In the blood: the myth and reality of genetic markers of identity

Mark A. Jobling[1], Rita Rasteiro[1,2] and Jon H. Wetton[1,2]

[1] Department of Genetics, University of Leicester, Leicester, UK

[2] School of History, University of Leicester, Leicester, UK

Mark A. Jobling: maj4@le.ac.uk

Abstract

Each of the 14 billion copies of the human genome in the world's population is unique, and forensic scientists use this fact to identify individuals. The differences between genomes are small (about 0.1% on average), but are not randomly distributed, and cluster in different indigenous groups of people (populations). So, despite the fact that small inter-population distances do not support a classical biological definition of races in humans, statistical methods are nonetheless claimed to be able to predict the population of origin of a DNA sample with considerable success. Such methods are employed in the genetic ancestry tests offered to the public, and particular genetic signatures, often in the male-specific Y-chromosome or maternally-inherited mitochondrial DNA, have become widely identified with particular existing or ancestral groups, such as Vikings, Jews, or Zulus, or even specific ancestral individuals. Here, we provide a primer on genetics and genomics, and describe the way in which genetic markers have become associated with particular groups. We describe the conflict between population genetics and individual-based genetics and the pitfalls of the over-simplistic interpretation of genetic data, arguing that although the tests themselves are reliable and scientific, the interpretations are unreliable and strongly influenced by cultural and other social forces.

Introduction

We begin this introduction by introducing ourselves: we are not sociologists, or political scientists, or social anthropologists, but geneticists, and therefore not part of the usual constituency of writers for this journal. We work in an interdisciplinary programme that attempts to marry genetics with history, archaeology and linguistics, so we are constantly learning that in different fields, the same words can have very different flavours. An example appears in this journal's title. No human biology or genetics journal these days would have 'race' on its masthead. But of course, this is not to say that human genetics has stayed out of the issue altogether; indeed, it has contributed much to the cultural construction of 'race', both before and since the dawn of the genomic era, marked in 2001 by the publication of the human genome sequence (International Human Genome Sequencing Consortium 2001). "Genetics is, for all its bells and whistles, simply another handmaiden recruited to bolster an eighteenth-century European world view", says Anne Morning (Morning 2014). Racists continue to use genetic data to support their ideologies (Kemp 2013), and DNA tests are increasingly used to link individuals to some past homeland or group, more or less real or imaginary.

In this paper we address the issue of markers of identity from a genetic perspective. We provide a primer on the technology and terminology of genetics and genomics, and then focus on the practice and pitfalls of genetic ancestry testing in its 'recreational' sense. We do not address the important and related area of the use of genetic ancestry testing in medical contexts, which has been widely discussed elsewhere (Duster 2003; Fullwiley 2007, 2008; Royal et al. 2010; Fujimura and Rajagopalan 2011). We argue that genetic methods based on the study of groups of individuals (populations), are reliable and respectable scientific tools, but that the practice of individual genetic ancestry testing is unreliable and powerfully influenced by cultural and other social forces. These include the prevailing bureaucratic classifications of race, popular obsessions with past historical groups such as the Vikings, and fascination with particular powerful male historical figures, such as Genghis Khan, who are reputed to have outfathered their peers.

Depending on how you look at them, human genomes are either all the same, or all different. After all, they are all human, and can be easily distinguished from the genomes of all other existing species, including our cousins the chimpanzees, with whom we share a mere 98.8% DNA sequence similarity (Chimpanzee Sequencing and Analysis Consortium 2005). The introduction to the publication of the human genome sequence in *Nature* (Dennis, Gallagher, and Campbell 2001) was headed by a quotation from the Universal Declaration on the Human Genome and Human Rights: "The human genome underlies the fundamental unity of all members of the human family". And yet each human genome is unique, differing by an average of about 0.1% (1000 Genomes Project Consortium 2012). This is the fact exploited by forensic geneticists when they target highly variable parts of the genome, and thus distinguish between individuals with a very high degree of certainty.

