

What Can The 90's Teach Us About Good Forensic Science

John Buckleton

The late 80's and 90's saw a rapid development in DNA technology. In most, if not all cases, there was agreement amongst the molecular biologists that this represented real progress. It is not intended to talk about this nor is the author qualified to do so. All that is intended is to contrast this situation with the interpretation issues of our time. In most cases the interpretation "arguments" were never settled. In many cases they were superseded by technology and in others they remain. It is reasonable to ask why we have been unable, by science alone, to settle the interpretation matters of our day?

WHAT IS THE QUESTION?

Perhaps the most important aspect of interpretation is to work out what question needs answering. A prominent commentator [1] has offered the following

"The germane question is: What is the frequency in the database"

There are a number of aspects of this statement that are worrying. First 'germane'. Oxford gives this as: "Relevant, pertinent to the matter or subject". It must certainly be agreed that it is important to ask relevant questions so it certainly seems important to define the germane question.

What is offered as the germane question: "What is the frequency in the database?"

Since this question is germane let us try answering it? Imagine that we have a 6 locus STR match. We search the database of say 200 persons (a plausible size in the USA) for this profile and typically do not find it? We can now answer the germane question: The frequency in the database is zero.

Your honour: "The germane question is: What is the frequency in the database? And, I tell you truthfully, that the frequency in the database is zero."

It is not that clear what the courts are going to do with this. It is also not clear why the audience that heard this statement thought it so good. The question, if taken literally, is not much use.

If it cannot be taken literally then can the question be improved? Perhaps the author didn't mean database, perhaps he meant population. Let us try:

"The germane question is: What is the frequency in the population?"

Here the last word has been changed from database (where the frequency is zero) to population. This seems quite a lot better. Armed with this, we approach our database of 200 persons and attempt to answer this "germane" question. What can we do next? What has happened is that we have changed from a database which is completely characterized and where we can count any frequencies in which we are interested to "population" where we do not have this knowledge. There is uncertainty about the population. We do not KNOW the frequency of the genotype in the population. We may even have difficulty defining the population. Answering this question of the frequency in the population is completely impossible from the set of 200 persons without some additional scientific equipment. We need to know how to assemble some comment about a population from a sample. This cannot be done armed only with the question and the data. We will need to accept that we are now estimating something and that there is uncertainty in the estimation process. We will need, at least:

1. A population genetic model.
2. A statistical theory to assess sampling error (random and systematic).

What may have been meant when he offered his “germane” question was: What was the product of the 12 allele frequencies in the database (with factors of 2)? Or perhaps: What is the product rule estimate from our database? It would be sad if the audience’s positive reaction was because they liked the simplicity of this statement or that they knew how to do it or that they didn’t have to learn anything new.

There is a marked contrast between this question and a different one offered in the 80’s:

What is the probability that this blood came from that person? [2]

How do these two questions differ?

1. What is the product rule estimate from our database?
2. What is the probability that this blood came from that person?

The second was set up deliberately to answer a question of relevance to the court. The first did not start from this formalized line of thinking. The second is the more “germane” question.

This is the lesson that can be taken from the analysis of this question: If you don’t have some firm principles on which to base your forensic thinking you can be left relying on your intuition. Sometimes this will be adequate but it will seldom stand up in comparison to a professional, scholarly and logical approach.

What is wrong with intuition. Isn’t all this interpretation stuff just common sense. I am after all a PRACTICAL Forensic Scientist. “Science is nothing but trained and organized common sense, differing from the latter only as a veteran may differ from a raw recruit: and its methods differ from those of common sense only as far as the guardsman’s cut and thrust differ from the manner in which a savage wields his club.[3]”

We need to decide: Will forensic science in the 00’s be an ad hoc, seat of the pants art, or a scholarly professional science? Shall we apply the best science of our day in interpretation just as we do in the technology or shall we let a love of overly simple approaches blunt our science. Is this love of simplicity consistent with the truth? Do we love it because we have sympathy for the jury or because of personal reasons. “Should we force science down the throats of those who have no taste for it? Is it our duty to drag them kicking and screaming into the 21st century? I am afraid it is.[4]”

