NISS

Conditional Genotypic Probabilities for Microsatellite Loci

Jinko Graham, James Curran, and Bruce Weir

Technical Report Number 101
August, 1999

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

Conditional Genotypic Probabilities for Microsatellite Loci

Jinko Graham[*,†], James Curran[†], and Bruce Weir[†,*]

[*]National Institute of Statistical Sciences, and
[†]Program in Statistical Genetics, Department of Statistics, North Carolina State University .

Running head: Conditional genotypic probabilities

Keywords: forensic DNA profiling,
microsatellite markers,
coancestry,
product-rule,
Dirichlet distribution.

Address for correspondence:
Dr. Bruce Weir,
Program in Statistical Genetics
Department of Statistics
North Carolina State University, Box 8203
Raleigh, NC 27695-8203.

Phone: (919)-515-3574
Fax: (919)-515-7315
Email: weir@stat.ncsu.edu

ABSTRACT

In forensic science, DNA profiles are used to characterize individuals and associate suspects with crimes. Modern DNA profiles are constructed using microsatellite markers, short tandem repeats of sequences of 2-5 base pairs. One tool for evaluating evidence based on DNA profiles is the match probability. The match probability is the conditional probability that a random person would have the profile of interest given that the suspect has it and that these people are different members of the same subpopulation. One issue in evaluating the match probability is hidden population differentiation, which can induce coancestry among members of the same subpopulation. Forensic assessments of DNA profiles which ignore such coancestry typically overstate the strength of evidence against the suspect. Theory has been developed to account for coancestry when evaluating match probabilites. Assumptions include a steady-state population of constant size, and a recurrent mutation model that may be unrealistic for microsatellite markers. We use a coalescent to investigate the robustness of the theory to a generalized stepwise model of microsatellite mutation, in growing rather than constant-sized populations. Demographic parameters are chosen to reflect historical human population estimates. Simulations demonstrate that, even when coancestry is taken into account, the probability of a matching profile can still be underestimated for common genotypes.

INTRODUCTION

Several authors (e.g. BALDING and NICHOLS 1994; BALDING and NICHOLS 1995; LANGE 1995) have discussed the need to account for the coancestry of individuals when assessing the evidential strength of matching DNA profiles in forensic identification. Matching profiles could reflect genetic homogeneity of a subpopulation, rather than guilt of the suspect. Hence, a fair assessment of DNA profile evidence should allow for the possibility that the suspect and perpetrator belong to the same subpopulation (WEIR 1994). Standard product-rule estimators of match probabilities (NRC 1992) assume that the effects of hidden population subdivision are negligible. However, BALDING and NICHOLS (1997) examined genetic correlations quantifying population differentiation among Caucasians, and concluded that coancestry was too large to be ignored. They found that product-rule estimators of match probabilities can, in many cases, overstate the strength of evidence against the suspect.

Theory has been developed to account for the effects of coancestry on match probabilities. A mutation model is assumed in which the allelic state after a mutation

event is independent of the state prior to mutation (WRIGHT 1951; GRIFFITHS 1979). Under this source-invariant mutation model, the joint allele probabilities within a population may then be expressed in terms of marginal allele probabilities and identity-by-descent measures appropriate to the genetic model. Joint allele probabilities determine match probabilities. BALDING and NICHOLS (1994) assume a subdivided population of constant size, and use a coalescent argument to arrive at expressions for the joint allele probabilities within a subpopulation. Subpopulations are not necessarily independent, because of migration and common history. Genetic replicates are therefore defined as the combined evolutionary history of subpopulations. These authors show that when the marginal probabilities of an allele have reached a steady state, the joint allele probabilities within a subpopulation match the moments of a Dirichlet distribution. Note that since the genetic model is formulated without reference to a base population, measures of identity by descent are defined in terms of the coalescence of ancestral lines, without intervening mutations or migrations. The measures therefore depend on the rates of mutation, migration, and coalescence. Coalescence rates, in turn, depend on population size. Hence, under constant population size and mutation and migration rates, the measures of identity by descent remain constant over time.