It is somewhere between these two levels – the unity of the species and the uniqueness of the individual – that the trouble lies. Because the variation among human genomes is non-randomly distributed in geographic space, this has led to the linking of particular genetic variants to particular ethnic groups, and the identification of genetic signatures as markers of identity for individuals or for groups of people. This is the topic we explore here.

Technological advances in DNA sequencing and the analysis of DNA variation have driven the recent rise of genomics in medicine, and the same technologies have been adopted by commercial testing companies which offer tests for genetic ancestry or deep family relationships to their customers. This activity is sometimes characterised as 'recreational genomics', although for some consumers with questions about their ancestry there may be more riding on a result than this label suggests. Our colleagues Scully et al. (this volume) distinguish between 'high-stakes' and 'low-stakes' clients of genetic ancestry testing companies. The former (for example, those seeking Native American roots) may be interested in potential benefits and rights as well as personal histories, while the latter (such as those fascinated by possible Viking ancestry) may be more light-hearted in their engagement.

Whatever their motivation, armed with only a credit card and an internet connection, anyone in the world can now have sophisticated analysis carried out

on their own genome via a tube of saliva. Many of these non-specialists post and share their data online, and some even participate in publications in scientific journals as 'citizen scientists' (Rocca et al. 2012). Thus the evolution of ideas about genetic markers of identity has a momentum of its own, driven by public and commercial interest, and largely unshackled by academic research: an exciting and sometimes alarming state of affairs.

A brief tour of the genome

Before going on, we set out salient background information about the human genome and how its components are inherited. The genome is made of deoxyribonucleic acid (DNA) – a long molecule built from units known as bases, or nucleotides. There are four types of bases (A, G, C and T), and their order along the DNA molecule, in the 2% or so of the genome that corresponds to our 25,000 genes, specifies the proteins that are made in our cells. DNA is usually double-stranded, with one strand wrapped around the other in the iconic double helix, and bases paired together in the core; A always with T, and G with C. One copy of the genome comprises about 3,200 million base pairs in total, and in almost all cells there are two genome copies, one inherited from the mother, and one from the father. The genome is not a single long DNA molecule, however, but is divided into 23 segments, the chromosomes, which lie in the nucleus of the cell. Of these, one pair (the sex chromosomes) determines sex and is inherited in a different manner to the rest (the autosomes): a female inherits an X chromosome from each of her parents, while a male always inherits an X from his mother and a sex-determining Y chromosome from his father. In addition to the chromosomes there is another small but essential piece of genetic information that all humans carry: mitochondrial DNA (mtDNA), which lies in the mitochondria, small energy-producing structures outside the nucleus of the cell. Because of events at fertilisation, when the egg contributes many mitochondria to the developing child but the sperm contributes none, mtDNA is inherited by male and female children from their mother only.

During the making of eggs and sperm, the 22 pairs of autosomal chromosomes swap segments in a reciprocal process known as recombination. For most of the genome the scrambling effect of recombination complicates the

passage of DNA through time. Thus, it does not make sense to ask from which one of your ancestors, many generations ago, your chromosome number 3 (for example) comes – it descends from multiple ancestors, as an unpredictable patchwork of segments. By contrast, mtDNA and most of the Y chromosome escape the recombination process, being passed intact down 'uniparental' lineages from ancestral foremothers and forefathers respectively. This simplifies the interpretation of mtDNA and Y lineages in populations, and has led to a particularly strong focus on these as markers for specific populations or groups that we explore below.

DNA variation exists on many scales and in many different forms, but for the purposes of this article we concentrate on its simplest form, the substitution of one base for another (for example, a C for a T). Such variants are known as single nucleotide polymorphisms, or SNPs for short. Because most of the genome appears to have no particular function (Graur et al. 2013), most SNPs are regarded as selectively neutral, in other words having no influence on observable traits or diseases. Some that we mention below, however, have a strong influence via effects on genes.