What then, are the remaining arguments in DNA interpretation. Can any depth of insight be gained by approaching these questions freshly, with a purely scientific frame of mind:

UNIQUENESS

A topic of some current interest is variously called uniqueness or individuality, although these terms may be poor. There has been some discussion of this by the FBI and most notably by Ian Evett and Bruce Weir[5]. It is not intended to try to add to the details of that discussion, but what to point out that both Bruce Weir and Ian Evett appear to be trying to deal with the question of: “What is meant when someone uses the words uniqueness or individuality?” This leads both to some interesting thoughts about the fact that all things are different (in some sense) but that what we are actually trying to do is determine whether two things have the same source. Contrast this with the FBI approach which appears to be to accept that there is some concept that they call uniqueness or individuality and then seek to define some objective criteria that meets this (defining the criterion subjectively of course).

If we take the FBI approach we can proceed with the following logic:

Your honour in my opinion this evidence meets an objective criterion for uniqueness. What is uniqueness? I’m not sure your honour but I’m sure we’ve met the requirements.

It is interesting to ask why the FBI are adding a “uniqueness” criterion in DNA and taking one out in fingerprints. Is this simply the evolution that evidence tends to go through with time? Is there any underlying principle driving these different trends?

PRODUCT RULE OR NOT

Probably the biggest single argument of the 90’s in DNA has been the assumption of independence.

There are two competing population genetic models in use for DNA evidence. These are “the product rule” and the “drift model.” Neither can model anything as complex as modern human populations exactly.

Many commentators in the USA use the product rule or a modification of it that appears as recommendation 4.1 in NRC II [6]. There are reasonable grounds to doubt this model, notwithstanding the support given to it by the National Research Council. It has been adopted by many laboratories because of its simplicity rather than an underlying belief in its validity. The product rule requires an assumption of within and between locus independence which cannot be exactly true. This assumption requires that everyone in the population is unrelated to everyone else. In fact in all real populations there is a complex web of relatedness. It is an interesting mind experiment [7] to take the modern human population of, say, 6 billion, and work out that we must have had 12 billion parents, 24 billion grandparents and so on. Since these people did not exist it is clear that we share ancestors. Modern theory has it that all modern human populations are descended from a small group of persons living, possibly in Africa, possibly 100,000 to 250,000 years ago. This simple thought experiment reveals that we must all be related.

Why are we calculating a frequency at all? The answer, typically, is that we are estimating the chance that someone other than the suspect has the profile. If we call this person the offender then we seek the chance that the offender has the profile, given that the suspect already has it. Since we have just argued that all people are related then the suspect and the offender are related. It does not make much sense therefore to use the product rule which assumes that they are not. The effect of low levels of relatedness in the population is not large, however it does seem wise to quantify it rather than ignore it. There are many other lines of logic that lead to the same conclusion: The product rule cannot be exactly true.

The product rule has a tendency to overstate the value of the forensic evidence by a small amount. Dr Budowle has termed this difference forensically insignificant. However it seems wise to let the court determine what is and what is not significant. It also seems wise to quantify the effect rather than dismissing it.

Nichols and Balding [8] drew this to our attention in 1991. New Zealand did not, comprehensively abandon the product rule in both simple stains and mixtures until 1995. What was the reason for the delay? We were not slow to read the paper, we were simply slow to accept the result. Is this an example of innate bias in our thinking or just bad science?

Some commentators, including the author, initially approached the problem of DNA profile interpretation from a purely statistical point of view, ignoring that they are genetic entities. This can lead to some very poor logic. Most prominent is the supposition, mentioned above, that independence testing somehow validates the product rule. It does not and cannot. Independence testing has almost no power to differentiate between the two models discussed here and often, when it does lend its weight one way or the other, it is toward the drift model. Many of us started making decisions on the DNA model based on statistical testing of a few hundred or a few thousand Caucasians. At this time we did not realise that such independence tests are quite powerless and that Caucasians are probably the race least likely to enlighten us.

