

## Genetic differences at four DNA typing loci in Finnish, Italian, and mixed Caucasian populations

(DNA typing/human population genetics/population substructure/ceiling principle/variable number of tandem repeats)

DAN E. KRANE\*, ROBERT W. ALLEN†, STANLEY A. SAWYER\*‡, DMITRI A. PETROV\*, AND DANIEL L. HARTL\*

\*Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110; †Chapman Institute of Medical Genetics, Children's Medical Center, 5300 East Skelly Drive, Tulsa, OK 74135; and ‡Department of Mathematics, Washington University, St. Louis, MO 63130

Communicated by Bruce Wallace, August 10, 1992

**ABSTRACT** Highly polymorphic segments of the human genome containing variable numbers of tandem repeats (VNTRs) have been widely used to establish DNA profiles of individuals for use in forensics. Methods of estimating the probability of occurrence of matching DNA profiles between two randomly selected individuals have been subject to extensive debate regarding the possibility of significant substructure occurring within the major races. We have sampled two Caucasian subpopulations, Finns and Italians, at four commonly used VNTR loci to determine the extent to which the subgroups differ from each other and from a mixed Caucasian database. The data were also analyzed for the occurrence of linkage disequilibrium among the loci. The allele frequency distributions of some loci were found to differ significantly among the subpopulations in a manner consistent with population substructure. Major differences were also found in the probability of occurrence of matching DNA profiles between two individuals chosen at random from the same subpopulation. With respect to the Finnish and Italian subpopulations, the conventional product rule for estimating the probability of a multilocus VNTR match using a mixed Caucasian database consistently yields estimates that are artificially small. Systematic errors of this type were not found using the interim ceiling principle recently advocated in the National Research Council's report [National Research Council (1992) *DNA Technology in Forensic Science* (Natl. Acad. Sci., Washington)]. The interim ceiling principle is based on currently available racial or ethnic databases and sets an arbitrary lower limit on each VNTR allele frequency. In the future the ceiling frequencies are expected to be established from more adequate data acquired for relevant VNTR loci from multiple subpopulations.

The discovery of hypervariable loci within the human genome is potentially one of the greatest contributions to forensic science in this century. Each human genome has been estimated to be heterozygous at  $3 \times 10^6$  nucleotide sites (1). Consequently, barring monozygotic multiple births, it is theoretically possible to uniquely identify individuals only from their DNA. Especially useful in DNA typing are highly variable multilocus probes (2, 3) and highly polymorphic single-locus variable number of tandem repeat (VNTR) probes (4). In both cases DNA types are determined by hybridization of the probes with genomic restriction fragments that have been separated by electrophoresis and transferred to suitable membranes.

The conventional method of estimating the probability of a VNTR match between two randomly chosen individuals has been based on the multiplication of the frequencies of the relevant VNTR alleles occurring in a database of a particular racial group (5). We will call this conventional method the "product rule." The product rule implicitly assumes that the

underlying database is drawn from a single randomly mating population with statistical independence of alleles within and between loci. It also assumes that the allele frequencies in the database are adequate estimators of the true allele frequencies in the sampled population. The validity of the conventional product rule has been challenged because analysis of blood group and enzyme-coding loci have given evidence of substructuring within the human races (6). Although data regarding the extent of ethnic variation at VNTR loci is still limited (7), the occurrence of significant differences in allele frequency would invalidate the product rule. The reason is that, if subpopulations differ in the frequencies of individual alleles or multilocus combinations of alleles, then the frequencies of the alleles are not statistically independent but correlated with ethnicity. In such a case, use of the product rule with an ethnically mixed database would usually be prejudicial to a defendant because it would consistently yield estimates of match probabilities that are artificially small (6). Especially for VNTRs, errors incurred through use of the product rule can be substantial due to the large number of alleles and the low frequencies at which many alleles occur.