Distribution of genetic variation within and between populations

Our genomes pass down to us from our ancestors, and along the way accumulate variation through mutation. This leads to the 0.1% average difference between individual genomes, but also to patterns of diversity that differ between human groups. We refer to these as 'populations' - sets of individuals defined according to some shared characteristic that may be social, cultural, historical or geographical. Examples we might come across in the genetic literature are Caucasians, Europeans, Germans, Indo-European-speakers, or Ashkenazi Jews - the labels are heterogeneous, and of course it is possible to belong to more than one group at the same time. Reading diversity patterns can allow us to interpret aspects of the histories of such groups, and thus genetics can be considered alongside traditional disciplines such as history, archaeology and historical linguistics as a method for studying the human past.

Studies of global samples show that most DNA variation in the autosomes (83–88%, depending on the sampling and methods) exists *within* populations,

rather than between them (Jobling et al. 2013) p. 329). When populations are grouped into their continental affiliations, which can be considered as proxies for traditional racial classifications, only 9–13% of variation exists between these groups. These findings have been taken to support the idea that the concept of human races has no biological basis (Barbujani et al. 1997), thus bolstering the position that has been characterised as 'race pragmatism' (Nelson 2008), in which race is a social construct, reflecting a power relationship rather than any inherent difference.

The geographical pattern of genetic diversity

You do not need to be a geneticist to realise that genetic variation is influenced by geography. The simplest models of population genetics assume panmixia – every human can mate with every other. But even in an era of internet dating and cheap air travel, any given person is most likely to marry and have children with people in their neighbourhood. So, for human populations 'isolation by distance' is a better model, in which individuals tend to move short distances and only rarely longer distances, implying that the closer two populations are geographically, the more similar they tend to be genetically. This model explains much of the general pattern of human genetic variation and its largely clinal (continuous) pattern.

Two additional factors influence genetic variation. The first is large-scale and ancient – the echo of the Out of Africa migrations that began about 50,000 years ago. The highest level of modern autosomal genetic diversity remains in Africa, and in indigenous populations around the world, diversity reduces steadily with over-land distance from Africa (Ramachandran et al. 2005), reflecting a 'serial founder effect' as populations expanded into Eurasia, Australia and eventually the New World. The second factor is the relative isolation of some regions by geographical features (seas or mountain ranges), and of some populations by cultural features (religion, language or marriage customs) that reduce the exchange of migrants with their neighbours.

Given the description above, it is not surprising that genetic analysis of non-continuously sampled human populations reveals a clustered pattern of diversity. This was shown in an influential paper of 2002 that applied a

clustering method (implemented in the computer program STRUCTURE (Pritchard, Stephens, and Donnelly 2000)) to autosome-wide data from more than 1000 individuals from 52 different human populations (Rosenberg et al. 2002). When the method was asked to find five genetic clusters, these corresponded to 'major geographic regions' (Africa, Europe plus the Middle East and South/Central Asia, East Asia, the Americas, and Oceania). This finding, despite the way the samples were chosen, the uncertainty over the 'best' number of clusters, and the slightness of the differences observed (Bolnick 2008), has been enthusiastically adopted by some as evidence to support the biological reality of races (Wade 2014). Such people can be characterised as race 'naturalists' (Hacking 2005), who argue that nature creates real differences between groups of individuals that science can then identify.

Subsequent studies of populations whose ancestry ostensibly originates from one place only ('non-admixed' populations) have shown that whole-genome SNP information matches self-defined ancestry at the level of European nations (Novembre et al. 2008), subregions (Elhaik et al. 2014) and even islands within the Orkney archipelago (O'Dushlaine et al. 2010).