WEIR and COCKERHAM (1984) proposed an estimator of the coancestry coefficient under a genetic model in which each subpopulation is constructed by randomly drawing individuals from a base population of infinite size. At the time of the base population, the probability of drawing an allele is assumed to be in steady state. The expected value of allele frequency in the subpopulation is therefore equivalent to the allele frequency in the base population. Thereafter, subpopulations are assumed to evolve under similar demographic conditions. Inference is conditional on allele frequencies in the base population, and subpopulations represent independent genetic replicates. In this prospective genetic model, identity by descent is defined with respect to the base population, and therefore decays with the time elapsed since the base population. Under the equilibrium of descent measures within subpopulations, higher-order descent measures can be written in terms of the pairwise measure of identity by descent (LI 1996). Then joint allele probabilities within a subpopulation have a form similar to those in the genetic model of Balding and Nichols, but are defined in terms of parameters in the prospective genetic model. LI (1996) used the resulting expressions to approximate joint allele probabilities early in the history of constant-size subpopulations, and found that the approximation performed well.

Both approaches invoke the equilibrium distribution of allele frequencies under the source-invariant mutation model. However, for the microsatellite alleles used in DNA profiling, a stepwise mutation model (e.g., FU and CHAKRABORTY 1998) may be more realistic. Under stepwise mutation, there is no stationary distribution of allele frequencies (MORAN 1975). In this paper, we investigate the robustness of the match probability theory to a stepwise model of microsatellite mutation, in growing rather than constant-sized populations.

METHODS

Demographic parameters: To simplify the analysis, we assumed the same demographic history for all subpopulations in our simulation study. The current number of 2×10^6 individuals in a subpopulation was chosen to be typical of the effective size of a modern population such as New Zealand Caucasians. Subpopulations were not of constant size over time, but instead underwent exponential growth. Each subpopulation arose 5000 generations before present (gbp) from 500 random individuals in a population of size 10000 individuals. Subsequently, each subpopulation evolved independently of the others, with no migration. Prior to 5000 gbp, the size of the meta-population giving rise to the subpopulations was constant at 10000 individuals, and there was no subdivision. Historical population sizes were based on estimates from the literature (e.g. HARPENDING *et al.* 1998; KRUGLYAK 1999), which suggest that the approximate date of the *Homo sapien* migration out of Africa is $G \approx 5000$ gbp, and that the effective population size prior to the migration was $N \approx 10000$ individuals.

Mutation model: Let π_{ij} be the probability that a mutation causes an allele size change from i to j. FU and CHAKRABORTY (1998) proposed a generalized stepwise mutation model in which π_{ij} depends on i and j only through $|i-j|$. A homogeneous distribution without constraints on allele size was preferred because it had fewer parameters and because only the relative sizes of alleles were known. Under their model,

$$\pi_{ij} = \begin{cases} \alpha P(1-P)^{j-i-1} & j > i \\ (1-\alpha)P(1-P)^{i-j-1} & j < i. \end{cases}$$

The parameter α describes the probability of an increase in repeat number; the size $|i-j|$ of the resulting change in allelic length has a geometric distribution with probability $P(1-P)^{|i-j|-1}$.

Table 1: Allele frequencies of microsatellite marker D8S1179 in New Zealand Caucasians.

	Allele length in repeat units									
	8	9	10	11	12	13	14	15	16	17
percent	2.13	0.45	11.63	7.72	14.54	33.33	17.56	9.40	3.02	0.22

Other parameters of the stepwise mutation model include the mutation rate μ, and the length A of the allele of the most recent common ancestor (MRCA) of the sample. We have selected a sample of size 1000 chromosomes. Simulations indicate that with high probability ($\sim$ .998) the sample MRCA coincides with the MRCA of the subpopulation. Even with a more modest sample size of 100 chromosomes, the probability is still very high ($\sim$ .980). Hence, A may also be viewed as the allelic length of the MRCA of the population. Given A and the realized ancestral tree, microsatellite mutations can be placed on the tree, from the root to the tips, as described by FU and CHAKRABORTY (1998). Conditional on the length of a segment of the tree, the number of mutation events on the segment is approximated by a Poisson random variable, with mean equal to the product of the mutation rate and the segment length.

