Genetic and isotopic analysis and the UK Border Agency

Can DNA analysis really be used to screen asylum seekers by identifying their country of origin? The immigration authority apparently believed so, and almost put such a scheme into immediate action – without, it seems, consulting academic scientists on the matter. David Balding, Michael Weale, Michael Richards and Mark Thomas examine a worrying story.

The UK Border Agency is concerned about people entering Britain and seeking asylum who claim to be refugees from strife-torn countries such as Somalia, when in fact they are opportunistic economic migrants from nearby but relatively peaceful countries such as Kenya. For a number of years the Agency has been using language analysis to help confirm or refute an individual’s claim that they come from one or other side of an international border. Recently, it has turned to genetic tests and isotope analysis.

A letter to Border Agency staff dated September 11th, 2009 suggests that implementation of these technologies was then well advanced. It refers to the “Human Provenance” pilot project that was due to start three days later, using samples that were voluntarily given by asylum seekers. A 15-page “Nationality Swapping” instruction document for Agency staff gave further details, and indicated that the trial would be limited to those claiming to come from Somalia but for whom language analysis suggested that this claim was doubtful.

Many of us are familiar with TV shows and commercial services that claim to be able to trace an individual’s genes back to a village in Wales, or Denmark, or somewhere in Africa. The DNA-based ancestry testing market worldwide is worth $26 million per year (2008 estimate1), and growing, and is widely perceived by the general public and media to be based on robust science2. For example, a mitochondrial DNA (mtDNA) match between a prehistoric skeleton and a local man living near Cheddar, western England, has been accepted by some historians as proving a deep genetic continuity of the British people dating back to prehistory, thus discounting a role for subsequent large-scale migration events3. Stable isotope analysis was widely claimed to have been instrumental in identifying the mutilated torso of a boy of African origin found in the Thames in 20014. So why not harness such apparently effective technology in the service of the state to help distinguish the genuine asylum seekers from the fraudulent?

Unfortunately, there is more hype than hard science to these reports. These examples can be found in media coverage and some secondary academic literature, but not in the primary scientific literature. As we explain below, and as often discussed in the pages of Significance,
the underlying sources of uncertainty are widely underestimated.

Isotopic analysis was originally developed in archaeology to help identify immigrants or “non-locals” in a burial population. More recently it has become widely used in ecology to track movements of migratory species such as birds, and to source animal products such as ivory. The method is based on using as a signature the ratio of two isotopes of an element, such as strontium, lead, oxygen or hydrogen. As animals and humans absorb these elements from food and drink, these isotopic signatures are more or less faithfully recorded in tissue such as bone, teeth, hair and nails. The value of the isotopic signature depends on the conditions in the locality from which the food and drink were obtained. Strontium, which is one of the main elements used for isotopic analysis in humans, is derived from soil, and hence from the underlying bedrock and from rainwater. Geologies of different ages usually have different strontium isotope values, so if, for example, you consumed bread made from wheat grown in a field on an old geological substrate then the strontium isotope signature obtained from your bone or hair could be very different than if your bread was made from wheat grown on a young substrate. Oxygen and hydrogen isotope signatures are mainly derived from water and largely reflect climate, so can be used to distinguish an animal that lived in a warm environment when their tissues were forming from one that lived in a cooler environment.

On the face of it, isotopic signatures seem to provide a straightforward way of inferring the provenance of both animals and humans, especially if multiple elements are used — lead and oxygen as well as strontium, for example. Unfortunately, things are not as simple as that. The isotopic signatures of strontium and other elements can be similar over very large areas, and even between remote parts of the world. Thus it is only in unusual cases, such as very old rocks from South Africa or Canada, that we may be able to pinpoint the exact origin of an individual based on their strontium values. Indeed, the common practice in archaeology is to measure the local strontium isotope signature (by sampling modern plants and animals) and then categorise the individuals under study simply into “local” or “non-local”. It is rare to attempt to pinpoint the source of the “non-locals”. Oxygen and hydrogen isotope signatures cannot distinguish between broadly similar climates. For example, Ireland, western France and Portugal all have similar rainfall oxygen isotope values. We also lack the detailed large-scale isotopic maps of different areas of the world that would be needed for proper inference of recent geographical origin.

