Mutations in Paternity

When a paternity test shows that a man is eligible to be the father at all but one locus, or all but a few loci, we consider the possibility that he may be the father but that a mutation (or maybe a few) has occurred.

The "two exclusion" rule

If the genetic pattern is inconsistent with paternity at one locus (often called an "exclusion", but that's a poor choice of word if you end up not excluding!) was traditionally (e.g. in serological times) not regarded as a demonstration of non-paternity, because of the possibility of mutation. However, mutations are quite rare in the traditional systems, so two inconsistencies were enough to exclude – to issue a judgement of non-paternity.

... and RFLP

The two-exclusion rule seems adequate for RFLP testing. However, it is not a good idea simply to ignore the inconsistent locus. A paternity index should be calculated for the locus, which takes into account the possibility of mutation.

... and STR

If a battery of a dozen STR systems is used for paternity testing, occasionally two inconsistencies can be expected even when the man is the father. On the other hand, we can also expect occasionally to see only two inconsistencies when the man is *not* the father.

Therefore calculating a paternity index for mutation is vital.

Treatment of mutations

If two inconsistent loci are deemed (nearly) enough to rule "exclusion" for the whole case, in effect two inconsistent loci together amount (nearly) to a combined PI=0. One inconsistent locus therefore deserves a very small PI, $0 < PI \ll 1$.

RFLP Approach

For RFLP's, I use the simple formula PI=μ, where μ is the observed rate of mutations/meiosis for the locus, when a locus shows a result inconsistent with paternity.

The rule that PI=μ is based on this simple model of mutation, assuming for the sake of illustration an obligatory paternal gene of Q:

X = P(man without gene Q will contribute Q)

= P(contributed gene will mutate) P(mutated gene will be a Q)

= μ P(Q).

Y = P(random sperm is Q) = P(Q).

X/Y = μ.

That is, the model assumes that the chance a gene will mutate to Q is proportional to the prevalence of Q genes in general, without regard to such possible complications as different probabilities for a small vs. a large or positive vs. negative mutation size change.

AABB formula for mutations

The AABB recommends a slightly different formula, for my facile analysis above is wrong. The evaluation of Y is slipshod. Better is

Y = P(paternal gene is Q and man has no Q)

= P(paternal gene is Q) $\cdot$ P(man has no Q)

= P(Q) $\cdot$ A

X/Y = μ/A,

where A is the *probability of exclusion*; the probability that a random man would have a pattern inconsistent with paternity at this locus. The extra complication does not change the result very much (since normally A is nearly 1) and the relative improvement is anyway dwarfed by the inherent inaccuracy of the method for estimating X.

The AABB however uses not the case-specific power of exclusion, A, but the mean power of exclusion, $\bar{A}$. I'm not sure why, but in either case the formula is correct on average (which mine is not), and in neither case is any of the three formulas anywhere near accurate.

Late breaking news: I have been told that as of December 2004 the official AABB recommendation is the model discussed on this page.
Later: But when I looked they seem equivocal, accepting of either my method or of "Fimmers" which would be the old RFLP way.

Mutation formulas not accurate

The general rule is that mutations change allele length by a small amount. Thus when the man mismatches the child allele by a small amount there is a lively chance of a mutation but the formula underestimates it, tending to help true fathers get off the hook. Conversely, when the alleles mismatch by a large amount, the same computation unfairly inflates the possibility of a mutation, perhaps unfairly overstating the evidence against a non-father.

Nonetheless, as regards RFLP, I don't have a better formula to suggest. Vigilance.

STR Approach

In the case of STR systems I think we can do a little better. For example, Brinkmann et al reported that of 23 STR mutations found, 22 were by a single step; one by a double step. Fourney et al (private communication) have similar experience, and mention that length-increasing and length-decreasing mutations are about equally frequent.

Therefore as a rule of thumb I suggest assuming that

- 50% of all mutations increase by one step
- 50% decrease by one step
- 5% increase by two steps
- 5% decrease by two steps
- 0.5% increase by three steps
- 0.5% decrease by three steps
- ... etc.

Never mind that these numbers add to more than 100%!

STR Formulas

- **Typical case**

  Unlike the RFLP case, the formula will depend on the actual alleles and possible patterns of sharing. Instead of trying to give a general treatise, I'll just illustrate with one typical example.

  Suppose the mother is PP, the child is PQ, and the man is Q'R, where Q' is s=1 or 2 steps smaller (or larger) than Q. As usual, let μ be the overall mutation rate for the locus, measured in mutations/meiosis.

  Then $X = P(\text{genotypes} \mid \text{true trio})$ is proportional to
  $P(\text{Q'R man transmits Q}) = P(\text{Q' is transmitted}) \times P(\text{mutation}) \times P(\text{mutation increases length}) \times P(\text{s steps})$
  $= (1/2)\,\mu\,(1/2)\,(1/10)^{s-1}$, and

  $Y = P(\text{genotypes} \mid \text{false trio})$ is proportional to

  > Note Apr 2010 This model is almost the same as the 1994 model of Di Rienzo *et al* (link above).

  $P(Q)$, the frequency of the Q allele among random sperm.

  Therefore I propose the approximations that the paternity index $X/Y = \mu / (4\,P(Q))$ if the mutation is 1 step,
  $X/Y = \mu /(40\,P(Q))$ if the mutation is 2 steps, etc.

  Please note that the above formula is just an example and the formula would be different if the genotype patterns were different, or if the model parameters were different (e.g. 2-step mutations may be 20 times rarer than 1-step rather than 10 times rarer).

- **More examples of STR Formulas**

mother	child	tested man	formula	remark
PP	PQ	Q'R	$\mu / (4\,P_Q)$ if the mutation is 1 step, $\mu /(40\,P_Q)$ if the mutation is 2 steps, etc.	
17	17 18	19	$\mu / (2\,P_{18})$	
12 14	12 14	15 16	$\mu / (4(P_{12} + P_{14}))$	ambiguous paternal allele
12 14	12 14	13 15	$3\,\mu / 4(P_{12} + P_{14})$	3 possible paternal mutations

15 16	15 17	15 16	$\mu(1+1/\underline{3.5}) / (4 P_{17})$	paternal or (less likely) maternal mutation
none	20.2 23	21.2 23.2	$\mu / (8 P_{20.2})$	23.2 –> 23 mutation too rare to consider (see **Lower bound**)
father	**child**	**tested woman**		
17	17 18	19	$\mu / (2 \bullet \underline{3.5} \bullet P_{18})$	maternity case

- **Strange example – mutation without μ**

mother	child	tested man	formula	remark
12 15	13 16	16	$1/(P_{16} + P_{13})$.	Since same (maternal) mutation possibilities are involved in both numerator and denominator of the LR formula, the μ's cancel and it is the same formula as if mother=child (i.e. no mutation). Thanks to LiHaiXia of Sun Yat-Sen University for a correction. 29July2009
12 14	13 15	15	$1/(P_{15}+P_{13}/2)$	Again the μ's all cancel because both numerator and denominator require maternal mutation. But there are 3 possible mutations for the denominator and only 2 for the numerator. So this case is *not* equivalent to any "no-mutation" pattern.