Ancestry testing and the reification of race

One consequence of the work described above has been the development of individual commercial ancestry tests that analyse the DNA of customers, and provide estimates of the proportion of their genomes that belong to various genetic clusters. Some versions of these tests use 'ancestry informative markers' (AIMs) (Shriver et al. 1997), small sets of genetic markers chosen because they show relatively large frequency differences between continental population samples. For example, AncestryByDNA's 'DNA Origins' test uses 144 SNPs to apportion a genome into just four 'ancestral groups' – European, Indigenous American, East Asian, and sub-Saharan African. The limited number of markers, and the false impression created of four biologically discrete groups of ancestral humans, have been critiqued before (Bolnick et al. 2007).

Both the number of DNA markers and the number of reference populations used in these tests have been increasing. In the test offered by 23AndMe (the '23' referring to the number of pairs of chromosomes), between

500,000 and 1 million SNPs across the autosomal genome are analysed – these are not AIMs, but an essentially unbiased set of DNA variants. The customer's SNP data are then compared with similar data from 31 populations, and partitioned into the major clusters. Ancestry estimates from these tests are always at the mercy of the quality and size of available databases (and hence, of the company carrying out the tests). Initially these were small and patchy, but 23AndMe has expanded its own databases by using the self-reported ancestry of some of its >750,000 (mostly US-based) customers to populate them, as well as published scientific information.

A recent academic study (Bryc et al. 2015) emerging from the same company compares the genetic ancestry of US customers who describe themselves as African Americans (5269), Latinos (8663), and European Americans (148,789). The authors report a positive correlation between self-identification prior to genetic testing, and the eventual proportions of African, European and Native American ancestry that the tests diagnose. However, customers who were shown by the tests to have less than 28% of African ancestry self-describe as European American; this contrasts with the 'one-drop' rule, a social mechanism and legal principle relegating admixed individuals to a subordinate group by designating 'black' anyone who had *any* known African ancestry (Daniel 2001). Interviews with African-American 'root-seekers' (Nelson 2008) have shown that they deploy genetic information not as proof of identity but as just one thread in a complex tapestry of ancestry in the context of their 'genetic aspirations'. It remains to be seen how far the results of genetic tests offered by companies such as 23AndMe will lead to changes in self-definition.

Such tests have the potential to contribute to the reification of race (Bolnick 2008; Shriver and Kittles 2008) as a biological phenomenon. Survey experiments by social psychologists (Phelan et al. 2014) show that such reification does indeed occur in the minds of many consumers, as a side-effect of the way that ancestry testing emphasises a genetic basis for differences between populations. Furthermore, in popular science literature (Wade 2014) the identification of genetic clusters has been explicitly taken as unequivocal biological support for the existence of major human races of the kinds proposed by the taxonomist Linnaeus and the anthropologist Blumenbach in the

eighteenth century, and which foreshadowed the classifications of 19th-century racial theory.

Genetic tests can also provide information about the proportion of a customer's genome descending from some unexpected ancestors – Neanderthals. Fossil evidence shows that anatomically modern humans and Neanderthals coexisted in close proximity in Europe and the Middle East for long periods of time, and it was much debated whether or not the two groups interbred. Advances in DNA sequencing technology allowed the genomes of ~45,000-year-old Neanderthals to be sequenced (Green et al. 2010), and with the sequence data in hand, it became possible to ask if parts of the genomes of modern humans could be traced back to them. Non-Africans have 1–4% Neanderthal ancestry, and commercial tests (such as 23AndMe's) now provide a customer's own Neanderthal ancestry proportion. Svante Pääbo, the pioneer of ancient DNA who led the Neanderthal genome sequencing, notes that when the scientific findings were published, 45 men (but only two women) wrote to him stating a belief that they were fully or partly Neanderthal, and offering their DNA for study. In addition, 12 women got in touch with Pääbo declaring their husbands to be Neanderthals, but only two men said the same of their wives. This reflects the popular but inaccurate stereotypes of dim-witted Neanderthals, and thus reinforces the idea that cultural forces strongly shape beliefs about DNA and ancestry. In fact, archaeological and palaeontological research shows that Neanderthals had larger brains than we do, and that they buried their dead, made tools, wore jewellery, and created art (Appenzeller 2013),

