the contamination process as follows: The experiments were of two types, A and B. In type A experiments, all 15 tubes were contamination free. In type B experiments, each tube had a 20% chance of being contaminated, independently of each other tube. Each simulated experiment had probability 75% of being of type A and had probability 25% of being of type B. In this way, the overall proportion of tubes which are contaminated is 5%, but the contaminated tubes tend to appear in clusters to a much greater degree than would be the case under the independence assumption. In the homozygous case (table 7), the error rate is about 1.5% when the sample is very small and is less than 1% otherwise. In the heterozygous case (table 8), the error rate remains less than 0.5%. In both cases, the frequency of conclusive results is somewhat less than it is when the contamination events are independent.

Improving the efficiency of PCR does not noticeably increase the accuracy of the procedure. In a simulation study similar to that reported in tables 1 and 2, we used the same values of the parameters as before, except that r = .001 and p = .999, so that the amplification was nearly perfect. There was a slight decrease in the average number of tubes needed, and the frequency of incorrect results was about the same.

The One-Tube Procedure

For purposes of comparison with the multiple-tubes procedure, we now describe the results of a simulation study of the standard one-tube procedure, in which

Table 7

Results of Multiple-Tubes Procedure: True Genotype Homozygous, True c = .05, and Contamination Events Dependent

No. of Fragments	Proportion of Samples Yielding Given Result						Average No. of Tubes
	Correct 4 Tubes	Correct 10 Tubes	Correct 15 Tubes	Total Correct	Incorrect	Inconclusive	
10	.0000	.0000	.0033	.0033	.0174	.9793	14.98
30	.0000	.1105	.7407	.8512	.0126	.1362	14.41
50	.0000	.4707	.4684	.9391	.0091	.0518	12.60
100	.0000	.7629	.1966	.9595	.0073	.0332	11.15
500	.0000	.7796	.1775	.9571	.0075	.0354	11.06

Note.—Results are of simulated typing experiments for various fragment sample sizes. For each size, 10,000 samples were generated. The values of the parameters are r = .1, p = .6, and m = 3. The estimated c is .10.

Table 8

Results of Multiple-Tubes Procedure: True Genotype Heterozygous, True c = .05, Contamination Events Dependent

No. of Fragments	Proportion of Samples Yielding Given Result						Average No. of Tubes
	Correct 4 Tubes	Correct 10 Tubes	Correct 15 Tubes	Total Correct	Incorrect	Inconclusive	
10	.0000	.0000	.0000	.0000	.0021	.9979	15.00
30	.0032	.5198	.3466	.8696	.0017	.1287	12.36
50	.0355	.8452	.0861	.9668	.0002	.0330	10.38
100	.1995	.7717	.0085	.9797	.0000	.0203	8.95
500	.9329	.0612	.0012	.9953	.0000	.0047	4.43

NOTE.—Results are of simulated typing experiments for various fragment sample sizes. For each size, 10,000 samples were generated. The values of the parameters are r = .1, p = .6, and m = 3. The estimated c is .10.

the entire DNA sample is amplified in a single tube. In this procedure, results are considered conclusive if the number of alleles detected is either one or two. Table 9 gives the results of 10,000 simulated experiments where r = .1, p = .6, m = 3, and c = 5%. In the homozygous case, because of contamination assumptions 6 and 7, the frequency of incorrect results is equal to c, regardless of sample size. A more detailed model of the contamination process would probably result both in the error rate decreasing with sample size and in the occurrence of some inconclusive results. The results in table 9 for the homozygous case are probably fairly accurate, however, for situations in which contamination overwhelms the sample. In the heterozygous case, contamination is likely to result in three alleles being detected when the contaminating allele does not match either of the sample alleles, so the frequency of inconclusive results is approximately equal to c, regardless of sample size. Thus, for sample sizes of more than about 50 fragments, the frequency of inconclusive results is higher with the one-tube procedure than with the multiple-tubes procedure. In the heterozygous case, the frequency of incorrect results with the one-tube procedure is higher than with the multiple-tubes procedure, when the sample size is 50 fragments or less, and is about the same otherwise. Most of these errors occur in cases where, by chance, the sample consists of many more fragments containing one allele than the other. In the heterozygous case, comparing the one-tube procedure with the multiple-tubes procedure reveals a basic difference in the characteristics of the two methods. When the sample contains the two alleles in unequal amounts, results from the multiple-tubes procedure are usually inconclusive, while results from the one-tube procedure are often false. In the homozygous case, the multiple-tubes procedure is more accurate than the one-tube procedure, for all sample sizes. In general,