Several alternatives to the product rule have been put forward. One alternative would be to estimate the frequency of occurrence of a given DNA profile as the reciprocal of the number of times the DNA profile actually occurs in a particular reference database (8). If the database consisted of  $N$  individuals, none of which exhibited the particular DNA profile in question, the estimated probability of a match would be taken as  $1/N$ . Multilocus matches in mixed databases are indeed uncommon. For example, only one three-locus VNTR match was found among  $7.6 \times 10^6$  pairwise comparisons in a racially mixed database of effective size 3900 individuals (9). By using these databases, if no match were observed with a particular three-locus VNTR profile, then the  $1/N$  method would yield an estimated probability of  $1/3900$  or  $2.6 \times 10^{-4}$ .

Another alternative to the product rule is the ceiling principle (8, 10), which modifies the conventional product rule so that each allele frequency used in the calculation is taken as the larger of (i) the maximum frequency observed for that allele in a sample of at least 100 persons from each of 15–20 relevant subpopulations and (ii) 0.05. For most genotypes, this method is likely to be conservative in the sense that the estimated match probability is generally expected to be greater than the true match probability within any particular ethnic group. In addition, the ceiling principle product rule will generally yield estimates of the match probabilities that are substantially smaller than the  $1/N$  method.

A recent report of the National Research Council, *DNA Technology in Forensic Science* (8), has endorsed the use of the ceiling principle for calculating match probabilities in forensic DNA typing. However, at present, relevant data

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: VNTR, variable number of tandem repeats; HWE, Hardy-Weinberg equilibrium.

from a sufficient number of ethnic groups is inadequate to apply the ceiling principle literally. As an interim method, the report of the National Research Council (8) suggests using a modified form of the ceiling principle in which the frequency of any allele used in the calculation is the larger of (i) the upper 95% confidence limit on the maximum frequency observed among at least three ethnically or racially distinct databases and (ii) 0.10. We will refer to this alternative as the interim ceiling principle.

Based on the experience with blood group and enzyme polymorphisms (11, 12), it might be expected that the level of subpopulation differentiation would differ among VNTR loci, and preliminary studies with VNTRs support this suggestion (13, 14). In the present paper, allele frequencies at four VNTR loci are reported in two European ethnic groups, Finns and Italians, and are compared with an ethnically heterogeneous ("mixed") Caucasian population from St. Louis, MO. The VNTR loci are *D2S44*, *D10S28*, *D16S85*, and *HRAS1*. Within the three databases, only one locus showed a marked departure from the Hardy-Weinberg equilibrium (HWE) (*HRAS1* in the mixed Caucasian database) and another locus showed a slight excess of single-banded phenotypes (*D16S85* in the mixed Caucasian database). Only one slight departure from pairwise linkage equilibrium (gametic phase balance) was observed (*D2S44* versus *D16S85* in the Italian database). However, the data indicate that allele frequencies at some VNTR loci can differ by nearly 4-fold between different ethnic subpopulations, which has important implications for the reliability of match probabilities calculated from the conventional product rule. Consequently, match probabilities for three-locus DNA profiles determined with the product rule using ethnically mixed databases often exaggerate the rarity of a match. The present data also allow, to our knowledge, the first practical comparison of the interim ceiling principle with the conventional product rule for estimating probabilities of random matches between individuals from particular subpopulations.

## MATERIALS AND METHODS

**Sources and Extraction of Chromosomal DNA.** DNA samples from residents of the St. Louis area were obtained as part of the parentage testing service offered by the St. Louis American Red Cross. The mixed database includes mothers and alleged fathers from paternity cases who categorized themselves as "Caucasian." The sample size was  $N = 1354$  individuals. Chromosomal DNA was extracted from peripheral blood samples using standard procedures (15).

High molecular weight genomic DNA from randomly chosen Italian and Finnish blood donors was obtained from dried blood stains on cotton cloth, which were kindly provided by Francesco Bertolini (Centro Transfusionale e di Immunologia dei Trapianti, Milan, Italy) and Gunnar Myllylä (Finnish Red Cross Blood Transfusion Service, Helsinki), respectively. The sample sizes were  $N = 79$  Italians and  $N = 73$  Finns. Chromosomal DNA was extracted from a portion of each stain by first mincing it into 1-mm squares. The minced material was then hydrated with extraction buffer [0.1 M Tris-HCl, pH 8.0/0.1 M NaCl/10 mM EDTA/2% (wt/vol) SDS/10 mM dithiothreitol/proteinase K (100  $\mu$ g/ml)] and incubated overnight at 37°C. The supernatant phase of each sample was then extracted once with phenol/chloroform, 9:1 (vol/vol), and once with chloroform/isoamyl alcohol, 24:1 (vol/vol). DNA was recovered from the aqueous phase by ethanol precipitation and centrifugation for 10 min at 12,000  $\times g$ .

