THE PATERNITY INDEX, POPULATION HETEROGENEITY, AND THE PRODUCT RULE

J.W. MORRIS[1] and D.W. GJERTSON[1,2]

[1]Long Beach Genetics, 2384 East Pacifica Place,Rancho Dominguez, CA 90220

[2]Department of Surgery, UCLA School of Medicine, Los Angeles, CA 90024

Determination of PI's Across Genetic Loci Without Use of the Product Rule

Recent controversy regarding the validity of the product rule in forensic DNA identification cases raises the possibility of living in a "product rule free" world. We have examined methods to determine PI's across genetic loci without using the product rule. In analogy with HLA, collection of haplotypes from family studies would permit such calculations from tables, but fixed bins and very large numbers of haplotypes would be required. Following a suggestion of C. Brenner that a "product rule free" PI can be obtained directly from multilocus VNTR phenotype tables, we have devised a simple method that relies only on counting: i) Determine, from direct counting, the proportion of individuals in the multilocus phenotype data base that match at all loci tested. This yields the "exclusion probability" term of the PI [J. Morris. 3rd Intl Symp on Hum Ident 1992, Promega p. 177-190, 1993]. ii) For each of the randomly matching phenotypes, determine from direct counting the relative "fit" to the mother-child pair compared to the phenotype of the tested man. The average relative "fit" gives the "goodness of fit" term of the PI. iii) Multiplication of the two terms yields the directly observed "gamete count" PI. For example, suppose in a paternity case tested at 3 VNTR loci, the tested man, heterozygous at all loci, matches the single paternal allele at each locus, and that search of the three locus data base yields 43 of 6146 non-fathers who also match. The exclusion probability PI is 6146/43 = 143. Of the tested man's eight equiprobable gametes, only one contains the paternal marker at each locus, yielding 1/8 as the numerator of the "goodness of fit" term. If the 43 matching non-fathers consist of 38 who match as well as the tested man and 5 who are apparent homozygotes at one of the three loci, the denominator of the "goodness of fit" term would be 1/43 (38•1/8 + 5•2/8), yielding a "gamete count" PI of 6146/48 = 128.

This method requires that genotype be directly determined from phenotype, so it cannot be used for systems such as HLA or ABO. "Motherless" cases would require direct observation of multilocus phenotype frequencies to obtain the "gamete count" PI, which is not practical for VNTR systems. The ability to obtain multilocus PI's independent of the product rule yields a method for validation of the product rule for multilocus PI (Fig I).

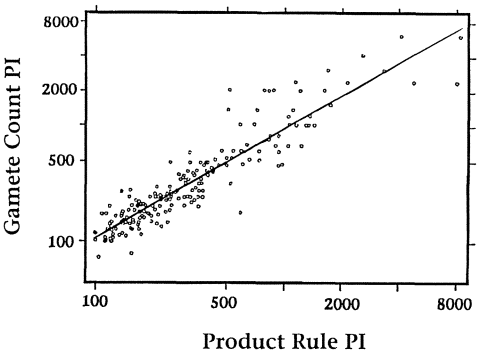

Fig 1 Correlation of gamete count and product rule PI. Data from 197 otherwise unselected cases for which product rule PI ≥ 100 for TBQ7 and EFD52. Gamete count PIs were obtained as described above from two locus data bases of size N=6146 (Caucasian), N=4148 (Hispanic) and N=2432 (Black). For one case with zero direct observation matches one match was assumed (Laplace's rule). 95% confidence intervals for slope and intercept of the regression line include 1 and 0, respectively.

Bioassay of Substructure of Paternity Populations

Critics have claimed that major US populations (Blacks, Caucasians, Hispanics) have substructure of sufficient magnitude to preclude use of intralocus (Hardy Weinberg) and interlocus product rules. It has been suggested that phenotype matching frequencies be compiled from data bases from the sub-population (ethnic group) of the suspect. Lacking such data bases, a ceiling principle could be adopted to allow for a "worst case" scenario. This view of population structure is equivalent to the affinal model "...the culprit is related to the suspect as closely as a spouse would be." [Morton, N.E. PNAS89(1992)2556]. For this model, Morton has shown that phenotype matching frequencies may be computed without ethnic specific data bases, provided that the inbreeding coefficient (α, F_{ST}), the standard estimate of population substructure, is available.