Table 9

Results of One-Tube Procedure

No. of Fragments	True Genotype Homozygous			True Genotype Heterozygous		
	Correct	Incorrect	Inconclusive	Correct	Incorrect	Inconclusive
10	.9504	.0496	.0000	.7673	.1938	.0389
30	.9491	.0509	.0000	.9326	.0165	.0509
50	.9486	.0514	.0000	.9480	.0021	.0499
100	.9487	.0513	.0000	.9519	.0001	.0450
500	.9485	.0515	.0000	.9482	.0000	.0518

NOTE.—Results are of simulated typing experiments for various fragment sample sizes. 10,000 samples for each size were generated. The values of the parameters are r = .1, p = .6, m = 3. The true c is .05.

the frequency of incorrect results is much lower with the multiple-tubes procedure, even when the model assumptions are violated to a moderate degree.

Varying the Limiting Ratio

Under our assumptions, an increase in m will result, in the heterozygous case, in an increase in the accuracy of both the one-tube and multiple-tubes procedures, since it is less likely that an allele actually present in the sample will be mistaken for contamination. The risk of contamination error does not increase, since we have assumed that contamination is always detected. The value of m does not matter in the homozygous case, since only one noncontaminating allele can be present.

In the heterozygous case, we ran some simulations for both the one-tube and the multiple-tubes procedures, with $m = 5$. The other parameter values were taken to be the same as in table 2 for the multiple-tubes procedure and the same as in table 9 for the one-tube procedure. There was virtually no difference in the performance of the multiple-tubes procedure when the sample size was either equal to 10 fragments or greater than 30 fragments. When the sample size was 20 or 30 fragments, the frequency with which the results were correct increased from 5% to 20% over the values in table 2, while the frequency of incorrect results remained about the same.

In the one-tube procedure the results were about the same as in table 9, when the sample size was 20 fragments or more. When the sample size was 10 fragments, the one-tube procedure was correct about 88% of the time, incorrect about 7.5% of the time, and inconclusive about 4.5% of the time. This represents an improvement over the results in table 9.

In summary, if m increases from 3 to 5, the advantage of the multiple-tubes procedure over the one-tube procedure is narrowed slightly for very small sample sizes of about 10 fragments or so and is increased slightly for samples of 20 to 30 fragments. For larger sample sizes, the advantage remains about the same.

Discussion

We have shown that the multiple-tubes procedure could offer reliable DNA typing in cases when the amount of available DNA is too small to be reliably typed with a conventional one-tube procedure. We have developed a statistical method for evaluating the significance of a typing result. The frequency of incorrect results is low whenever $c < 10\%$. The procedure remains effective even when there is the possibility that some tubes may have contamination great enough to overwhelm the sample. When the assumption of independence of contamination events is strongly violated, the procedure may be less reliable. When the number of gene fragments in the sample is 30 or more, conclusive results are usually obtained. When the sample contains two alleles in unequal amounts, the results are generally inconclusive rather than false.

The one-tube procedure is fairly accurate in the heterozygous case when the sample is about 50 fragments or more. When the sample is smaller, the one-tube procedure is subject both to errors due to contamination and to errors due to random fluctuations in the proportions of the two alleles in the sample. In the homozygous case, the one-tube procedure will be subject to errors even when the sample is fairly large, as long as contamination great enough to overwhelm the sample is a possibility. The multiple-tubes procedure is clearly superior to the one-tube procedure for typing small DNA samples, and it is also somewhat better when the number of fragments is large.