Another important issue is the length of time that it takes for a change of geographical location to be recorded in body tissues. The UK Border Agency proposal is to take samples of hair and nails from asylum seekers for isotopic analysis. These tissues have a relatively rapid turnover in the human body, and at best will record the isotope signatures of food and drink consumed during the few months before arriving in the UK. Information from further back in time would require bone biopsies, an invasive and damaging procedure that would be ethically untenable. Isotope signatures would therefore seem of limited value as an indicator of an asylum seeker’s real country of origin.

What then of DNA analysis, the second scientific strand apparently proposed by the Border Agency? Unlike isotopes, the genotypic variants we carry today are unchanged from the day we were born. They are also, with few alterations, the same variants that we inherited from our parents, and they from theirs. This of course is the basis for genetic ancestry testing; but it also poses problems for the inference of recent origins. Someone may have been born in one country — say, Somalia — and be a bona fide citizen of it, yet if their parents or their recent ancestors were immigrants — say, from Kenya — then they will carry the genetic signatures of that more distant ancestry, not of their current nationality.

The evidence points to the history of the human race being a history of migration. The ancestors of modern non-Africans began to migrate from Africa to other parts of the world perhaps between 60 000 and 100 000 years ago, and we humans have been on the move ever since, both within and outside Africa. Indeed, recent studies of ancient DNA and radiocarbon dates suggest episodes of dramatic human migration activity over the last 10 000 years. This has led to patterns of spatial variation that are remarkable for their uniformity, not their diversity, relative to other primate species. Like the pattern seen in isotopes, the frequencies of genetic variants typically vary only gradually with distance, making precise localisation difficult. There can, however, be pockets of more substantial variation, due to small population size and genetic isolation, or recent long-distance migration.

As with isotopes, one would need a very extensive sampling programme to ascertain the subtleties of the spatial pattern of genetic variation. To do this properly would require not only a sample from each country but also appropriate within-country sampling to account for regional variation as well as religious, cultural and linguistic groupings. Currently, there is no sufficiently comprehensive sample repository of this type. The requirement for informed consent can hinder systematic sampling, and the available studies typically rely on convenience samples. Furthermore, many of the geographically specific DNA collections that are available to academic researchers are unlikely to be available for use by the Border Agency or any branch of government, because
this would not have been covered by the informed consent governing the use of those collections.

Even without any study of genetic or demographic history, it is widely recognized that international borders, and particularly those in Africa, often do not respect genetic boundaries, nor have they acted as effective barriers to migration. Many semi-nomadic or pastoral groups straddle modern borders and move with seasons and rainfall from one side to the other and back. Indeed, the separation of culturally and sometimes genetically similar groups by an international border has often contributed to tension and conflict. Even distinct cultures either side of a border provide no guarantee of genetic distinctness: both theory and observation tell us that a relatively small level of migration, perhaps accompanying trade or wars, can suffice to generate relative genetic homogeneity between culturally distinct neighbouring groups.

Genetic data does have one advantage over isotopic data: while there are only a handful of elements available for isotope analysis, there are millions of variable sites in the human genome. If a large proportion of these were analysed, this would help to average out the uncertainties introduced at a single locus by "genetic drift": the random walk in the frequency of a genetic variant over space and time that is brought about by the stochastic nature of reproduction. From the documents they released, it appears that at least initially the Agency proposed to use only mtDNA and Y chromosome markers. The former trace only the maternal lineage (your mother, your mother’s mother, your mother’s mother’s mother and so on); the latter trace only the paternal line. To appreciate the loss of information arising from this, note that your paternal and maternal lineages include neither your mother’s father nor your father’s mother, and at every generation further back in time the proportion of ancestors in that generation who are included in these lineages approximately halves. You have up to 16 great-grandparents, but these two techniques between them will provide information from just two.

At a single locus, migration and genetic drift means that the same mtDNA or Y chromosome types might be found at appreciable frequencies in Norway and India, in Mongolia and Turkey, in Africa and Yorkshire. Furthermore, the frequency distributions within these ranges are "lumpy", with interspersed regions of high frequency. For example, a Y chromosome haplotype that is common in North Wales, but rare or absent in the rest of the UK and indeed most of Europe, occurs at appreciable frequency in the eastern Mediterranean. Could this have resulted from a male migrant from the Mediterranean who arrived in North Wales and pleased at least one and perhaps several of the local ladies, and thereby passed on many of his lady-pleasing genes to subsequent generations? Careful use of these genetic systems can allow researchers to make inferences about past demographic events at the population level, but, for the reasons outlined above, their use in most forms of individual ancestry testing can be considered as little more than genetic astrology. Neither a Y chromosome test nor an mtDNA test will allow us to say with confidence that somebody came from, for example, Somalia instead of Kenya.