For the simulation study, we chose mutation model parameters $\mu = 5 \times 10^{-4}$, $A = 9$, $\alpha = 0.720$, and $P = 0.999$, consistent with microsatellite allele frequencies shown in Table 1. Reported allele frequencies are based on a sample of 454 New Zealand Caucasian offenders. Selected parameter values are consistent with estimates from the literature. Microsatellites have a high mutation rate of $\sim 10^{-4}$ to 10^{-3} per generation (GYAPAY *et al.* 1994). Most observed mutations result in a change of a single repeat unit (WEBER and WONG 1993; DIRIENZO *et al.* 1994), with a tendency towards increasing allelic length (BRINKMANN *et al.* 1998).

Allelic associations: Following the notation of EVETT and WEIR (1998), consider a microsatellite locus **A** with alleles A_i of length i repeat units in a randomly mating subpopulation. Let p_i and P_{ij} denote, respectively, the probability of drawing an allele A_i, and the probability of drawing an individual with genotype $G = A_iA_j$ in the subpopulation at present. In both genetic models, p_i is assumed to be in steady state. Under the source-invariant mutation model, genotype probabilities may then be described by

$$P_{ii} = \theta p_i + (1 - \theta)p_i^2$$

$$P_{ij} = (1-\theta)2p_ip_j \qquad i \neq j,$$

where the coancestry coefficient θ is specific to the genetic model. In the genetic model of WEIR and COCKERHAM (1984), θ is the probability that two alleles drawn from the subpopulation at present are identical by descent with respect to the base population. In the genetic model of BALDING and NICHOLS (1994), θ is the probability that two alleles from the same subpopulation coalesce with no intervening mutation events on the lines of descent. However, for the high mutation rates typical of microsatellite markers, both measures are virtually identical given the demographic parameters of this simulation study. Figure 1 shows the coancestry coefficient, measured first with respect to the base population at 5000 gbp, and then without reference to a base population. At the selected mutation rate $\mu = 0.0005$, both coancestry coefficients are ~ 0.008. Coancestry coefficients were determined empirically based on 10^7 coalescent replicates for a random pair of chromosomes from a subpopulation.

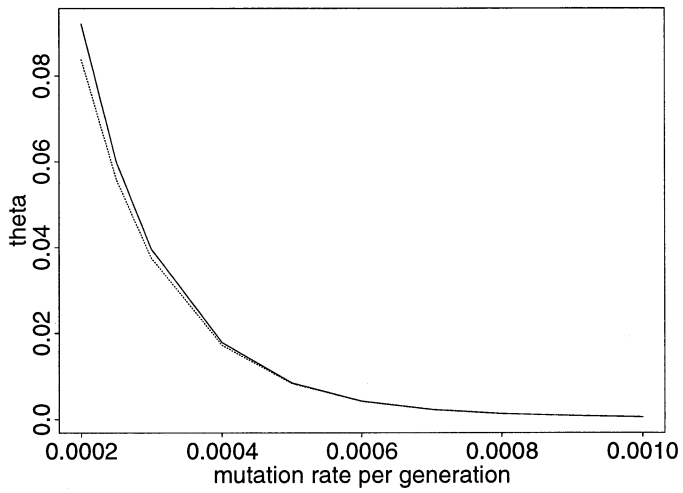

Figure 1: Coancestry coefficient versus mutation rate. Solid, pairwise probability of identity by descent without reference to a base population; dotted, with reference to the base population at 5000 gbp.

To gain insight into the adequacy of the source-invariant mutation model for microsatellite DNA profiles, we compared association parameters describing P_{ij} under the stepwise mutation model to the analogous quantity under the source-invariant mutation

model. Associations were determined empirically, based on 10^7 coalescent replicates for a random pair of chromosomes within a random sample of 100 chromosomes. The within-subpopulation correlation for an allele of length i repeat units is

$$\theta_{ii} = \frac{P_{ii} - p_i^2}{p_i(1 - p_i)}.$$

Under the source-invariant mutation model, this correlation coincides with the coancestry coefficient θ; hence $\theta_{ii} \equiv \theta$. More generally, however, θ_{ii} can vary with allele length i. Another measure of association within a subpopulation, between two alleles of different lengths $i \neq j$, is

$$1 - \theta_{ij} = \frac{P_{ij}}{2p_i p_j}.$$

In a stepwise mutation model, θ_{ij} is expected to vary with the allelic states (BALDING and NICHOLS 1994). Such variation is not accommodated by the source-invariant mutation model, which constrains $\theta_{ij} \equiv \theta$. As a diagnostic for the fit of the source-invariant mutation model, we examined the departure of θ_{ii} and θ_{ij} from the coancestry coefficient θ.