**Restriction Enzyme Digestion.** Chromosomal DNA samples (0.5–2.0  $\mu$ g in 200  $\mu$ l) were digested to completion with 10–20 units of *Pvu* II restriction endonuclease (United States Biochemical). After digestion the sample was adjusted to 0.2 M

NaCl, and the DNA fragments were precipitated with 2.5 vol of 95% ethanol. The digested DNA was recovered by centrifugation at 12,000  $\times g$  for 10 min, washed once with 70% ethanol, resuspended in electrophoresis buffer, and subjected to electrophoresis in 15  $\times$  20 cm agarose gels (0.8%) equilibrated in TAE (40 mM Tris acetate/2 mM EDTA, pH 8.3).

**DNA Blot Analysis.** After electrophoresis, gels were blotted onto nylon membranes (Biodyne B; BRL) for 6 h at room temperature in 0.4 M NaOH according to recommendations of the supplier. Blots were briefly rinsed in 2 $\times$  SSC (1 $\times$  SSC is 0.15 M NaCl/15 mM sodium citrate, pH 7.0) containing 0.1% SDS, baked in a vacuum oven for 2–4 h at 80°C, washed for 30 min at 65°C in 0.1 $\times$  SSC containing 0.1% SDS, and stored at 4°C between hybridizations.

**Hybridization and Washes.** Blots were hybridized to DNA probes randomly labeled with [<sup>32</sup>P]dCTP (16). The following four probes were used: YNH24 (locus *D2S44*) (4), TBQ7 (locus *D10S28*) (4), 3'HVR (locus *D16S85*) (17), and pHRAS (locus *HRAS1*) (18). Hybridizations were performed at 65°C essentially as described in Church and Gilbert (19) with radiolabeled probe at 1  $\times$  10<sup>6</sup> cpm/ml ( $\approx$ 1 ng/ml). The blots were washed for two 10-min periods in 2 $\times$  SSC containing 0.1% SDS at room temperature and for two 30-min washes at 65°C in 0.1 $\times$  SSC containing 0.1% SDS. Washed blots were exposed to x-ray film (Kodak XAR, Sigma) with intensifying screens at –80°C for 2–5 days.

**Data Acquisition.** Estimates of band sizes in a DNA profile were made using a sonic digitizer (Compugene, St. Louis). Molecular weight standards used for size estimates included a mixture of *Hind*III and *Bst*EII fragments of  $\lambda$  DNA that had been end-labeled with [<sup>32</sup>P]dCTP by using the Klenow fragment of DNA polymerase I. Each autoradiogram was digitized twice, and estimated band sizes in the readings were averaged. Reproducibility of band size estimates was typically  $\pm 0.6\%$  for duplicate readings of a given sample. Included on each gel was a sample of reference DNA composed of digested genomic DNA from a pool of three individuals. The reproducibility of estimated band sizes for the reference DNA sample present on different gels averaged approximately  $\pm 1.0\%$ .

**Statistical Analyses.** In principle, each VNTR allele corresponds to a different size *Pvu* II fragment, but in practice fragments that are too similar in size cannot be distinguished. Accordingly, we define the "allele" corresponding to any specified fragment size as including any genomic DNA sequence yielding fragments within a  $\pm 2.5\%$  window around the specified size. The "allele frequency" is determined by counting the number of chromosomes yielding DNA fragments within the  $\pm 2.5\%$  window. The  $\pm 2.5\%$  window was determined empirically from the actual reproducibility of the typing system used by the St. Louis Red Cross, and it conforms to recommended quantitative standards for forensic measurements (8).

