

# For justice, share DNA databases

Giving scientists access to offenders' DNA profiles is the only way to ensure their fair use in court

Linda Geddes

WHEN a defendant's DNA appears to match DNA found at a crime scene, the probability that this is an unfortunate coincidence can be central to whether the suspect is found guilty. The assumptions used to calculate the likelihood of such a fluke – the “random match probability” – are now being questioned by a group of 41 scientists and lawyers based in the US and the UK.

These assumptions have never been independently verified on a large sample of DNA profiles, says the group. What's more, whether some RMPs are truly as vanishingly small as assumed has been called into question by recent insights into DNA databases in the US and Australia.

The group, led by Dan Krane of Wright State University in Dayton, Ohio, is demanding access to CODIS – the US national DNA database, which contains over 7 million profiles – so that they can test the assumptions behind RMPs. They have outlined their arguments in a letter, which was published in *Science* in December (vol 326, p 5960). “The national US database is a truly enormous source of data,” says signatory Larry Mueller of the University of California, Irvine (UCI).

Such research could reveal if incorrect RMPs are prompting jurors and judges to attach undue weight to DNA evidence, possibly leading to miscarriages of justice. Even if these fears are not borne out, independent checks on the DNA held in large databases like CODIS are vital to maintaining

confidence in DNA evidence presented in courts all over the world, the group says. Access would also allow the number of errors in CODIS to be measured.

DNA evidence, considered the gold standard in forensic science, is typically used in two ways: to link a known suspect to a crime, or to find new suspects – known as a “cold hit” – by searching for a match in a DNA database of known criminals.

Before a match can be sought, a profile is generated from a DNA sample by analysing specific

**“Offender databases reveal shocking differences between what you see and what you would expect”**

locations on the chromosomes, called loci, and looking at short sections of non-coding DNA, known as short tandem repeats (STRs), which vary between individuals. An RMP is then arrived at using the estimated frequencies of these STRs, or alleles, at all the loci investigated. The more loci that are analysed at once, the more comprehensive the profile and the smaller the RMP. Labs in the US typically look at 13 loci, while UK labs tend to look at 10.

One thing that researchers would like to use CODIS to verify is whether the allele frequency estimates are correct. Most of these estimates are based on data from small studies conducted during the early years of DNA forensics. But there are signs that these studies did not capture the true frequencies of certain alleles



in some populations, which could mean that the RMPs presented in court are wrong. “When you look at real offender databases you see that there are shocking differences between what you actually see and what you would expect to see,” says Krane.

The first clue that something might be amiss came in 2005, when limited data was released from the Arizona state database, a small part of CODIS. An analyst who compared every profile with every other profile in the database found that, of 65,493 profiles, 122 pairs of profiles matched at nine out of 13 loci and 20 pairs

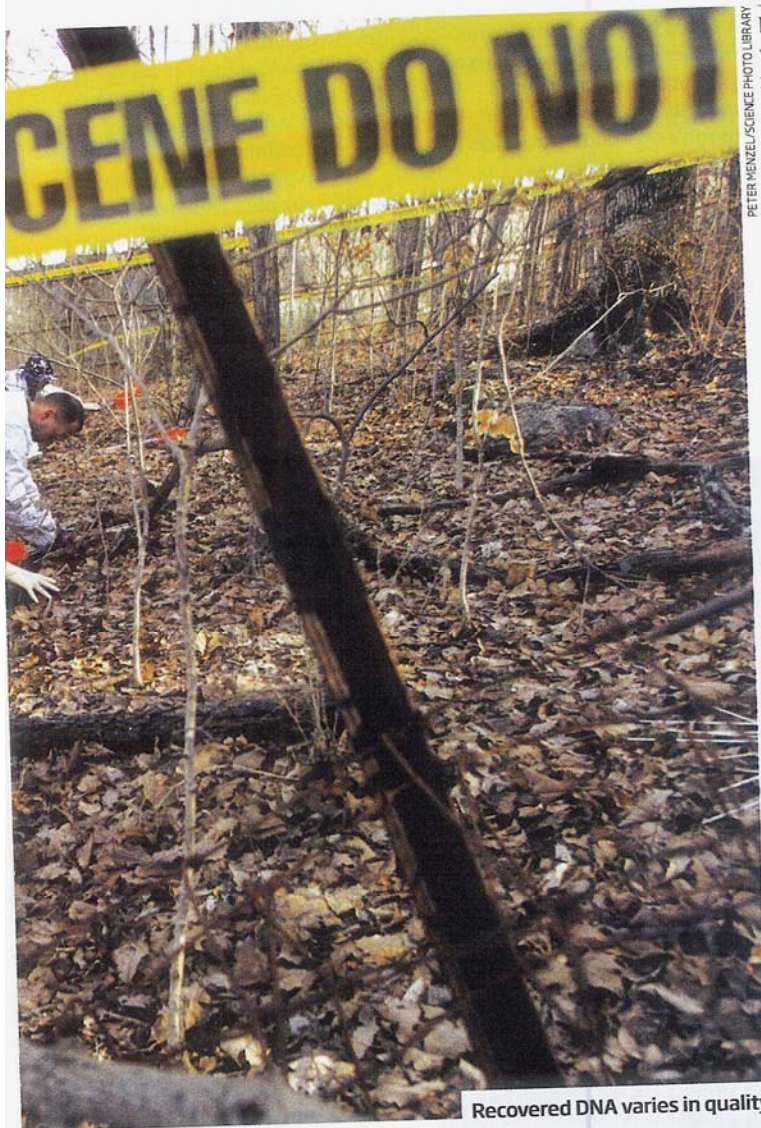
matched at 10 loci, while one pair matched at 11 loci and one more pair matched at 12 loci. “It surprised a lot of people,” says signatory Bill Thompson of UCI. “It had been common for experts to testify that a nine-locus match is tantamount to a unique identification.”