Since it is a statistical approach, the multiple-tubes procedure has the advantage of providing a quantitative measure of the degree of support for each possible genotype. It may also be useful to apply this kind of statistical approach to methods of genetic haplotyping of DNA samples by using single-molecule PCR (Ruano et al. 1990; Stephens et al. 1990).

Acknowledgments

We would like to express our appreciation to Henry Erlich for many helpful discussions. This research was partially supported by NSF grant DMS90–05833 and NIH grants GM3620, GM36745, and HG00328.

Appendix

Proof of Theorem 1: Let N be the number of tubes. The proof proceeds by induction on N. If $N = 1$, and if neither allele 1 nor allele 2 is detected, or if both are, then $n_1 = n_2 = 0$, so, conditional on $n_1 + n_2$, n_1 has the binomial distribution with zero trials. If exactly one of alleles 1 or 2 is detected, then $n_1 + n_2 = 1$, and, since the detected allele is just as likely to be 1 as 2, n_1 has the binomial distribution with parameters 1 and 1/2. Now let N be a number of tubes for which the theorem is true. We show that the theorem is true for $N + 1$. Let $t = 1$ if allele 1 is detected in the $N+1$st

tube while allele 2 is not. Let $t = 2$ if allele 2 is detected in the $N+1$st tube while allele 1 is not. Let $t = 0$ otherwise. Let n_1 be the number of tubes, among the $N + 1$, which contain allele 1 but not allele 2. Let n_2 be the number of tubes, among the $N + 1$, which contain allele 2 but not allele 1. Let n_1' be the number of tubes, among the first N, which contain allele 1 but not allele 2. Let n_2' be the number of tubes, among the first N, which contain allele 2 but not allele 1. Let $k = n_1 + n_2$. If $k = 0$, then, conditional on $n_1 + n_2$, n_1 has the binomial distribution with zero trials. If $k > 0$, let i be a nonnegative integer, $i \leq k$. Then

$$P(n_1 = i|n_1 + n_2 = k) = P(n_1 = i, t = 0|n_1 + n_2 = k) + P(n_1 = i, t = 1|n_1 + n_2 = k) + P(n_1 = i, t = 2|n_1 + n_2 = k) \quad [(A1)]$$

The three terms on the right-hand side of equation (A1) can be written as follows:

$$P(n_1 = i, t = 0|n_1 + n_2 = k) = P(t = 0|n_1 + n_2 = k) P(n_1' = i|t = 0, n_1' + n_2' = k), \quad [(A2)]$$

$$P(n_1 = i, t = 1|n_1 + n_2 = k) = P(t = 1|n_1 + n_2 = k) P(n_1' = i - 1|t = 1, n_1' + n_2' = k - 1), \quad [(A3)]$$

$$P(n_1 = i, t = 2|n_1 + n_2 = k) = P(t = 2|n_1 + n_2 = k) P(n_1' = i|t = 2, n_1' + n_2' = k - 1). \quad [(A4)]$$

Using the induction hypothesis, and the fact that $E(n_1|n_1 + n_2 = k) = k/2$, we obtain

$$P(n_1 = i, t = 0|n_1 + n_2 = k) = \frac{N+1-k}{N+1} \frac{k!}{i!(k-i)!} (1/2)^k, \quad [(A5)]$$

$$P(n_1 = i, t = 1|n_1 + n_2 = k) = \frac{k/2}{N+1} \frac{(k-1)!}{(i-1)!(k-i)!} (1/2)^{k-1}, \quad [(A6)]$$

$$P(n_1 = i, t = 2|n_1 + n_2 = k) = \frac{k/2}{N+1} \frac{(k-1)!}{i!(k-i-1)!} (1/2)^{k-1}. \quad [(A7)]$$

Substituting the expressions on the right-hand sides of equations (A5), (A6), and (A7) into the right-hand side of equation (A1) and performing some algebra yields

$$P(n_1 = i|n_1 + n_2 = k) = \frac{k!}{i!(k-i)!} (1/2)^k, \quad [(A8)]$$

which proves the theorem.

