

How fakable is DNA evidence?

Dr Julian Huppert, University of Cambridge

Summary

DNA samples are increasingly used in forensics, and generally held to be reliable evidence, at least of presence. However, a recent paper from an Israeli team shows that it is possible to fake DNA samples. They demonstrate two things.

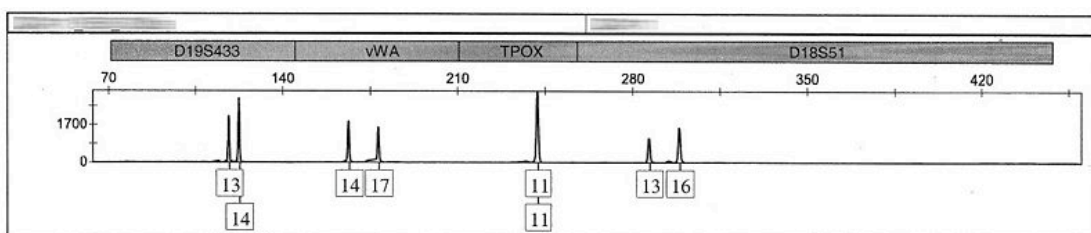
- 1) Given a DNA sample from an individual (eg from a cigarette butt) it is possible to generate a sample of another bodily material (eg blood or saliva), that will appear through standard DNA testing to be from that individual.
- 2) Given an individual's database profile ('DNA fingerprint'), it is possible to construct DNA to match that profile, and then place the DNA in a sample, which will then test positive for that individual. Access to the individual is not required.

These results firstly mean that DNA evidence can relatively easily be faked using current standard technologies, and secondly that the existence of a DNA database makes it possible to frame an individual with no previous contact. This raises very serious concerns about the forensic approach currently used in testing, the reliability of such data, and the risks of having a DNA database.

Paper reference: D. Frumkin, et al., Authentication of forensic DNA samples, *Forensic Sci. Int. Genet.* (2009), doi:10.1016/j.fsigen.2009.06.009

Introduction

When DNA evidence is collected forensically, specific regions are studied, rather than the whole sequence. These are locations where there are repeats of the same DNA sequence, and where the number of such repeats varies between humans. Since, in most cases, humans have two copies of each location (one per chromosome, inherited from each parent), this gives two numbers for each location, showing how many copies are present. So, for example, in the image below (taken from Wikipedia), only four loci are shown, and this individual has 14 copies at vWA in one chromosome, and 17 in the other. Given sufficient regions of this sort, each of which has sufficient variation in length across the population, it is possible to match two samples as probably coming from the same individual. In addition, a position is used to accurately identify the gender of the individual. This is a powerful technique, and has been increasingly used since its invention. It is frequently treated as clear evidence of the presence of an individual, if not their involvement in a crime.



Creating new samples

It is typically assumed that DNA samples are accurate, and result from genuine biological material. Frumkin et al show that this is not the case. Defined regions of DNA can be amplified using very standard techniques (typically PCR), that allow an effectively unlimited amount of material to be generated from a starting sample. Using this technique, given any source of DNA (eg a hair, a cigarette butt or dried saliva on absorbent material), a large amount of synthetic DNA can be produced, with the same sequence as the original sample. (See figure 1 in the paper – A is the original sample, B and C are artificial versions, and are indistinguishable).

To produce an artificial sample, this synthetic DNA needs to be combined with a sample from which the original DNA has been removed. This can be easily done by centrifugation, to produce (for example) red blood cells with no DNA, or cell-free saliva extract. These can then be mixed with the artificial DNA, and left at a scene. Figure 2 in the paper illustrates the results obtained using this technique; artificial DNA was placed on a gun (panels marked 1), and a bloodstain was created, which passed forensic tests for blood and was identified as belonging to the original whose DNA was amplified, rather than the individual whose blood was used (panels 3,F).

The technology to perform this is readily available. The required equipment is minimal, and would be found in any undergraduate teaching laboratory, hospital or biotech/pharmaceutical company. Some small pieces of DNA would need to be purchased, but these are readily available from many companies, and are very cheap (a few pounds each). I estimate that in my own laboratory, I could produce samples such as this in around a day, and in principle could process many samples at once. This technique would pose absolutely no problems at all for even slightly organised crime, and could be done by an aware individual.

The consequences of this are that DNA evidence cannot be taken as clear proof of presence of an individual. While it has always been possible in principle to plant any forensic evidence (eg by collecting hairs, saliva, blood etc), this would require access to these materials in sufficient volume to leave behind. However, this new technique means that any access at all to DNA samples is sufficient to generate an unlimited collection of incriminating materials that can be left at a scene.

Using a database to create fakes

The technique described above requires access to an individual's DNA at some stage. However, Frumkin and co-workers show that even this is unnecessary if there is a database available containing an individual's DNA profile. Standard techniques allow the production of a library, which contains copies of the DNA required to emulate all known variants at each of the positions tested. For the CODIS system (used in the US, and similar to the SGM+ system used in the UK), only 425 versions are needed to recreate any desired profile.

Once the library is produced, it is very simple and rapid to produce any desired quantity of DNA that will match a particular profile. This synthetic DNA can then be combined with a DNA-free biological sample as described above.

Frumkin et al demonstrate the feasibility of this technique, (figure 1D), producing a match to a known (female) individual. They also demonstrate the control available, by producing a profile identical in all respects to that individual, but is clearly identifiable as being a male (1E). They then create a saliva sample for that male pseudo-individual, which they leave on a ski mask, and is identified by standard forensic methods to be saliva from this male pseudo-individual.

This technique means that artificial samples can be produced for anyone whose profile is on a DNA database, or whose DNA profile is somehow made available. While there is a certain degree of complexity in making the library required for this, it could be performed in any reasonably equipped laboratory, and would be very accessible to any organised crime organisation. Once the library had been produced, however, selecting the correct combination of components and then amplifying them is straightforward. The library itself could be reused to produce many samples for different individuals.

Limitations

There are limitations in these techniques, and forensic approaches could be modified in order to limit the risk of such artificial approaches. Firstly, the samples themselves could be studied in greater detail, so as to detect, for example, that a salivary sample contains no cells. However, this is time consuming, and will be very hard for samples that are dry or otherwise in poor condition.

Secondly, more loci could be used for identification of a sample. The more positions that are studied, the harder it is to create materials that will pass every test. In particular, the size of the library required to fake samples from a profile alone would grow rapidly. However, this would increase the expense of the forensic tests, and I think would not make faking significantly harder. Ultimately, it might be possible to sequence the entirety of the DNA, some 3 billion bases. This would be very hard to fake from the profile alone (though not with access to the actual DNA). However, the current cost of whole-genome sequencing is around 10,000 USD per sample, so could not be used on a wide scale.

Lastly, Frumkin et al describe a method they have developed that can discriminate between natural and artificial DNA sequences. They have a company, Nucleix Ltd, that can perform tests on such samples. If this (or another approach) was used as standard in forensic assays, then sample faking in this manner could be prevented. However, this would increase significantly the cost of forensic tests, and is not currently performed.

About the author

Dr Julian Huppert is an RCUK Fellow in Computational Biology at the University of Cambridge, in the Institute of the Physics of Medicine. His research interests are in the structure and function of DNA, and so he is familiar with standard techniques for manipulating and amplifying DNA. He has no direct experience of forensic DNA work. This document should be treated as commentary on the paper by Frumkin, and is not the result of any confirmatory research. He can be contact on jlh29@cam.ac.uk or 07876 192 177.