Three tests for HWE frequencies and linkage equilibrium (also called gametic-phase balance) were carried out as follows: (i) examination of the intraclass correlation coefficient for the alleles within individuals, using the bootstrap to estimate significance levels (20), (ii) checking for an excess of single-band VNTR phenotypes, using contingency-table tests, and (iii) determination of whether randomly chosen pairs of alleles occur independently in genotypes (20). These tests are "omnibus" tests of averages across all alleles, and hence nonsignificance of the omnibus test is no guarantee of HWE or linkage equilibrium for any particular allele or pairs of alleles. To compare the distributions of fragment sizes in the Finnish, Italian, and mixed Caucasian databases, two statistical tests were performed on the frequency distributions (i) comparison using a two-sided Kolmogorov-Smirnov test (21) and (ii) comparison using a block test. In the block

tests, for each pairwise comparison, the range of fragment sizes was first divided into 10 regions ("blocks") of equal length. Adjacent blocks were then combined until all had a minimum of eight entries total, with at least three entries in each cell. Then  $\chi^2$  tests for homogeneity among the blocks were carried out.

**Match Probabilities.** Estimates of the probability of matching DNA profiles between two randomly chosen individuals were determined by application of the conventional product rule as well as by the interim ceiling principle. In the product rule, for a single locus, the probability of a genotype that is apparently homozygous for any specified allele is estimated as  $2p_i$ , where  $p_i$  is the frequency of the specified allele (i.e., all chromosomes yielding fragments within a  $\pm 2.5\%$  window) in the reference database. The frequency  $2p_i$  is used instead of  $p_i^2$  to correct for the fact that some VNTR loci exhibit a significant excess of single-band phenotypes (22) (for example, in the present study, the loci *D16S85* and *HRAS1* in mixed Caucasians). The probability of a genotype that is apparently heterozygous for any specified pair of alleles is estimated as  $2p_i p_j$ , where  $p_i$  and  $p_j$  are the estimated frequencies of the specified pair of alleles in the reference database. The overall multilocus probability is estimated as the product of the estimated frequencies for each individual locus. The interim ceiling principle was applied essentially as recommended by the National Research Council (8): homozygous genotype frequencies were estimated as  $2p_i$  and heterozygous genotype frequencies were estimated as  $2p_i p_j$ , using, for each allele, the larger of (i) the greatest value of the 95% upper confidence limit of the frequency of the allele in the Finnish, Italian, or mixed Caucasian databases or (ii) 0.10.

## RESULTS

**HWE.** Adherence to the expectations of the HWE at VNTR loci is often cited as evidence of the absence of population substructure (23, 24). This inference is generally invalid because the conventional statistical test for the HWE based on the frequency of homozygotes is low in statistical power (6). On the other hand, departures from the HWE can sometimes be taken as evidence that substructuring does exist in a population (25). Tests for the HWE were performed on each locus in the Finnish, Italian, and mixed Caucasian databases examined in this study. For each locus tested in all databases, the only exceptions to HWE frequencies were *HRAS1* and *D16S85* in the mixed Caucasian database. The locus *HRAS1* had a large within-locus intraclass correlation coefficient ( $r$ ) for fragment lengths ( $r = 0.26$ ;  $P < 0.01$ ) and also had a highly significant excess of single-banded phenotypes ( $P < 0.001$ ). The locus *D16S85* also had a slight excess of single-banded phenotypes ( $P = 0.02$ ) but did not deviate significantly from the HWE in the other statistical tests. While the findings with regard to an excess of single-banded phenotypes are statistically significant, we do not regard them as necessarily indicating an excess of homozygotes in the general population, since such departures can be caused by technical artifacts like DNA bands of low molecular weight migrating off the gels during electrophoresis (22).

**Linkage Equilibrium.** To determine whether the four VNTR loci studied give evidence of lack of statistical independence across loci, tests for linkage equilibrium were also performed. Only one departure from linkage equilibrium, of borderline statistical significance (intraclass correlation  $r = 0.08$ ;  $P = 0.03$ ), was found for any pairwise comparisons of loci (*D2S44* versus *D16S85* in the Italian database).