Predicted match probabilities: BALDING and NICHOLS (1994) showed that under the source-invariant mutation model joint allele probabilities within a randomly-mating subpopulation match the moments of a Dirichlet distribution, provided that the marginal allele probabilities are in steady state. They expressed joint allele probabilities in terms of marginal probabilities p_i and the coancestry coefficient θ, and used them to derive formulae for match probabilities. For a suspect ($\mathcal{S}$) with genotype G_S and perpetrator ($\mathcal{P}$) with genotype G_P, equations (1) give the expressions for genotypes A_iA_i and A_iA_j, $i \neq j$, respectively (EVETT and WEIR 1998).

$$\begin{aligned} \Pr(G_P = A_iA_i \mid G_S = A_iA_i) &= \frac{[p_i + \theta(2 - p_i)][p_i + \theta(3 - p_i)]}{(1 + \theta)(1 + 2\theta)} \\ \Pr(G_P = A_iA_j \mid G_S = A_iA_j) &= \frac{2[p_i + \theta(1 - p_i)][p_j + \theta(1 - p_j)]}{(1 + \theta)(1 + 2\theta)}. \end{aligned} \tag{1}$$

Empirical match probabilities were compared to those predicted by the equations using empirically determined values of p_i and assigned values of $\theta = 0.010, 0.050$, 0.100, and 0.150. Empirical values were based on 10^7 coalescent replicates for a random sample of four chromosomes, within a random sample of 1000 chromosomes drawn from a subpopulation. Empirical match probabilities were calculated by dividing the observed probability of

drawing two members of a subpopulation with the given genotype by the probability of that genotype.

Estimated match probabilities: We also examined the bias, over coalescent replicates, of the product rule estimator and an estimator based on the Dirichlet equations (1). The Dirichlet estimator is formulated unconditionally, over repeated realizations of populations or sets of populations. In contrast, the product rule estimator is formulated conditional on the observed population. However, bias of the product rule estimator across coalescent replicates should reflect a tendency towards bias at the fixed population level.

Typically, forensic databases are constructed using convenience samples from a limited number of subpopulations. To mimic such data, we simulated the ancestry of random samples of 1000 chromosomes from each of 5 subpopulations with demographic histories as described above. For each coalescent replicate, the samples were used to build a database of microsatellite allele frequencies. The overall database frequency f_i of an allele of size i repeats was used to estimate the expected frequency p_i. The product-rule estimator of match probability is $2f_if_j$ for a suspect and perpetrator with genotype G_iG_j, $i \neq j$, and f_i^2 for a suspect and perpetrator with homozygous genotype G_iG_i. An estimator taking into account coancestry was constructed by substituting database frequencies f_i for p_i, and a moment estimate for θ into the Dirichlet equations (1). We chose a moment estimator of θ which is easy to calculate and combines coancestry information across subpopulations (WEIR 1996). BALDING and NICHOLS (1997) introduced a Bayesian approach to modelling variation in θ among subpopulations in order to address the possibility that subpopulations may have different degrees of coancestry, owing to differing demographic histories. However, in the current study, all subpopulations were simulated to have the same coancestry coefficient. Hence, modelling of variation in θ is unnecessary.