## Unexpected matches

Similar tests have since been conducted on the Illinois state database (of 220,000 profiles, 903 pairs matched at nine or more loci) and the Maryland state database (of the 30,000 profiles,

## In this section

- Crystal mountains found on moon, page 10
- Monogamy may suit men after all, page 12
- Battling the ocean's "red devils", page 14



PETER MENZEL/SCIENCE PHOTO LIBRARY

Recovered DNA varies in quality

32 pairs matched at nine loci, and three matched on all 13 loci).

One possibility is that some are duplications of the same profile in the databases – although this is not the case with the Arizona matches. Alternatively, assumptions about the frequency of alleles in populations, such as how independent these variations are of each other, might be wrong. If this is the case, access to the database is vital if these assumptions are to be corrected. “We need to learn how DNA profiles cluster by race, ethnicity and even geography,” says Krane. A third possibility is that the

surprisingly high number of matches found in these databases is the result of large numbers of relatives in the database, who are more likely to have similar DNA profiles than non-relatives. This could mean that in areas of the US and other parts of the world with more closely related populations, the RMPs may need to be tweaked.

So if CODIS provided new knowledge of the frequency of certain alleles in related or unrelated people, what would the subsequent adjustments of RMPs lead to? Even with such tweaks, in cases where all 13 loci are matched, the chances of it

being a coincidence will still be vanishingly small. But a 13-loci match is not always possible.

If only small amounts of DNA are recovered from crime scenes, or if samples are degraded or mixed with other people's DNA, the number of loci available for comparison is often much lower than 13. This means that the statistical weight attached to a match is lower – and the probability of a coincidental match higher. “I would say 5 to 10 per cent of database searches involve evidence profiles with fewer than 10 loci and/or that are mixtures,” says Mueller.

For such cases, RMPs will be much higher, so tweaks to these estimates could make a big difference to how a jury interprets them. “I’ve been involved in cases where these are 1-in-67 or 1-in-83,” says signatory Bill Shields of the State University of New York at Syracuse. “If those numbers are off by 50 per cent, then that could make a big difference to a jury.”

Bruce Budowle, former senior scientist at the FBI, which controls CODIS, argues that fears sparked by the Arizona database are overblown. Selecting a known suspect's profile and comparing it against a crime scene profile is a bit like taking a person whose birthday is 9 January and calculating the chance that a specific other person shares that birthday, which is about 1 in 365. The comparisons made within the Arizona database were the equivalent of asking how many people in a room share any birthday – a different statistic altogether. With just 23 people, for example, the probability that any two share any birthday exceeds 50 per cent. With 60 people, it is nearly 100 per cent.

The signatories insist that this “birthday problem” can't explain all the matches, however. In 2008, Mueller developed a computer model of the Arizona database that showed that the birthday problem could account for a few, but not all of the matches (*Journal of Genetics*, vol 87, p 101).

Access to DNA databases is not just about preventing potential miscarriages of justice. In 2003, when Krane was given limited access to the DNA database for the Australian state of Victoria as part of the inquest into the death of a toddler, he noticed a cluster of 32 profiles that seemed to match at 17 of the 18 alleles tested for. This was odd because far fewer matched at just 16 alleles – you would expect the opposite to be the case. Krane says the most likely cause is mistakes made

### “We need to learn how DNA profiles cluster by race, ethnicity and even geography”

when the samples were entered into the database, which he estimates may be present in as many as 1 in 1000 samples.

Access to CODIS would reveal if it contains errors, too, which could be causing investigators searching for a cold hit to miss potential suspects. “If you have mistyped an allele or a locus, then you have a person in a database whose profile would not match his own DNA,” says signatory Bicka Barlow at the San Francisco Public Defender's Office.

Will the FBI grant scientists access to CODIS? Director of the FBI Laboratory, Christian Hassell, says he appreciates the concerns the *Science* letter raises. “We are exploring ways to investigate some of the topics,” he adds. But he has turned down the request for access, citing concerns about genetic privacy.

The letter's signatories point out that medical researchers who work with DNA overcome privacy issues regularly, for example by signing an agreement promising not to divulge the data and taking certain security measures.

Without external scrutiny of the databases, doubts will remain, Mueller argues. “All of this... can be resolved by letting scientists have access to the data to do what they need to do.” ■