**Allele Frequency Distributions.** It has been noted that population substructure is generally easier to detect by comparing allele frequencies directly than by tests of the HWE or linkage equilibrium in mixed databases (6). Hence, population substructure between the Finnish and Italian

ethnic groups in this study might be reflected in differences in allele frequencies at the VNTR loci. The distribution of DNA fragment sizes for the loci *D10S28* (Fig. 1 A–C) and *D2S44* (Fig. 1 D–F) in the Finnish, Italian, and mixed Caucasian databases is shown. For all four loci, at least one difference between the databases was statistically significant in block tests or two-sided Kolmogorov–Smirnov tests or both (Table 1). The hypothesis that the Finnish and Italian databases have the same allele frequency distributions can be rejected for *D10S28* and *HRAS1*, even though the sample sizes are not large.

In addition to the overall significant differences among all four VNTR loci, many notable differences in the frequency of specific alleles at some loci between populations were also found. Most pronounced were those at the *D10S28* locus (Fig. 1 A–C), where the frequency of particular alleles differed almost 4-fold (e.g., the 5.43-kilobase allele, with a frequency 0.134 in the Finnish database and 0.034 in the Italian database; and the 3.94-kilobase allele, at 0.037 in the Finnish database and 0.137 in the Italian database).

Differences in allele frequency distributions among subpopulations imply that individuals chosen at random from the same subpopulation are more likely to have matching genotypes at one or more loci than individuals chosen from different subpopulations. This effect can be seen in the present databases by comparing the number of three-locus VNTR matches (using the  $\pm 2.5\%$  match window for each band) for the three most polymorphic loci in this study (*D2S44*, *D10S28*, and *D16S85*). A total of 1349 mixed Caucasians, 70 Italians, and 29 Finns were tested for all three loci. Among the 909,226 possible pairwise comparisons in the mixed Caucasian database, there were two matches at all three loci, for an overall frequency of  $2.2 \times 10^{-6}$ . There were also two matches at all three loci among the 2415 possible pairwise comparisons among Italians, for an overall probability of  $8.3 \times 10^{-4}$  (i.e., 376-fold greater than the match probability among mixed Caucasians). No matches of all three loci were found among the Finnish sample, but in this case only 406 pairwise comparisons were possible.

**Estimating Match Probabilities.** Population subdivision means that match probabilities calculated from the product rule using a mixed database (or a database from an incorrect ethnic group) will usually yield estimates that are artificially small (6). To assess this effect quantitatively, match probabilities were generated from the conventional product rule using the allele frequencies of *D2S44*, *D10S28*, and *D16S85* in

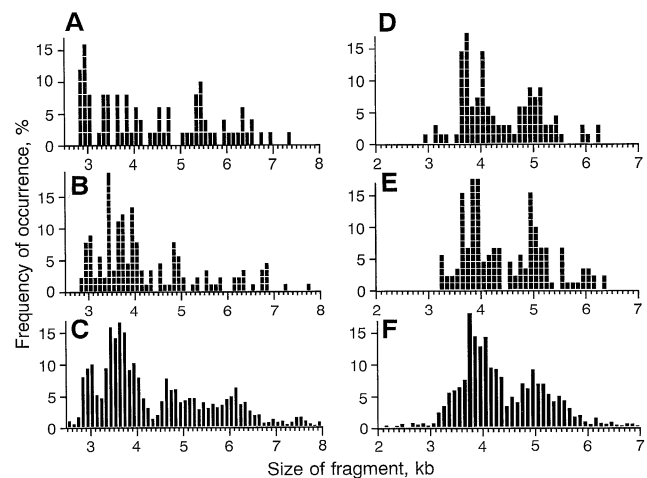


FIG. 1. Distribution of estimated DNA fragment sizes among the Finnish (A and D), Italian (B and E), and mixed Caucasian (C and F) databases for the loci *D10S28* (A–C) and *D2S44* (D–F). kb, Kilobase(s).