Individual genes as markers of identity

The discussion above describes the use of large numbers of genome-wide SNPs that are individually uninteresting and lacking functional consequences, but collectively deliver information about descent from many different ancestors to provide diffuse markers of identity. However, some individual SNPs do have functional consequences, and are the focus of genetic tests. Examples include variants causing red hair, or blue eyes, or ear-wax type (Yoshiura et al. 2006) ('wet', as found in most Africans and Europeans, or 'dry', as in most East Asians and Native Americans). The perspicacious reader may be thinking that a mirror

and a cotton-wool bud are perfectly adequate tools for diagnosing these things, and that a sophisticated gene test seems unnecessary. Perhaps for someone who self-identifies as 'Celtic' the red hair gene test gives additional legitimation and reinforcement over and above what the mirror can provide.

More interesting, perhaps, is a test for the 'warrior gene' offered by FamilyTreeDNA. The enzyme monoamine oxidase A (MAOA) degrades a subset of neurotransmitters including serotonin, epinephrine, and norepinephrine - molecules that transmit information from one neuron to another. Adjacent to the *MAOA* gene is a region of DNA that controls how much enzyme is produced, and a common variant of the length of this region (called 3R) leads to reduced production of enzyme compared to other common versions (Sabol, Hu, and Hamer 1998). The gene lies on the X chromosome, so males, who have only one X, show the simplest relationship between the version of the gene they carry and its behavioural consequences. Men carrying the 3R version (the 'warrior gene') are more likely to respond aggressively to maltreatment or stress (Caspi et al. 2002). Despite charging almost a hundred dollars for the 'warrior gene' test, the testing company calls the association between gene variant and behaviour a 'factoid', and best used as a 'cocktail conversation starter'. Nonetheless we might wonder if the results of the test have any influence on the behaviour of people who are tested; the possible influence of the 3R variant was used in 2009 as part of a successful criminal defence in the USA (Brooks-Crozier 2011), and made the difference between 32 years' imprisonment and the death penalty.

Mitochondrial DNA and the Y chromosome trace matrilines and patrilines

As outlined above, there are parts of the genome that, unlike the autosomes and the X chromosome, are not inherited from both parents, and that escape the reshuffling process of recombination. These are mtDNA and the Y chromosome, together constituting just 2% of the genome, but passing down maternal and paternal lines and diversifying only through the accumulation of mutations.

Mitochondria have an extraordinary history, originating as bacteria that were taken up by a single-celled precursor of all modern multicellular organisms about 1.5 billion years ago. They were provided a secure symbiotic haven within the larger cells, and in return contributed biochemical machinery to produce

energy. Most of the DNA of these ancient bacteria has been lost, but a vestigial bacterial genome remains. In humans, this is a circle of 16,500 DNA base pairs, packed with genes that are needed in cellular energy production, and passed by mothers to all of their offspring. Determining all or part of this mtDNA genome sequence reveals SNPs. For a number of reasons, including the proximity of the energy-generating system and its chemical by-products, mtDNA undergoes more frequent mutation than the bulk of the genome, and this leads to high sequence diversity and many SNPs.

The Y chromosome forms the paternal counterpart to mtDNA, differing from it in that, while all of us carry mtDNA, only males carry the Y chromosome because of its sex-determining role. It is also a much larger piece of DNA, about 60 million base pairs in size, potentially providing much more scope for DNA variation to occur. Two segments of the Y chromosome do undergo recombination with the X chromosome, and therefore have the complex histories seen in autosomes. These segments are ignored in genealogical and population studies. The non-recombining majority is often called the male-specific region of the Y, or MSY.