A purely statistical approach cannot therefore be correct. It is necessary to consider that these genetic profiles have an evolutionary history [9]. For these reasons the second model, “the drift model” (recommendation 4.2) is much preferred. Those people adhering to the product rule have, in all probability, not thought through the logic of their support for it.

INDEPENDENCE TESTING

It has become customary to perform independence testing before adopting the product rule or recommendation 4.1. It is important to note that independence testing cannot prove independence. All it can ever do is find dependence. In fact many datasets do exhibit mild dependence, however the power of modern statistical tests with realistic datasets is very low and their ability to find small disequilibria is weak. Therefore it is not reasonable to infer that independence is true after independence testing.

Do not be fooled by the term “no statistically significant departures were found.” This means that testing did not find departures, not that they are not there. They may require a larger dataset to be found. From population genetic arguments we know that there must be some level of departure from independence. Just because you didn’t find a needle in the haystack does not mean that there isn’t one there.

There are several practices adopted by some practitioners that are particularly unwise. Perhaps most unwise is the practice of post hoc rationalization. As an example we discuss Budolwe [10] et al. Who, when testing the FBI’s HBGG locus, appear to have combined the A and C alleles. There is a significant departure from independence if the alleles are not combined, caused largely by a single CC homozygote¹. This result does not appear in the paper and would take careful repeating from original data to find. Budowle et al. conclude that the data are “consistent” with Hardy-Weinberg expectations which is an unreasonable statement considering the (presumably) post hoc combination of two alleles. Whether or not we accept that this combination was acceptable it is not reasonable to use this data as evidence FOR independence when it is equally, or more so, evidence FOR subpopulation effects

The presence of one datum that, largely by itself, gives a departure from independence on testing is a common event. Such an event may sometimes be spurious or may be evidence of subpopulation effects. It is not reasonable to make one inference above the other on the basis of independence testing alone.

There are actually a number of post hoc data treatments that can occur. The process is often to perform independence testing and then find which data are causing the departures. Scrutiny (post hoc) of this data may reveal, a typographic error, a typing error, a reason to doubt the ethnicity of this sample, or that the sample has rare alleles. If the error is typographic or typing then it can be corrected however it must be remembered that this datum was singled out and checked, the other data have not be so scrutinized. Again if the ethnicity of the sample is in doubt it can be removed but again the remaining samples have not been subject to similar scrutiny. Some offending data remain even after this treatment. Budowle et al’s approach with one of these appears to have been to combine it with the most common allele. This is acceptable practice in some testing situations. However given the arsenal of weapons available to the independence tester to “make departures go away” it is not reasonable to use this approach to validate the assumption of independence.

Another well used method for explaining departures is to invoke “multiple testing”. This is a very realistic problem. If we operate at the 5% significance level then about 5% of our tests should exhibit departures even if they are not there. If we perform 100 tests we expect about 5 of them to show departures by chance alone. A statistical practice that is sometimes applied is the Bonferroni correction. This correction changes the significance level based on the number of tests. There are a number of theoretical objections to this practice.

First the Bonferroni correction is only applicable when the hypothesis being tested is the same. In our multiple testing situation the hypotheses are all different. For instance we test the hypothesis that locus 1 is in Hardy-Weinberg equilibrium. Next we test that hypothesis that locus 2 is in Hardy-Weinberg equilibrium and so on. These are different hypotheses.

¹ The C allele is rare, therefore a CC homozygote should be very rare.

Second by adjusting the significance level we lower the power of an already powerless test.

It is better practice to attempt to draw an inference from all the p-values². There are a number of ways to do this but a simple graph appears to work. An example is given in figure 1.