References

Arnheim N, Li F, Cui X (1990*a*) PCR analysis of DNA sequences in single cells: single sperm mapping and genetic disease diagnosis. Genomics 8:415–419

Arnheim N, White T, Rainey W (1990*b*) Application of the polymerase chain reaction to organismal and population biology. Bioscience 40:174–182

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

Cui X, Li H, Goradia TM, Lange K, Kazazian HH, Galas DJ, Arnheim N (1989) Single sperm typing: determination of genetic distance between the $^{G}\gamma$ globin and parathyroid hormone loci. Proc Natl Acad Sci USA 86:9389–9393

Erlich H (ed) (1989) PCR technology: principles and applications for DNA amplification. Stockton, New York

Erlich H, Gelfand D, Sninsky J (1991) Recent advances in the polymerase chain reaction. Science 252:1643–1651

Goradia TM, Stanton VP, Cui X, Aburatani H, Li H, Lange K, Housman DE, et al (1991) Ordering three DNA polymorphisms on human chromosome 3 by sperm typing. Genomics 10:748–755

Higuchi R, Blake E (1989) Applications of the polymerase chain reaction in forensic science. Banbury report 32: DNA technology and forensic science. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Innis MA, Geland DH, Sninsky JJ, White TJ (eds) (1990) PCR protocols: a guide to methods and applications. Academic Press, New York

Kogan SC, Doherty M, Gitschier J (1987) An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. N Engl J Med 317:985–990

Kwok S, Higuchi R (1989) Avoiding false positives with PCR. Nature 339:237–238

Landegren U, Kaiser R, Sanders J, Hood L (1988) A ligase-mediated gene detection technique. Science 241:1077–1080

Li H, Gyllensten U, Cui X, Saiki R, Erlich H, Arnheim N (1988) Amplification and analysis of DNA sequences in single human sperm. Nature 335:414–417

Li H, Cui X, Arnheim N (1991) Analysis of DNA sequence variation in single cells. In: Ableson JN, Simon MI (eds) methods: a companion to *Methods in Enzymology,* vol 2 (Arnheim N [ed]). Academic Press, San Diego, pp 49–59

Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. Methods Enzymol 155:335–351

Navidi W, Arnheim N (1991) Using PCR in preimplantation genetic disease diagnosis. Hum Reprod 6:836–849

Ruano G, Kidd K, Stephens J (1990) Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules. Proc Natl Acad Sci USA 87:6296–6300

Saiki R, Bugawan TL, Horn GT, Mullis KB, Erlich HA (1986) Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. Nature 324:163–166

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, et al (1988) Primer directed amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Saiki R, Scharf S, Faloona F, Mullis K, Horn G, Erlich H, Arnheim N (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350–1354

Stephens JC, Rogers J, Ruano G (1990) Theoretical underpinning of the single-molecule-dilution (SMD) method of direct haplotype resolution. Am J Hum Genet 46:1149–1155

Ugozzoli L, Wallace RB (1991) Allele-specific polymerase chain reaction. In: Ableson JN, Simon MI (eds) methods—a companion to *Methods in Enzymology,* vol 2 (Arnheim N [ed]). Academic Press, San Diego, pp 42–48

von Beroldingen CH, Blake ET, Higuchi R, Sensabaugh GF, Erlich HA (1989) Application of PCR to the analysis of biological evidence. In: Erlich H (ed) PCR technology: principles and applications for DNA amplification. Stockton, New York pp 209–223

White T, Arnheim N, Erlich H (1989) The polymerase chain reaction. Trends Genet 5:185–189