Because of the lack of recombination, SNPs can be used in a simple way to build a genealogical tree of mtDNA or MSY sequences (Underhill and Kivisild 2007), rooted by comparison with the sequence of our closest living animal relative, the chimpanzee. If we follow the mtDNA tree from the tips of its branches towards its root, the number of mtDNA types reduces as branches join together. Eventually we reach a point where all modern human mtDNA sequences coalesce to a single ancestral maternal lineage. This sequence really existed in an ancestral woman, who, through Biblical analogy, has been called ‘mitochondrial Eve’. Unlike the Biblical character, Eve was not the only woman around, but lived among a lot of other would-be Eves. The modern mtDNA sequences that join the tree closest to its root are all sub-Saharan African, so Eve was African, and this supports the idea that the origin of modern humans was in Africa. Notably, given the fact that autosomal DNA shows a small proportion of Neanderthal ancestry, there is no trace among modern humans of Neanderthal mtDNA, which, if it was contributed in the first place, has been lost. There are close parallels with mtDNA in the construction and features of the phylogenetic

tree constructed from MSY sequences: again SNPs define different types that can be arranged in a simple branching tree. Again the root lies in sub-Saharan Africa (in a long-dead man predictably dubbed 'Y-chromosomal Adam'), and no trace of Neanderthal Y chromosomes is found among modern humans.

Daughters of Eve, sons of Adam

Mitochondrial DNA types are more differentiated geographically than most autosomal DNA sequences, with particular branches of the mtDNA tree ('haplogroups') tending to be concentrated in particular parts of the world. Haplogroups are named after letters of the alphabet, and this led Bryan Sykes to take the seven major European mtDNA haplogroups (H, J, K, T, U, V, X) and label their respective ancestors the 'seven daughters of Eve' – Helena, Jasmine, Katrine, Tara, Ursula, Velda and Xenia. This work resulted in a popular science book (Sykes 2001) that included florid imaginings of the lives of these ancestral women, and a commercial DNA testing company, Oxford Ancestors, which uses mtDNA testing to connect customers to their 'clan mothers'.

The geographical differentiation of MSY sequences is even higher than that of mtDNA, and in fact the highest of any part of the genome. This is partly due to 'genetic drift' – some men have many sons, while some have none at all, and this can lead to shifts in the local frequencies of MSY haplogroups. Another factor is due to marriage practices: most societies are traditionally patrilocal (Burton et al. 1996), where men are more likely to stay close to their birthplaces upon marriage than women, leading to local accumulations of particular MSY types. Haplogroups in the MSY tree are also known by letters of the alphabet (Y Chromosome Consortium 2002), and names with biblical or ethnic flavours have been given to 'clan-fathers', including Abel, Cain, Jahangir, Krishna, Quetzalcoatl, Ruslan and Seth (Oppenheimer 2003). Happily, given the ethnic stereotyping and oversimplification of history that these names imply, they have not caught on.

Maternal and paternal markers of identity in diasporas

Mitochondrial DNA testing has been particularly focused on US citizens connected to the African diaspora. This is partly due to the nature of sexual interactions in the age of colonialism and slavery. African Americans are more

likely to retain genetic signals of their African ancestors in maternally inherited mtDNA and the X chromosome (inherited twice as often from mothers compared to fathers), than in autosomes, or the Y chromosome, which frequently originates from European men (Lind et al. 2007). This has been exploited in DNA ancestry tests aiming to restore old ties, or forge new ones, between African-Americans and Africans (Nelson 2008). The genetic testing company African Ancestry has focused on this issue, using mtDNA testing (again, limited by the size and comprehensiveness of its databases) to claim that the individual African tribes could be identified from which the maternal ancestors were taken as slaves, and publicising their tests with the help of African-American celebrities. On this basis Oprah Winfrey declared herself a Zulu, Chris Rock discovered descent from the Udeme people of northern Cameroon, and Forest Whitaker ended up being made an honorary chief of Igboland in Nigeria on the basis of his mtDNA sequence. The certainty associated with these claims is at odds with the small size and limited geographical coverage of the comparative databases that support them (Greely 2008).