Table 1. Statistical significance (*P* values) of comparisons of VNTR allele distributions

Locus	Finns versus Italians	Finns versus mixed Caucasians	Italians versus mixed Caucasians
<i>D2S44</i>	—	—	0.01 <sup>†</sup>
<i>D10S28</i>	0.02* (2 × 8)	0.02* (2 × 5)	—
<i>D16S85</i>	—	0.01* (2 × 6)	0.05 <sup>†</sup>
<i>HRAS1</i>	0.01 <sup>†</sup>	<10 <sup>-8†</sup>	<10 <sup>-6†</sup>

\*Block test. The numbers in parentheses are the number of rows and columns in the array.

<sup>†</sup>Two-sided Kolmogorov–Smirnov test (21).

each of the three databases. The results are illustrated in Fig. 2. Fig. 2A shows the match probabilities for individuals matched against their own ethnic database (“cognate database”) as compared with the same individuals matched against a different ethnic database (i.e., Finns against Italian database or Italians against Finnish database). In using the inappropriate ethnic database, 77% of the match probabilities are artificially small; 34% of these are off by a factor of >10, and 4% are off by a factor of >100. Use of the mixed Caucasian database (Fig. 2B) is an improvement in terms of the magnitude of the discrepancy, but 80% of the estimated match probabilities are still nonconservative (i.e., too small), and 22% of the latter are off by a factor of >10.

Fig. 2C compares the match probabilities estimated from the interim ceiling principle (8), using all three databases,

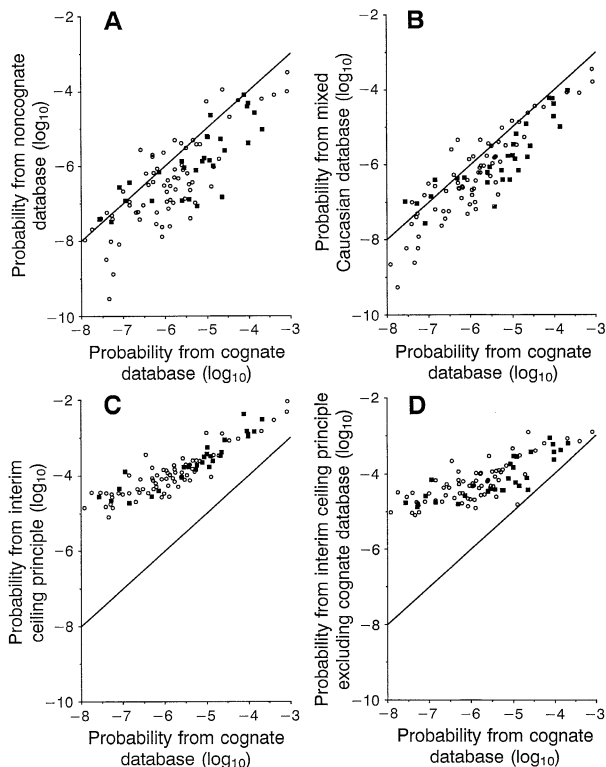


FIG. 2. Comparison of estimated match probabilities using the conventional product rule and the interim ceiling principle. Lines indicate “correct” probability if the cognate database is assumed to be appropriate. ■, Finns; ○, Italians. (A) Product rule estimates using the cognate database (Finnish database for Finns and Italian database for Italians) against estimates using the noncognate database (Italian database for Finns and Finnish database for Italians). (B) Product rule in cognate database against mixed Caucasian database. (C) Product rule in cognate database against interim ceiling principle using the Finnish, Italian, and mixed Caucasian databases. (D) Product rule in cognate database against interim ceiling principle with cognate database excluded.

with those estimated from the product rule using the cognate database. The estimates from the interim ceiling principle are conservative, usually by a factor of 10 or more. This margin of safety would seem to be justified by the results in Fig. 2D, in which the interim ceiling principle has been applied after elimination of the cognate database. Estimates without the cognate database are less conservative, particularly when match probabilities in the cognate database are  $>5 \times 10^{-6}$ .