For disputed paternity, an analogous model is: the mother is related to the father as closely as a spouse would be. The actual ethnicity of the tested man is irrelevant for calculation of PI in VNTR systems - we know his phenotype and can accurately infer his genotype. What might be relevant is the ethnicity of other possible fathers, but since the relationship between mother and tested man reflects the relationship between mother and possible fathers, the issue of substructure (inbreeding) of the paternity population can be directly addressed from paternity casework. One result of inbreeding is increased homozygosity of offspring (Wahlund's rule). Estimation of inbreeding on this basis has two important limitations. First, at each autosomal locus the offspring is one of four possible results, so that Wahlund's method detects only 25% of similarities between mother and father. Second, the frequency of homozygosity expected on the basis of random mating and the frequency of silent alleles are critical to the calculations. For VNTR loci the former calculation depends on the level of discrimination of close alleles (delta), which can only be roughly estimated, and independent estimates of silent (null) alleles are difficult to come by for forensic (but not paternity) populations. One further limitation: excess homozygosity

determined from data bases compiled from mothers and tested men yields estimates of substructure one generation removed from the question at hand for disputed paternity (but not for disputed identity). Following the method of Morton, we have derived equations that yield estimates for $\alpha(F_{ST})$ from a determination of difference in matching of alleles between mother and tested man in each paternity case compared to matching between mother and random men:

$$PE_{M/AF}/PE_{M/RM} = \begin{matrix} 1\text{-}4\alpha \text{ (mother heterozygous)} \\ 1\text{-}2\alpha \text{ (mother homozygous)} \end{matrix}$$

For each paternity case, the tested man is considered to be a possible father of the mother, and the mean (motherless) exclusion probability ($PE_{M/AF}$) is compiled. This is compared to the same calculation treating each of the random individuals in the data base as a possible father of the mother ($PE_{M/RM}$). Excess sharing of alleles between mother and tested man would result in decreased probability of excluding the tested man as the father of the mother. This method has none of the limitations of Wahlund's method.

Estimates of α averaged from matching at 3 VNTR loci and calculated at three different values of delta are given in the table. N is the total number of mother-tested man comparisons made for each race. For the systems studied, our resolution is about $\delta = 2\%$ x MW. We use $\delta = 3\%$ x MW for casework; $\delta = 1\%$ x MW corresponds to a discrimination of less than one repeat unit. The mean values of α were computed for each race by averaging values obtained at loci D2S44(YNH24), D10S28(TBQ7) and D17S26(EFD52) [all restrictions HaeIII]. All estimates of α are close to 0, suggesting for each race that mothers and possible fathers are, for practical purposes, randomly paired. Similar results (based on fewer comparisons) were obtained for HLA and also for D1S47 [pL336, PstI].

Robustness of Estimates of α

δ	Caucasian	Hispanic	Black
1%	.0076	.0045	.0018
2%	.0016	-.0009	-.0010
3%	.0000	.0004	.0032
N	4751	3139	1823

Pairwise Independence of Loci

We tested for pairwise independence of phenotype matches among the five loci and three races. Overall, no departure from independence was found. Following the exposition of Morton, estimates of corrections needed in 4 DNA locus and 5 locus (4 DNA + HLA) phenotype matches were determined. For $\delta = 2\%$ x MW averaged over the three races the product rule would overstate the rarity of a four locus DNA phenotype match by a factor of 1.23, and for a five locus phenotype match (including HLA) by a factor of 1.32. Overstatement of PI values is estimated by the square root of these factors: the product rule would overstate 4 locus and 5 locus PI's by factors of 1.11 and 1.15 respectively.Use of even mildly conservative binning criteria would provide more than adequate compensation for both disputed identity and disputed paternity.