RESULTS

Figure 2 shows the probability of drawing an allele A_i of length i repeat units from a subpopulation at present. The marginal distribution has a longer right tail, with a mode of 13 repeat units, and a mean of approximately 16 units. Over 90% of the time, an allele is between 9 and 28 repeat units in length. The mode and longer right tail of the distribution are consistent with the ancestral allele $A = 9$ and the parameter $\alpha = 0.720$ describing the probability of an increase in allelic length given a mutation event. Over

time, the mode of allele frequencies within a subpopulation tends to drift towards higher repeat numbers. Long ancestral trees generally have more such drift and a larger spread of allele lengths than shorter trees. Shorter ancestral trees result in more tightly clustered lengths, closer to the ancestral allele. As predicted (MORAN 1975), the spread of allele lengths in a subpopulation tends to be more stable than the mode, which can occasionally drift towards high repeat numbers. In fact, most variation in allelic length ($\sim$ 80%) is observed across coalescent replicates rather than within a replicate.

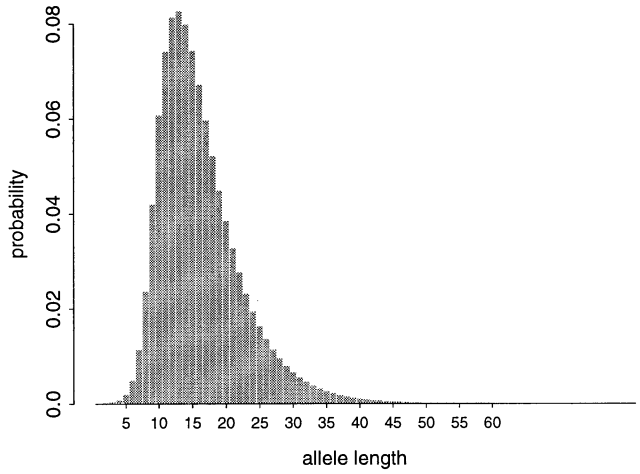

Figure 2: Probability of sampling an allele of a given length.

Allelic associations: Allelic correlations θ_{ii} are plotted in Figure 3. The stepwise mutation model introduces excess correlation, above the correlation of $\theta = 0.008$ (the coancestry coefficient) that would hold under the source-invariant mutation model. The average correlation weighted by allele frequency is $\sum_i p_i \theta_{ii} \approx 0.095$. Correlation is high for alleles of length 9 and 10 repeat units, which are associated with shorter ancestral trees. Short ancestral trees have alleles that tend to be more tightly clustered in length. Correlation is lowest for alleles with very low repeat numbers, which tend to derive from long ancestral trees carrying alleles with a wider range of lengths. Further simulation results indicate that, as expected, correlation is diminished at higher mutation rates and, when $\alpha = 0.5$, drops off symmetrically from the ancestral allele length of $A = 9$.

Figure 4 displays associations $1 - \theta_{ij}$ in the natural logarithmic scale for selected

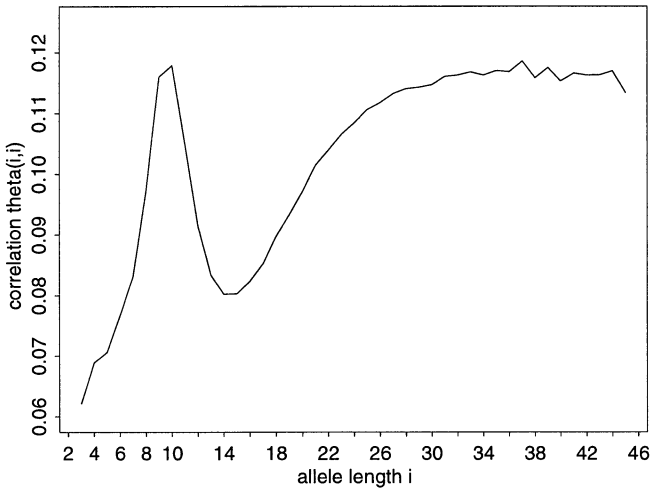

Figure 3: Within subpopulation correlations θ_{ii} for allele A_i.

genotypes A_iA_j, $i < j$. The rarer the allele, the stronger the association with alleles of similar but unequal length. Such positive association is at odds with the negative Dirichlet association between distinct alleles that is predicted by the source-invariant mutation model. Overall, $\sum_{ij} P_{ij}\theta_{ij} \approx -3.6$.