Another diaspora, that of the Jewish people, is the focus of a company called iGENEA, which on its website asks potential clients 'Are you Jewish?'; this, on the face of it, is a question about religion and culture, but iGENEA instead offers a mtDNA test. Jews are an example of a 'transnational isolate' – a population that is geographically widespread, but through religion and customs maintains a degree of genetic distinctiveness that is detectable at the population level with autosomal SNPs, mtDNA or Y chromosomes (Ostrer and Skorecki 2013). A 2014 UK television series, *Dead Famous DNA*, undertook analysis of DNA claimed to originate from a number of dead celebrities including Elvis Presley, Marilyn Monroe and Charles Darwin. These odd bedfellows were joined by Eva Braun - hair from a brush allegedly belonging to her was tested, and its mtDNA claimed by the presenter (Hodgson 2014) to be 'a specific sequence that's only shared by Ashkenazi Jews'. The sequence apparently belongs to haplogroup N1b1 (Sherwin 2014), which is indeed associated with Ashkenazi Jews, but far from exclusively, and at a frequency of only ~10% (Behar et al. 2008). So the DNA information, if we are to believe it, is inconclusive, and at the individual level provides ambiguous information at best.

Specific MSY types have also been attached to Jewish populations, in particular the so-called Cohanim lineage, originally defined by a study of Jewish priests (Thomas et al. 1998) in whom it was found at high frequency. Subsequent higher-resolution studies show a small number of well-defined lineages (including an 'extended Cohanim' lineage) present in Ashkenazi and non-Ashkenazi Cohanim, dated to approximately 3–4 KYA (thousand years ago) based on accumulated DNA variation within the lineages and an estimate of mutation rate (Hammer et al. 2009). The 'Are you Jewish?' service offered by iGENEA include tests for these lineages. The Lemba people of South Africa and Zimbabwe, who claim Jewish origins based on a long-standing oral history and a number of apparently Judaic customs, were shown to carry the Cohanim lineage at a frequency of about 9% (Thomas et al. 2000). This was widely reported to corroborate the Lemba's origins story (Vickers 2010), but in their case the presence of this lineage has not been sufficient to validate a claim to be Jewish (Shimona 2003 p. 178). Subsequent higher-resolution studies of Y-chromosomal lineages in the Lemba (Soodyall 2013) have been interpreted as consistent with an origin in the Middle East, but the absence of the 'extended' Cohanim lineage taken as evidence against Jewish ancestry. This case serves to show that interpretations of markers of genetic identity are not fixed: they can shift with time, and they depend on who the carriers are.

Oxford Ancestors has also offered 'Tribes of Britain' tests, while 'Britains DNA' (sic. – the absence of an apostrophe is deliberate but mystifying) assigns individual MSY types to origins including 'Pictish', 'Teutonic', 'Hebridean Viking', and 'Anglo-Saxon'. The evidence on which these links are made is unpublished and so cannot be assessed, but is probably based on datasets in which particular Y-chromosome types are present at relatively high frequencies in particular geographical regions today. These then become 'homelands' for the Y-chromosome types (Bowes 2014), and a link is made to a past population or tribe identified in written sources and applied uncritically to genetic groupings, sometimes supported by attempts to estimate the ages of lineages. In truth, we have very little idea of who the Picts were, and certainly no idea whatever about the kinds of Y chromosomes they carried.

The problem of multiple ancestors

Unlike the autosomal ancestry tests, mtDNA and Y-chromosome tests, whether they link an individual to clan-mother Helena or to the Zulus or the Ashkenazim, focus on only one ancestor out of many. As an example, if Oprah Winfrey's African maternal ancestor was enslaved in 1600, then some 14 generations (assuming ~30 years per generation (Fenner 2005)) have elapsed since then. The number of Oprah's ancestors doubles each generation into the past, so in principle she had more than 16,000 ancestors living in 1600. It seems arbitrary to focus on only one of them. The only possible justification might be that Oprah's mtDNA ancestor is connected to her via an uninterrupted chain of mother-daughter relationships, and thence the shared life experience of femaleness. For MSY testing, the analogous patrilineal link may have additional resonance in patriarchal cultures where, as if often the case, property or status is transmitted down patrilines.