Recent studies have shown the potential for genome-wide data to be more precise than mtDNA or Y chromosome data in predicting

Figure 1. (a) Predicted locations for each of 1,387 European individuals based on genotypes at over half a million DNA sites. The codes indicate the stated country of origin of the individuals’ four grandparents (individuals with grandparents of mixed national origin or of non-European origin were excluded; the country of the individual’s birth was used if the grandparental information was unavailable). The filled circles use the same colour scheme and indicate the locations used to train the assignment method. (b) Distribution of prediction accuracy by country. Reproduced by permission from Macmillan Publishers Ltd: Novembre, J. et al., Genes mirror geography within Europe, Nature, 456, 98–101. (Copyright 2008)
geographic origin.\textsuperscript{13–15} These studies have used multivariate statistical methods, typically principal components analysis, to map European individuals to their country of origin based on their DNA data at hundreds of thousands of genetic markers. Most individuals can be assigned to within several hundred kilometres of the country where they were actually sampled, or which they gave as the origin of their immediate ancestors (see Figure 1). These impressive results point to the potential of modern genetic systems to infer relatedness of individuals, and hence population structure, which in turn is indicative of demographic history. However, it is important to note that these studies were not based on random samples of individuals but on convenience samples, and some of the samples were pre-filtered to remove individuals of mixed ancestry. While genome-wide genetic data does hold much promise for identifying genetic relationships among groups, the underlying problems remain concerning incomplete sampling and how to deal with individuals resulting from recent migrations and mixed ancestry. Even complete knowledge of an individual’s genes cannot tell you with certainty his or her country of birth.

Why did the UK Border Agency apparently fail to grasp these points? We believe that intense public and media interest has led to a distorted view of the field of human genetic history. DNA-based ancestry testing companies and organisations make claims about individual ancestries, or indeed the nature of human ancestry in general, that are misleading and misrepresent the underlying uncertainties. By telling somebody that they “come from” some particular ethnic, racial or national group, they also tend to reinforce the notion that these categories have long-term genetic validity, when in fact each of us has a Celtic princess or a Viking warlord, or a Zulu chief.

A worrying feature of the Border Agency’s proposals was the apparent lack of engagement with the scientific research community. From the letter referred to above it seems clear that implementation of genetic and isotope analysis was imminent, without any indication of trials or mention of underpinning research. We are aware of no university researchers who were consulted, and none have subsequently come forward.

The bad science of DNA ancestry testing has gone far beyond unverifiable stories that a client is descended from a Celtic princess.

Thanks to the recognition of the dangers by the journal Science\textsuperscript{16}, the Home Office has moved quickly to distance itself from the implications of its letter, saying that while this trial is being undertaken, “no decisions on individual cases will be made using these techniques, and they will not be used for evidential purposes”, and that “Only after review by the Forensic Science Regulator will the techniques be considered for use in asylum investigations”\textsuperscript{11}.

Of course, isotope and DNA analysis can convey information about an individual’s origin. The problem is whether it is possible to evaluate this information in a way that is useful to Border Agency staff and fair to asylum seekers. In view of the problems outlined above, we do not think this is achievable now or in the near future. Although the draft guidelines stated that case managers “must not rely solely on the isotope and DNA test results” and that “The results from the language analysis must be afforded appropriate weight in relation to the isotope and DNA results”, no guidance was given as to what was “appropriate” – nor does it seem apparent to us what guidance could usefully be given.

References


David J. Balding is at University College London Genetics Institute. Michael Weale is at the Department of Medical and Molecular Genetics, Kings College London. Michael Richards is at the Department of Anthropology, University of British Columbia and at the Max Planck Institute for Evolutionary Anthropology, Leipzig; and Mark G. Thomas is at the Evolutionary Biology Centre, Uppsala University, and the Research Department of Genetics, Evolution and Environment at University College London.