These diagnostics indicate that the source-invariant mutation model does not fully capture the pairwise distribution of alleles under a stepwise mutation model. It is therefore reasonable to expect that the joint distribution of three and four alleles, and hence the predicted match probabilities, would also be misspecified. In the next section, we investigate the impact of the stepwise mutation model on match probabilities predicted by the source-invariant mutation model.

Predicted match probabilities: Figure 5 shows empirically determined match probabilities, and the probabilities predicted by the source-invariant mutation model with various values of θ in the equations (1) for selected genotypes A_iA_j, $i = 13$ and $j \geq i$. For the common genotypes, predicted match probabilities systematically underestimate true match probabilities, except when θ is taken to be very high. For example, at $\theta = 0.150$, the predicted match probability for genotype A_{13}, A_{14} is approximately correct, but those for the A_{13}, A_{13} homozygote and the less common heterozygotes are too conservative.

Estimated match probabilities: Figure 6 shows empirically determined match probabilities, and the expected values of match probability estimators for selected

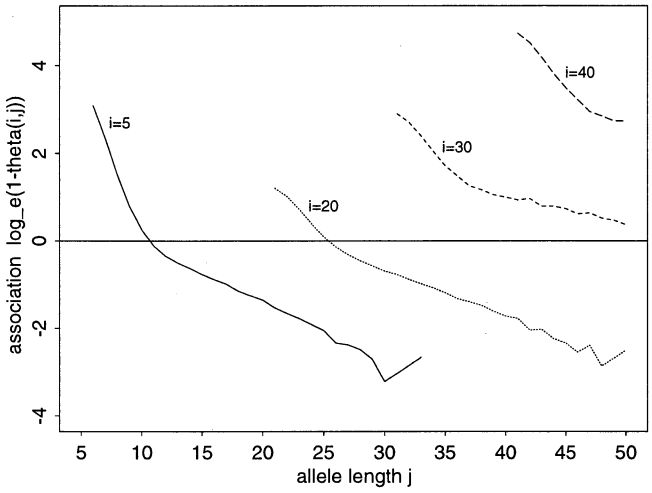

Figure 4: Within subpopulation associations $\log_e(1-\theta_{ij})$ for alleles $j > i$; solid line, $i = 5$; dotted, $i = 20$; short-dashed, $i = 30$; long-dashed, $i = 40$.

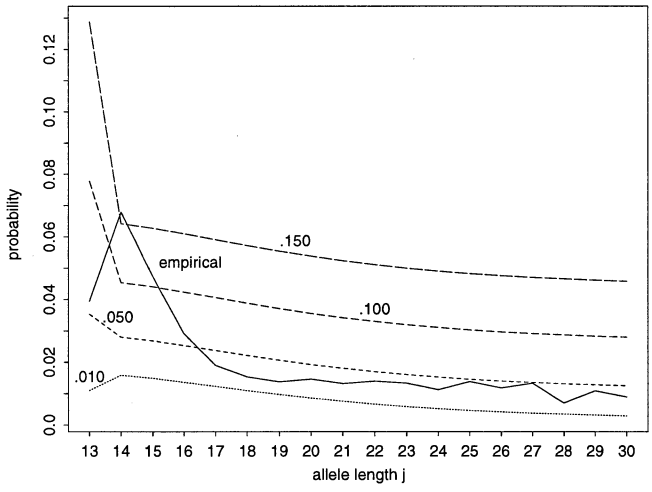

Figure 5: Empirical and predicted match probabilites for selected genotypes A_iA_j, $i = 13$ and $j \geq i$. Solid line, empirical probabilities; dotted, predicted probabilities $\theta = .010$; short-dashed, $\theta = .050$; medium-dashed, $\theta = .100$; long-dashed, $\theta = .150$.

genotypes A_iA_j, $i = 13$ and $j \geq i$. The product-rule estimator is systematically biased, with a tendency to underestimate match probabilities. The Dirichlet estimator is less biased, but still tends to underestimate match probabilities for common genotypes. For example, estimated match probabilities for a suspect with the more common heterozygous genotype A_{13}, A_{14} are expected to be about 51% and 31% of the true match probability, for the Dirichlet and product-rule estimators, respectively. Suppose profile data are available on five unlinked microsatellite loci, all with mutation parameters similar to the locus considered here. Then, in the case that the suspect carried the common genotype at all five loci, we would expect match probabilities to be underestimated by a factor of $.51^5 = 3 \times 10^{-2}$ with the Dirichlet estimator, and by a factor of $.31^5 = 3 \times 10^{-3}$ with the product-rule estimator, assuming statistical independence of alleles at unlinked loci. For ten loci, we would expect underestimation by factors of about 1×10^{-3} and 9×10^{-6}, respectively.

