Matching and Partial Matching

When two autosomal DNA profiles are compared, at each locus they may

- Match

- Partially Match

- Mismatch

To associate a person with an item of evidence, only matches will be relevant.

Partial matches, however, may suggest the involvement of a relative of the person whose profile has been compared to the evidence profile.

Debates Surrounding Forensic Genetics

DNA evidence has such potential power that it has been subjected to much scrutiny, especially over matching probability statements.

There have been commentaries in the press (Los Angeles Times: LAT Home > Articles > 2008 > July > 20 > California | Local).

There have been ill-founded commentaries online (https://www.maa.org/external_archive/devlin/devlin_10_06.html) "How big a population does it take to produce so many matches that appear to contradict so dramatically the astronomical, theoretical figures given by the naive application of the product rule? The Arizona database contained at the time a mere 65,493 entries. Scary isn't it? "

Some published discussions are shown on the next slides.

Paper by Edward Ungvarsky

Ungvarsky E. 2007. What does one in a trillion mean? GeneWatch 20:10-14.

The author is a Public Defender in Washington DC. He is worried about extreme numbers, and uses the match/partial match situation in the Arizona database to question the statistical calculations producing those numbers.

"The modified product rule [using theta] has not been subject to rigorous empirical testing." This statement ignores the work reported in the paper by Weir he cites in reference 18.

Paper by Bruce Weir

Weir BS. 2004. Matching and partially-matching DNA profiles. Journal of Forensic Sciences 49:1009-1014.

All 14,768 profiles in an Australian database were compared to each other: 109,039,528 pairs of comparisons. There were 9 loci scored for these profiles, and no 9-locus matches found. There were 7,323 pairs of five-locus matches though, and the observed numbers of 5-out-of-5 matches could be predicted well on average by the product rule, and conservatively so if $\theta = 0.01$.

"It is very likely, for example, that there are already 9-locus matches within combined U.S. offender databases."

Paper by Bruce Weir

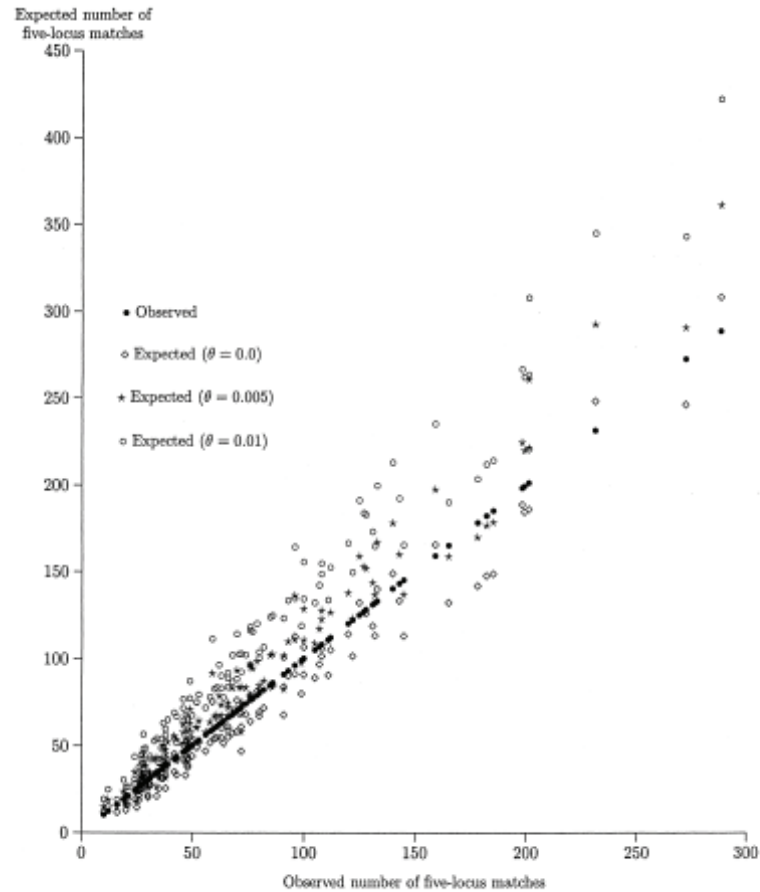Expected number of
five-locus matches

FIG. 1—*Observed and expected numbers of five-locus matches.*

There are 126 different sets of five loci, for which the number of matches ranged from 10 to 280.

5

Paper by James Curran

Curran J. 2010. Are DNA profiles as rare as we think? Or can we trust DNA statistics? Significance 7:62-66.

This discussion was designed to explain the 9-out-of-13 locus match among the 2,144,633,778 pairs of comparisons for the 65,493 profiles in the Arizona DNA database.

"These matches are predicted by statistical and population genetic theory. Rather than invalidate the models used in DNA identification, they actually strengthen them."

Paper by Laurence Mueller

Mueller LD. 2008. Can simple population genetic models reconcile partial match frequencies observed in large forensic databases¿ Journal of Genetics 87:101-108.