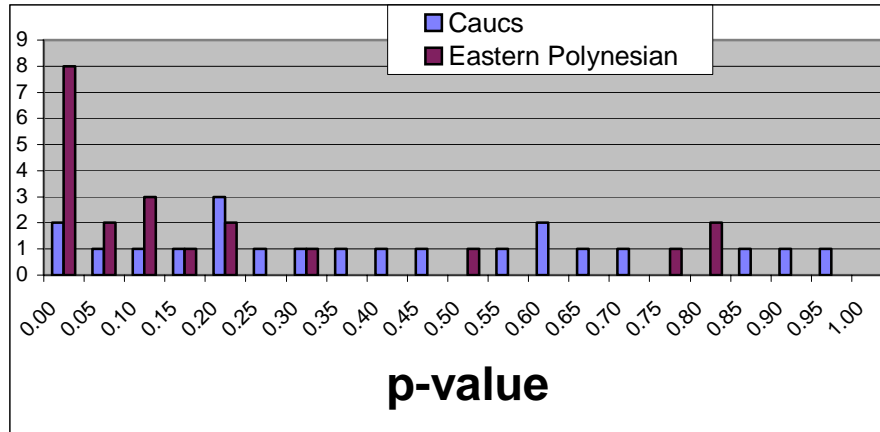


Fig 1. A plot of the p-values obtained for Hardy-Weinberg and Linkage testing on 3290 Eastern Polynesians (Maori and Cook Islanders) and on 1764 New Zealand Caucasians

This figure shows the p-values crowding down towards the low end for eastern Polynesians: A typical result of disequilibrium. It is likely that most researchers would conclude that this is an example of disequilibrium (and we have edited out the identical twins and the pair of matching brothers).

Figure 1 shows the p-values with a less obvious tendency towards the low end for the New Zealand Caucasians. The mean of this set is 0.40 not much less than the expected value of 0.50 if the population was in fact in Hardy-Weinberg and linkage equilibrium. It may be possible to explain the three “significant” results in some way or apply the Bonferroni correction. However a more realistic, and genetically correct approach would be to interpret this as the expected result from mild disequilibrium.

These two results are typical of datasets that we have examined worldwide although they have more data than average. Typically disequilibrium in Caucasians is small whereas in other races it is usually larger.

Presented here was a very short summary of some issues that have occurred in the 90’s in DNA interpretation. Disagreement on them can often lead to heated debate. Even now there is little international consistency on the main points mentioned here. The USA is largely using the product rule. Many other areas are not. It is argued that this is not because reasonable persons may hold opposing view but rather because there is no agreed method of scientific decision making in forensic science. “Simplicity” is often viewed as an excuse for avoiding superior science in the interpretation field, although seldom is this excuse used in the technology field. It has been argued that there is not even agreement over the question to which an answer is sought even though the published literature is replete with well written articles giving a coherent approach. We can only hope that “When men understand what each other mean, they see, for the most part, that controversy is either superfluous or hopeless [11].”

1. Budowle B. National Forensic Science Training Centre, Florida, 1996.
2. Evett IW. (1983) What is the probability that this blood came from that person: A meaningful question? *J Forensic Sci Soc* 23:35-59.
3. TH Huxley.
4. George Porter 1986
5. Evett, I.W. and Weir, B.S. (1998). *Interpreting DNA evidence*. 2nd. Ed. Sinaur, Sunderland, MA.

² The p-value is the probability of obtaining data this, or more, extreme if the hypothesis is true.

6. National Research Council (1996). The evaluation of forensic DNA evidence. *National Academy Press*, Washington, D.C.
7. BS Weir per comm
8. B Budowle pers comm.
9. Nichols RA and Balding DJ. (1991) Effects of population structure on DNA fingerprint analysis in forensic science. *Heredity* 66:297-302.
10. BS Weir pers comm.
11. Budowle B, Lindsey JA, Decou JA(check reference I can't read my fax properly), Koons BW, Guista SM, Com..... JFS 40:45-54, 1995.
12. Cardinal Newman 1839 in a sermon at Oxford.