Stanford Continuing Studies, Course MATH 07, Fall 2005

Three views of mathematics

LECTURE 9

Keith Devlin

STANFORD UNIVERSITY

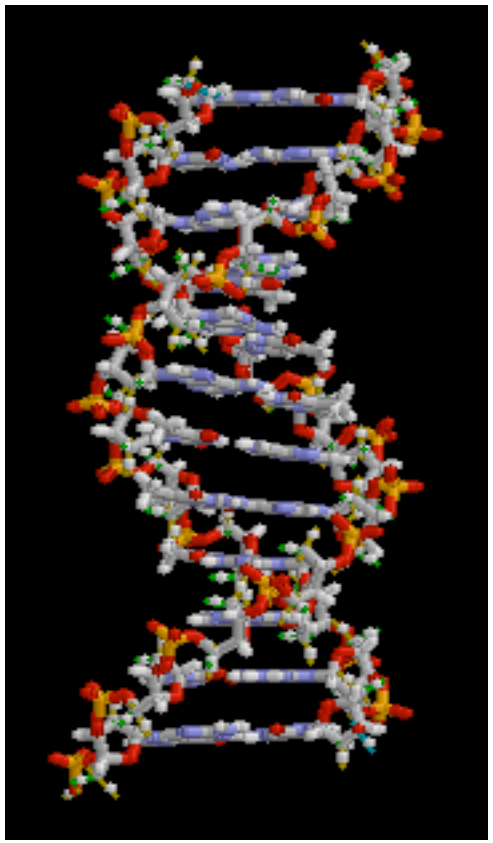
devlin@stanford.edu

The Math of DNA Profiling in Cold Hit cases

- *United States versus Raymond Jenkins*
- 1999 murder case in Washington, D.C.
- An initial suspect, who was charged, was subsequently eliminated after his DNA profile (taken on the 13 loci used by the FBI's CODIS system) did not match that obtained from blood samples taken at the crime scene.
- A "Cold Hit" search of the CODIS database failed to find a match to the crime scene sample profile.
- A Cold Hit search of the Virginia convicted offender database (101,905 8-loci profiles) produced a match with Raymond Jenkins.
- A 13-loci profile taken from Jenkins under a search warrant was found to match the profile of the crime scene samples.
- Jenkins was arrested on January 13, 2000.

DNA Profiling - the science

- DNA profiling
- Double helix molecule, four kinds of bonds: adenine (A), thymine (T), guanine (G), and cytosine (C).
- Roughly 3 billion of these nucleotide bonds in a human DNA molecule.
- Signature: … AATGGGCATTTTGAC …
- Human DNA can be regarded as made up of 46 chromosomes.
- Each chromosome can be regarded as made up of about 550 genes.
- A gene is thus a locus on a chromosome.
- DNA profiling is done using the patterns of the genes (loci) at certain specific sites.



DNA Profiling - the procedure

- The probability of a random match at one locus is about 1/10.
- The random match probability (RMP) is obtained using the product rule to multiply probabilities.
- The probability of a random match at 8 loci is $1/10^8$, i.e. 1 in 100 million.
- The probability of a random match at 13 loci is $1/10^{13}$, i.e. 1 in 10 trillion.
- [The government figure in the Jenkins case of 1 in 26 quintillion seems excessively high.]
- The FBI CODIS database (**CO**mbined **D**NA **I**ndex **S**ystem) stores DNA profiles based on 13 loci.

DNA Profiling in a Cold Hit process

National Research Council reports

- NRC I (1992): “The distinction between finding a match between an evidence sample and a suspect sample and finding a match between an evidence sample and one of many entries in a DNA profile databank is important. The chance of finding a match in the second case is considerably higher. … The initial match should be used as probable cause to obtain a blood sample from the suspect, but only the statistical frequency associated with the additional loci should be presented at trial (to prevent the selection bias that is inherent in searching a databank).”
- NRC II (1996): “When the suspect is found by a search of DNA databases, the random-match probability should be multiplied by N, the number of persons in the database.”
- The statistic NRC II recommends using is generally referred to as the “database match probability” (DMP).

A contrary view

- Dr. Peter Donnelly, Professor of Statistical Science at the University of Oxford. "… after a database search, the DNA evidence … is somewhat stronger than in the setting in which the suspect is identified by non-DNA evidence and subsequently found to match the profile of the crime sample. … I disagree fundamentally with the position of NRC II. Where they argue that the DNA evidence becomes less incriminating as the size of the database increases, I (and others) have argued that in fact the DNA evidence becomes stronger. … The effect of the DNA evidence after a database search is two-fold: (i) the individual on trial has a profile which matches that of the crime sample, and (ii) every other person in the database has been eliminated as a possible perpetrator because their DNA profile differs from that of the crime sample. It is the second effect, of ruling out others, which makes the DNA evidence stronger after a database search…"

The DNA profiling debate

Five methods suggested in the Jenkins case:

- Report the RMP alone. (The government's desired approach)
- Report the DMP alone. (The NRC II recommendation)
- Report both the RMP and the DMP. (Government fallback)
- Use a Bayesian approach (*à la* Donnelly)
- Follow the NRC I recommendation, and report the RMP for confirmatory loci only.

- The *Frye* test (based on a 1923 ruling): admissible scientific evidence must be based on a "well-recognized scientific principle or discovery [that is] sufficiently established to have gained general acceptance in the particular field to which it belongs".