This paper shows results of simulations of databases of size 65,493 with varying numbers of pairs of full siblings. Some results are shown on the next page. The paper makes a strong case for access by independent scientists to offender databases in order to conduct numerical experiments on the data.
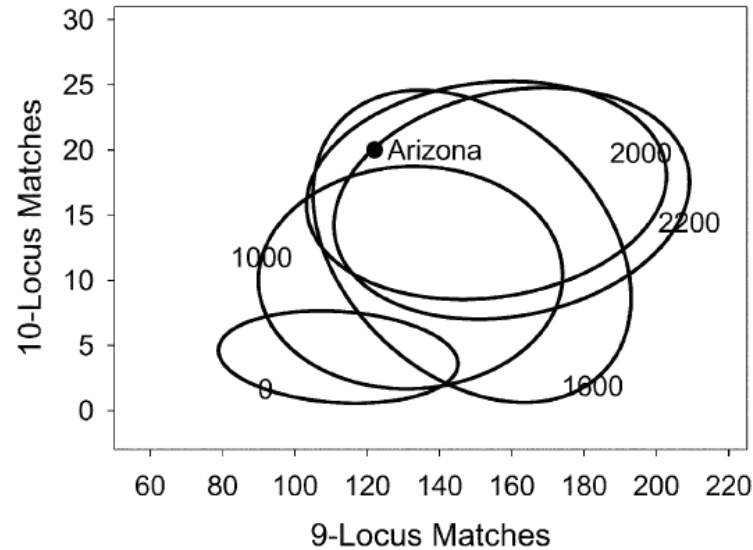
Paper by Laurence Mueller

Figure 8. 95% confidence ellipsoids for simulations in which θ was set to 0.015 and the number of full sibs varied. The number on each ellipsoid corresponds to the number of pairs of sibs present in the simulated databases.

A good fit between simulations and the actual database is found if the proportion of sibling pairs in the database is between about 3% and 9%.

Paper by David Kaye

Kaye DH. 2009. Trawling DNA databases for partial matches: What is the FBI afraid of? Cornell Journal of Law and Public Policy 19:145-171.

The author is a Law Professor with substantial knowledge of the science and application of genetic evidence. He makes a strong case for access to the CODIS database for empirical studies of matching proportions.

This paper provides a well-documented discussion of the Arizona database result, and of the legal and ethical aspects of making databases available for study.

Paper by David Kaye

The current CODIS database has over 14 million profiles (roughly 11 million from offenders, 1 million from arrestees and 1 million from evidence samples). This is about 4% of the population. It would allow about 10^{14} pairwise comparisons.
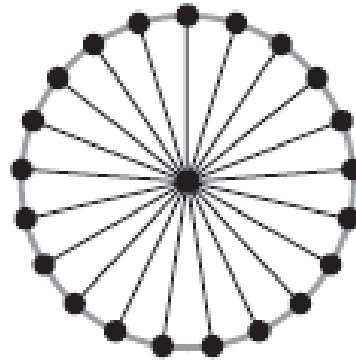
A clickable map showing numbers for each state is at http://www.fbi.gov/about-us/lab/biometric-analysis/codis/ndis-statistics

Washington State has 240,602 offender profiles, 0 arrestee profiles, and 5,522 evidence profiles in December 2014.
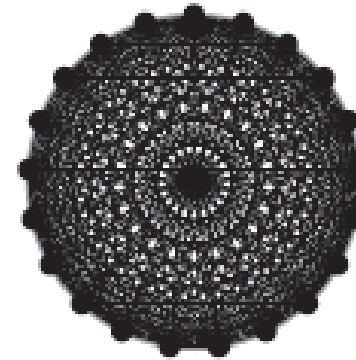
Paper by David Kaye
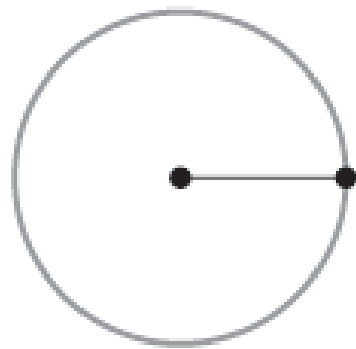
Scene to suspect: 1 pair

Scene to database: n pairs

All pairs in database: almost n^2 pairs

Figure 1. Comparisons in a confirmation case, an ordinary database trawl, and an all-pairs database trawl. Each line represents a comparison. There is 1 comparison in the first situation, n in the second, and about half of n^2 in the third.

A good illustration of the birthday problem.

DNA Identification Act of 1994: 42 USC 14131

(3) maintained by Federal, State, and local criminal justice agencies (or the Secretary of Defense in accordance with section 1565 of title 10) pursuant to rules that allow disclosure of stored DNA samples and DNA analyses only

(A) to criminal justice agencies for law enforcement identification purposes;

(B) in judicial proceedings, if otherwise admissible pursuant to applicable statutes or rules;

(C) for criminal defense purposes, to a defendant, who shall have access to samples and analyses performed in connection with the case in which such defendant is charged; or

(D) *if personally identifiable information is removed, for a population statistics database, for identification research and protocol development purposes, or for quality control purposes.*