How to Deal with Mutations in DNA-Testing

R. Fimmers*, L. Henke**, J. Henke***, M.P. Baur*

* Institute for Medical Statistics, University of Bonn, Sigmund-Freud-Str. 25, Germany

** Institut für Blugruppenforschung, Otto-Hahn Str. 39, Düsseldorf

*** Institut für Blugruppenforschung, Hohenzollernring 57, Köln

INTRODUCTION

Some of the DNA single locus polymorphisms, which are presently used in paternity testing have a comparably high mutation rate. Consequently changes in fragment size occur quite often, during the transmission from parent to child. An isolated event like this may not be interpreted as an exclusion, especially if the mutation rate for the system in question is high. In combination with other information from blood-group-, HLA- and DNA-systems one would like to include the evidence against paternity from a possible mutation into the global likelihood statement. This requires the ability to calculate likelihood values for "mutation" patterns. From the theoretical point of view this is no problem, as will be seen in the following text. The problem arises with the estimation of the parameters, which describe the mutational event. A more global approach has to be used.

CALCULATION OF LIKELIHOODS

Figure 1 presents a typical "exclusion" or "mutation" band pattern from an arbitrary DNA single locus system. To calculate the likelihood of this band pattern it is necessary to include the possibility of mutations into the formal genetic model. It will be assumed,

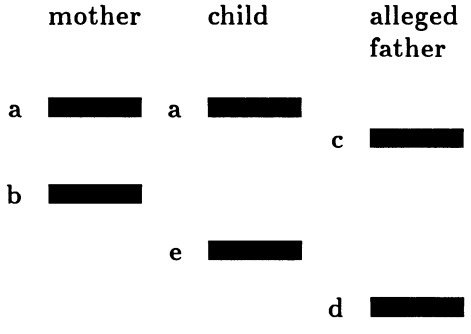

Figure 1: "Mutation" band pattern

that an allele x may change (its fragment size) to an allele y, while it is transmitted from parent to child. m_{xy} may denote the probability of such a mutational event. Including this possibility of change of fragment size, several explanations for the above band pattern in Fig. 1 possible under the assumption of paternity. One (unlikely) example is, that allele b is transmitted from the mother to the child and changes to allele e and that c is transmitted from the father to the child and changes into allele a. The probability of

Advances in Forensic Haemogenetics 4
Edited by Ch. Rittner and P. M. Schneider

Table 1: Possible explanation of the band pattern

maternal band → filial band	paternal band → filial band	transmission probability
$a \longrightarrow a$	$c \longrightarrow e$	$\frac{1}{2} \cdot \frac{1}{2} m_{ce}$
$a \longrightarrow a$	$d \longrightarrow e$	$\frac{1}{2} \cdot \frac{1}{2} m_{de}$

this event is $\frac{1}{2}m_{be} \cdot \frac{1}{2}m_{ca}$. This event is very unlikely, because it requires two mutational events. If all explanations, which require more than one "mutation" are considered to be negligible, because of their low probability, only two possible explanations remain (Table 1). The likelihood for the band pattern, under the assumption of paternity, can therefore be calculated as

$$X = \underbrace{2f(a)f(b)}_{\text{mat. phenotype}} \cdot \underbrace{2f(c)f(d)}_{\text{pat. phenotype}} \cdot \underbrace{\frac{1}{2} \cdot \frac{1}{2}(m_{ce} + m_{de})}_{\text{transm. prob.}} \tag{1}$$

The likelihood under nonpaternity can be calculated as

$$Y = 2f(a)f(b) \cdot 2f(c)f(d) \cdot \left[f(e)\frac{1}{2}(1 + m_{ba})) + f(a)\frac{1}{2}(m_{ae} + m_{be}) \right] \tag{2}$$

$$\approx 2f(a)f(b) \cdot 2f(c)f(d) \cdot f(e) \cdot \frac{1}{2} \tag{3}$$

and consequently the likelihood ratio is

$$\frac{Y}{X} = \frac{2f(e)}{m_{ce} + m_{de}} \tag{4}$$

The problem with the application of this formula is, that the specific mutation rates m_{xy} cannot be estimated. They are by far smaller than the overall mutation rate and depend on the two fragment sizes. Small changes in the fragment size seem to be more likely than larger ones.

In order to evaluate this type of band pattern in terms of likelihood ratios, one has to choose a less specific model. A mutation rate, which can be estimated concerning a singlelocus DNA-system is the overall paternal mutation rate μ, which may differ from the maternal mutation rate. More precisely μ is the probability of the occurrence of a "mutation" band pattern under paternity. At this level only one information is drawn from the band pattern. Either the pattern is compatible with paternity or not. The frequencies of the alleles which are involved in a pattern are not taken into account. The resulting likelihoods are the same for all "mutation" band patterns for a given system.

Under nonpaternity the probability of a "mutation" band pattern is $1 - r$, where r is the probability of a (wrong) inclusion. r can be estimated from empirical data or calculated from the allele frequencies f.

$$r = \sum_{a \in \{alleles\}} f(a) \left[2f(a) - f(a)^2 \right] \tag{5}$$

The likelihood ratio, which can be calculated using μ and r is

$$\frac{Y}{X} = \frac{1 - r}{\mu}. \tag{6}$$

APPLICATION TO DATA

This method was applied to data for the probes MS1, MS31, MS43, G3, YNH24. The estimated paternal mutation rates and "wrong inclusion" probabilities are summarized in Table 2. MS1 has the highest paternal mutation rate, but the smallest number of "wrong"

Table 2: Estimation of μ and r, likelihood values

	μ		r		$\frac{Y}{X}$	EM	W
	n	%	n	%			
MS1	591	4.57	281	2.14	21.41	11.33	0.045
MS31	600	2.17	300	6.67	43.08	11.63	0.023
MS43	594	0.51	298	7.72	182.71	12.26	0.005
G3	589	0.51	276	3.26	189.93	12.28	0.005
YNH24	349	0.86	179	6.70	108.54	12.04	0.009

inclusions, which both may be related to a high degree of polymorphism.

The resulting likelihood values (likelihood ratio $\frac{Y}{X}$, EM-value and W-value) are also presented in Table 2. They depend mainly on the mutation rates. The influence of the "wrong inclusion" probability is low. Smaller mutation rates lead to higher likelihood ratios, that means higher evidence against paternity. A "mutation" band pattern for MS43 or G3 has a higher evidence against paternity than a similar pattern for MS1.

DISCUSSION

To be able to calculate likelihoods and likelihood ratios for "mutation" band patterns from DNA single locus polymorphisms requires to incorporate the possibility of mutations (fragment size changes) into the formal genetic model. This can only be done in a more global approach, where only a part of the available information is used. The resulting likelihoods are valid for all "mutation" patterns for a given system. The frequencies of the alleles or fragments which occur in the band pattern cannot be taken into account, unless more specific mutation rates can be estimated.

The resulting likelihood values make it possible, to combine the "mutation" information from a DNA single locus system with the evidence from other systems. A problem may be seen in the fact, that the available information cannot be used completely. A small change in the fragment size should be more likely than a large change. The evidence against paternity should be therefore different for different band patterns. The likelihood values, which are calculated here have to be regarded as a kind of average for the total system in question. They should be handled carefully, because they may over- or underinterpret the true evidence of paternity. To be conservative in the sense of the not excluded father, one should perhaps use low estimators (lower confidence limits) for the mutation rates and the probabilities of "wrong" inclusion.