## DISCUSSION

Among the four VNTR loci (*D2S44*, *D10S28*, *D16S85*, and *HRAS1*) studied in Finns, Italians, and mixed Caucasians, with few exceptions, the genotypes occurred in frequencies consistent with the assumption of the HWE and pairwise linkage equilibrium among the alleles. However, for every locus, the distributions of allele frequencies differed significantly in at least one comparison among the databases, and there were many marked differences in the frequencies of individual alleles. These results support the contention that conventional statistical tests for the HWE and linkage equilibrium are generally weak in their power to detect population substructure (6, 25). As demonstrated here, a more reliable way to detect substructure is through direct comparison of the subpopulations themselves. Furthermore, the degree of genetic differentiation among subpopulations is expected to differ among loci. Although general considerations based on blood groups and enzyme-coding genes suggest that the comparisons in Fig. 1 should be typical (6), other VNTR loci and other European ethnic groups may reveal some exceptions.

In our application of the interim ceiling principle, we used the Finnish, Italian, and mixed Caucasian databases. In agreement with the National Research Council (8), we would not endorse the use of ethnically mixed racial databases (e.g., mixed Caucasians, mixed Blacks, and mixed Hispanics) except as an interim measure. Reservations are justified on at least two grounds (6): (i) each of the major races consists of heterogeneous subpopulations within itself, and it may often be the case that the differences between average allele frequencies across races are smaller than the differences among subpopulations within races; and (ii) some racial designations are mere census terms that are virtually meaningless from a biological standpoint (e.g., the census term “Hispanic” includes people of variable amounts of Spanish, native American Indian, and African ancestry). On the other hand, with respect to the Finnish and Italian data analyzed here, the interim ceiling principle has a sufficient margin of safety (Fig. 2C) that it remains conservative even when the cognate database is not included (Fig. 2D). Additional data will be needed to ascertain whether this feature of the method remains valid with other VNTRs and other ethnic databases. We must also emphasize that there is an important practical limitation in taking, as the standard of comparison, the result of the conventional product rule as applied to the cognate database. In particular, this procedure need not yield a “correct” (i.e., unbiased) estimate of the match probability, even when applied to the cognate database, because (among other reasons) ethnic groups defined on the basis of geography or national boundaries can still harbor substantial genetic variation even among their own subpopulations (26).

The ceiling principle is based on the multiplication of estimates of allele frequencies and, hence, the validity of an estimated match probability could still be affected by the extent to which allele frequencies are precisely independent. Since the human population consists of multiple subpopulations of finite size, rather than a single random-mating unit of infinite size, then the HWE and linkage equilibrium cannot be expected to hold precisely (although many departures from independence will be too small to be statistically significant

in databases of 200–1200 individuals). There are also theoretical reasons why different VNTR loci may not be perfectly independent even in an infinite homogeneous population. Individuals that already match at one or more VNTR loci will, on the average, be more likely to share remote common ancestors than randomly chosen individuals and, hence, they are more likely to match at additional VNTR loci. Although the effect is not large, this situation also violates the assumption of independence. As we have shown in this report, the interim ceiling principle is a significant improvement over the conventional product rule for estimating the probability of matching DNA profiles. However, there is no theory, that we are aware of, that guarantees that estimated match probabilities will always be conservative (biased in favor of the defendant) when there are small departures from the HWE or multilocus linkage equilibrium. Although these considerations may adversely affect estimates when the match probabilities are extremely small, they are not likely to seriously compromise most estimates derived from the interim ceiling principle.

Apart from the issue of estimating match probabilities, our finding of multiple three-locus matches in the mixed Caucasian and Italian databases contrasts with the result of Risch and Devlin (9). These authors report no three-locus matches within any racial group in a Federal Bureau of Investigation database and only 1 three-locus match among  $7.6 \times 10^6$  pairwise comparisons among all their databases. If we assume a three-locus match rate of 1 per 454,500 pairs (our data for mixed Caucasians), Risch and Devlin (9) might have been expected to find 16 or 17 such matches instead of only 1 match. However, their tests involved a different set of VNTR loci than ours (only the locus *D2S44* is in common), and there is also some question whether some matches may have been erroneously eliminated from the Federal Bureau of Investigation database (27). In any event, our three-locus matches are real. Among the Italians, both pairs of matching individuals differed from each other at the *HRAS1* locus. Among the two pairs of matching mixed Caucasians, the *HRAS1* type is available for only one pair, and it differed between the individuals; for the other pair, the individuals appeared for paternity testing more than a year apart, their names and addresses are different, as are their thumbprints and facial features as judged from photographs present in the original files. Hence, it seems very unlikely that the DNA samples are derived from the same individual. On the other hand, even if we exclude the one matching pair of Caucasians on which *HRAS1* data are unavailable, our observed proportion of three-locus matches is still an order of magnitude greater than that reported by Risch and Devlin (9). Considering that three-locus matches were also observed in the Italian database, the discrepancy is even greater.