Given that the product rule estimator underestimates the true match probability, it is not surprising that for common genotypes so does the Dirichlet estimator. Predicted match probabilities for common genotypes A_iA_j involve larger marginal probabilities p_i and p_j in the numerator of equations (1). Larger p_i and p_j reduce the importance of the coancestry coefficient in the expression, and make predicted match probabilities more similar to those under the product rule.

DISCUSSION

There are several aspects of population genetics that require genotype probabilities and conditional genotype probabilities. In forensic assessments of DNA profiles, a theory of conditional genotype probabilities has been developed to account for the effects of coancestry on match probabilities (BALDING and NICHOLS 1995). Assumptions include a source-invariant mutation model with a stationary allele frequency distribution (WRIGHT 1951; GRIFFITHS 1979) and populations of constant size. However, for microsatellite DNA profiles, a stepwise mutation model would seem more realistic; for such a model there is no stationary distribution of allele frequencies (MORAN 1975). We have investigated the robustness of the theory to a generalized stepwise model of mutation, in growing populations. Although a variety of demographic and stepwise mutation models may be applied in a coalescent framework, we have opted for simple models as useful first approximations. While our results confirm that it is important to account for coancestry, they also cast doubt on the growing use of the Dirichlet approximation to the distribution

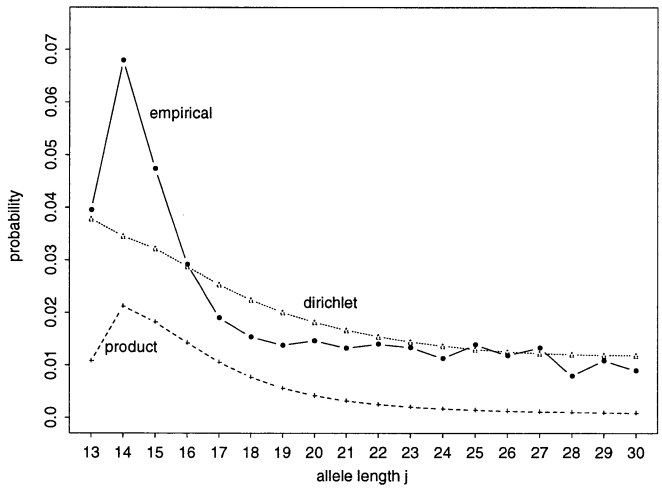

Figure 6: Empirical match probabilites, and expected values of match probability estimators for selected genotypes A_iA_j, $i = 13$ and $j \geq i$. Solid line, empirical probabilities; dotted line, expected value of Dirichlet estimator; dashed line, expected value of product estimator.

of microsatellite allele frequencies. We find that, although the coancestry coefficient is small ($\theta = 0.008$), the product-rule estimator is systematically biased, with a tendency to underestimate match probabilities. As shown in Figure 6, the estimator which takes into account coancestry is less biased, but still tends to underestimate match probabilities for the more common genotypes. However, as shown in Figure 5, such underestimation may be avoided by setting the coancestry coefficient to be very high. The price for such corrections is overly conservative predictions for rarer genotypes. For example, in the current study, some predicted match probabilities were more than three times the empirical value.

