Posted 5/11/08

# BEAT THE ODDS, GO TO JAIL

### *DNA random match probabilities may be overstated*

   *By Julius (Jay) Wachtel*.  Looking for a growth industry?  Think genetics. With more than one million profiles, California's DNA databank is the third largest in the world, trailing only those of the FBI and Great Britain. At its 1990 debut the GoldenState's database only kept track of sex offenders, but it has since expanded to include everyone convicted of a felony.  What's more, starting next year DNA specimens will be collected from every adult *arrested* for a felony, a move that should increase the databank's size by 390,000 profiles *each year*.

   When DNA got its start there weren't databanks, so police had to have someone in mind to make crime scene DNA useful.  Now it's possible to run unknown DNA through massive databanks like California's hoping for a "cold hit."  A recent example is the case of John Puckett, a previously convicted rapist who is appealing his conviction on a thirty-year old rape/murder. An expert testified that there was only one chance in 1.1 million that the match between Puckett's DNA and the crime scene sample could have happened at random.  With a probability of error that low, prosecutors suggested there was only one explanation: both samples came from the same source.  Not unexpectedly, jurors agreed, sending the 70-year old to prison.

   Since the human genome is exceedingly large, DNA is only typed at thirteen known places ("loci") in the strand.  Each location has two chemical sequences ("alleles"), one inherited from each parent.  Scientists have determined how often specific loci/allele combinations occur in different populations, such as Caucasian males.  Single combinations are commonplace and can be present in one out of every three or four persons.  Multiple loci/allele combinations occur less frequently.  In this example the probability of randomly selecting a DNA profile with four specific loci/allele combinations is 14 in 100,000.

   Just like with fingerprints, a single dissimilarity between DNA profiles means that they're not from the same person.  If no differences are observed a sufficient number of identical loci/allele combinations must be present to suggest that they have a common origin.  How many is enough? There's no set answer. Five and six loci/allele combinations can yield probabilities of a random match in the one-in-a-million range; while seven or more can generate probabilities in the hundreds-of-millions, billions, trillions or even quadrillions.  (For an online tool that lets users run a sample profile, click here.)

When a suspect is independently developed a subsequent DNA match obviously carries enormous weight. Still, the DNA match alone is *not* a probability of guilt -- it's an estimate of the likelihood that DNA *drawn at random* will match the profile of crime scene DNA. (Probability of guilt requires that all other pertinent factors be considered. This requires use of Bayes' theorem.)  Random match probabilities also assume that we only draw *once* from a population.  But that's not what happens in cold hits. No one knows whether the match that cooked Puckett's goose came on the computer's first draw or last (at the time California had 338,000 DNA profiles online.)  Had the expert witness followed the recommendation of the National Academy of Sciences he would have multiplied the random match probability of one in 1.1 million by the number of draws (338,000), yielding a true random match probability -- in effect, the chance of mistakenly identifying an innocent person -- of *one in three.*

Interestingly, the expert told a reporter that he didn't mention the adjustment, which he agreed was a superior approach, because the judge wouldn't allow it.  After the trial jurors said that the probability of one in 1.1 million was a key factor in deciding to convict. Asked if correcting it might have affected the verdict a juror said, "of course it would have changed things. It would have changed a lot of things."

Bigger DNA databases will yield more matches. While that seems beneficial, more profiles mean  more draws, so the probability that matches may be caused by chance will increase. Of course, random match probabilities with denominators that approach or exceed the population of the U.S. or the planet will remain noteworthy. In any event, understating the probability that a match might point to the wrong person is no solution.  At least one expert has already warned that an invaluable tool for freeing the innocent -- DNA -- could inadvertently become an instrument of wrongful conviction.

Only days ago Puckett's appeal was argued before the California Supreme Court. Its decision is expected soon. In the meantime keep away from the lottery.  The probability of hitting it is so low that if you do, it could be evidence that you fixed it!