We are grateful to Beverly Hoover and Addean Pearson for technical assistance.

1. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980) *Am. J. Hum. Genet.* **32**, 314–331.
2. Jeffreys, A. J., Wilson, V. & Thein, S. L. (1985) *Nature (London)* **316**, 76–79.
3. Jeffreys, A. J., Wilson, V. & Thein, S. L. (1985) *Nature (London)* **314**, 67–73.
4. Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. & White, R. (1987) *Science* **235**, 1616–1622.
5. Budowle, B., Giusti, A. M., Wayne, J. S., Baechtel, F. S., Fournay, R. M., Adams, D. E., Presley, L. A., Deadman, H. A. & Monson, K. L. (1991) *Am. J. Hum. Genet.* **48**, 841–855.
6. Lewontin, R. C. & Hartl, D. L. (1991) *Science* **254**, 1745–1750.
7. Lander, E. S. (1991) *Am. J. Hum. Genet.* **48**, 819–823.
8. National Research Council (1992) *DNA Technology in Forensic Science* (Natl. Acad. Sci., Washington).
9. Risch, N. J. & Devlin, B. (1992) *Science* **255**, 717–720.
10. Lander, E. S. (1991) *Am. J. Hum. Genet.* **49**, 899–903.
11. Lewontin, R. C. (1972) *Evol. Biol.* **6**, 381–398.
12. Mourant, A. E., Kopec, A. C. & Domaniewska-Sobczak, K. (1976) *The Distribution of Human Blood Groups and Other Polymorphisms* (Oxford Univ. Press, New York).
13. Balazs, I., Neuweiler, J., Gunn, P., Kidd, J., Kidd, K. K., Kuhl, J. & Mingjun, L. (1992) *Genetics* **131**, 191–198.
14. Sajantila, A., Budowle, B., Ström, M., Johnsson, V., Lukka, M., Peltonen, L. & Ehnholm, C. (1992) *Am. J. Hum. Genet.* **50**, 816–825.
15. Miller, S. A., Dykes, D. D. & Polesky, H. F. (1988) *Nucleic Acids Res.* **16**, 1215.
16. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
17. Tarman, A. P., Nicholls, R. D., Wetherall, D. J., Clegg, J. B. & Higgs, D. R. (1986) *EMBO J.* **5**, 1857–1863.
18. Baird, M., Balazs, I., Giusti, A., Miyazaki, L., Nicholas, L., Wexler, K., Kanter, E., Glassberg, J., Allen, F., Rubinstein, P. & Sussman, L. (1986) *Am. J. Hum. Genet.* **39**, 489–501.
19. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
20. Weir, B. S. (1992) *Genetics* **130**, 873–887.
21. Hollander, M. & Wolfe, D. A. (1973) *Nonparametric Statistical Methods* (Wiley, New York), 219–227.
22. Odelberg, S. J., Plaetke, R., Eldridge, J. R., Ballard, L., O'Connell, P., Nakamura, Y., Leppert, M., Lalouel, J.-M. & White, R. (1989) *Genomics* **5**, 915–914.
23. Deka, R., Chakraborty, R. & Ferrell, R. E. (1991) *Genomics* **11**, 83–92.
24. Chakraborty, R. & Daiger, S. P. (1991) *Hum. Biol.* **63**, 571–588.
25. Hartl, D. L. & Clark, A. G. (1989) *Principles of Population Genetics* (Sinauer, Sunderland, MA).
26. Lewontin, R. C. & Hartl, D. L. (1992) *Science* **255**, 1054–1055.
27. Sullivan, P. J. (1992) *Science* **256**, 1743–1744.