It is clear that allelic associations must be taken into consideration when estimating match probabilities for microsatellite profiles. However, as shown in Figures 3 and 4, these associations are inadequately characterized by the coancestry coefficient. Estimation procedures formulated under the source-invariant mutation model will therefore be ineffective. One alternative suggested by the current study is a coalescent-based estimator. For a given microsatellite locus, available data from well-characterized populations could be used to estimate the appropriate mutation parameters. FU and CHAKRABORTY (1998)

describe one such analysis. Estimated parameters could then be used to evaluate match probabilities empirically, in conjunction with a variety of plausible demographic histories for the population of the suspect. However, for a given mutation model, such a procedure would only be as good as the parameter estimates. The statistical properties of estimators of mutation parameters are uncertain, and require further investigation.

ACKNOWLEDGEMENTS

This work was supported in part by National Institutes of Health grant GM-45344 to North Carolina State University and by National Science Foundation grants DMS-9208758, DMS-9700867, and DMS-9711365 to the National Institute of Statistical Sciences, and in part by a postdoctoral fellowship from the New Zealand Foundation for Research in Science and Technology (FORST) to JMC. We thank John Buckleton and the Institute of Environmental Science and Research Limited of New Zealand (ESR) for the data, and Dr. Ian Painter for helpful discussions.

LITERATURE CITED

BALDING, D. J., and R. A. NICHOLS, 1994 DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int **64**: 125–140.

BALDING, D. J., and R. A. NICHOLS, 1995 A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity, pp. 3–12 in *Human Identification: The Use of DNA Markers*, edited by B. S. WEIR. Kluwer Academic Publishers, Dordrecht.

BALDING, D. J., and R. A. NICHOLS, 1997 Significant genetic correlations among Caucasians at forensic DNA loci. Heredity **78**: 583–589.

BRINKMANN, B., M. KLINTSCHAR, F. NEUHUBER, J. HÜHNE, and B. ROLF, 1998 Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet **62**: 1408–1415.

DIRIENZO, A., A. C. PETERSON, J. C. GARZA, A. M. VALDES, M. SLATKIN, and N. B. FRIEMER, 1994 Mutational processes of simple sequence repeat loci in human populations. Proc Natl Acad Sci USA **91**: 3166–3170.

EVETT, I., and B. WEIR, 1998 *Interpreting DNA Evidence*. Sinauer Associates, Inc., Sunderland, Massachusetts.

FU, Y.-X., and R. CHAKRABORTY, 1998 Simultaneous estimation of all the parameters of a stepwise mutation model. Genetics **150**: 487–497.

GRIFFITHS, R. C., 1979 A transition density expansion for a multi-allele diffusion model. Adv Appl Prob **11**: 310–325.

GYAPAY, J., J. MORISETTE, A. VIGNAL, C. DIB, C. FIZAMES, P. MILLASSEAU, S. MARC, G. BERNARDI, M. LATHROP, and J. WEISSENBACH, 1994 The 1993-94 Généthon human genetic linkage map. Nature Genet **7**: 246–249.

HARPENDING, H. C., M. A. BATZER, M. GURVEN, L. B. JORDE, A. R. ROGERS, and S. T. SHERRY, 1998 Genetic traces of ancient demography. Proc Natl Acad Sci USA **95**: 1961–1967.

KRUGLYAK, L., 1999 Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nature Genet **22**: 139–144.

LANGE, K., 1995 Applications of the Dirichlet distribution to forensic match probabilities. Genetica **96**: 107–117.

LI, Y.-J., 1996 . *Characterizing the structure of genetic populations.* Ph. D. thesis, North Carolina State University.

MORAN, P. A. P., 1975 Wandering distributions and the electrophoretic profile. Theor Popul Biol **8**: 318–330.

NRC 1992 *DNA technology in forensic science.* National Academy Press, Washington, DC.

WEBER, J., and C. WONG, 1993 Mutation of human short tandem repeats. Hum Mol Genet **2**: 1123–1128.

WEIR, B., 1996 *Genetic Data Analysis II.* Sinauer Associates, Inc., Sunderland, Massachusetts.

WEIR, B. S., 1994 The effects of inbreeding on forensic calculations. Annu Rev Genet **28**: 597–621.

WEIR, B. S., and C. COCKERHAM, 1984 Estimating F-statistics for the analysis of population structure. Evolution **38**: 1358–1370.

WRIGHT, S., 1951 The genetical structure of populations. Ann Eugen **15**: 323–354.