For autosomal testing, the problem is reversed – in geographical terms, the test provides an average location based on multiple ancestors. For example, an 'admixed' person whose parents could trace back their ancestry for many generations respectively in Scotland and Denmark would be placed in the middle of the North Sea. The company Prosapia Genetics sells a Geographic Population Structure (GPS) service (Prosapia Genetics 2014), which claims that it can take autosomal DNA data and trace an ancestral link from a customer's 'current home' to their 'genetic home', down to the level of an individual village. Yet for most customers, a claim that all of their ancestors can be traced back to one small place cannot be taken seriously – the 'genetic home' is an imaginary one.

Genetic identity through links to past individuals

As well as links to intangible ancestors and groups, mtDNA and the Y chromosome are sometimes employed to make connections to specific individuals in the past. An influential though academically unpublished example was the claim from Bryan Sykes that the mtDNA sequence of the 9000-year-old skeleton from Cheddar Gorge in south-west England, and known as Cheddar Man, matched that of Adrian Targett, a history teacher living in a town just a few miles away (Arthur 1997). Aside from the question of whether the sequence retrieved

from the bones, discovered in 1903, was really that of Cheddar Man himself rather than some subsequent museum curator, the lack of detail on the sequence itself makes it hard to judge the significance of the match. If, as Wikipedia claims, it belongs to the type known as U5, then given a European frequency of 9% (Richards et al. 2000), thousands of people in Britain are expected to share the same sequence today. The Cheddar Man saga is significant because it has been used to contribute to an idea of indigeneity or continuity. This idea has been used by the far right in their attempts to define who should be considered 'indigenous' British (Kemp 2013) - in the words of Nick Griffin, former leader of the British National Party, 'people who have been here overwhelmingly for the last 17,000 years. We are the aborigines here' (Anstead 2009).

Companies offer DNA tests that claim to link customers to past celebrities including Marie Antoinette of France (Jehaes et al. 2001), Tsar Nicholas II (Gill et al. 1994), and the outlaw Jesse James (Stone, Starrs, and Stoneking 2001). Two founders of common Y-chromosome lineages are also listed among 'Famous DNA' services. These are Genghis Khan, whose alleged Asian patrilineal descendants carry some 0.5% of all Y chromosomes worldwide (Zerjal et al. 2003), and the warlord Niall of the Nine Hostages, claimed to have lived in fifth- or sixth-century Ireland, whose descendants have been said to constitute one in five of men in northwestern Ireland (Moore et al. 2006). Men undertaking these tests appear keen to identify with a fecund warlord of the past; certainly Genghis Khan is of great importance in Mongolia, as illustrated by the fact that, when ordered to adopt surnames by the government, a large proportion chose his clan-name, Borjigin (Magnier 2004). But, while men whose Y chromosomes match the 'Khan' and 'Uí Néill' types certainly descend from two prolific male ancestors, the identities of these ancestors remain unclear without the extraction and analysis of ancient DNA from attested remains.

A link between Y chromosomes and patrilineal surnames

One aspect of the Y chromosome that sets it aside from mtDNA is its link with surnames. Most European surnames, and many surnames worldwide (King and Jobling 2009a), are patrilineal, so form a cultural counterpart to the biological inheritance of the Y. This relationship has been explored most thoroughly in

academic studies for surnames of the British Isles, with groups of men sharing surnames in Ireland (McEvoy and Bradley 2006) and England (King and Jobling 2009b) being shown to have a much higher than average probability of sharing Y chromosome types. In rare English names such as Attenborough (which has around 1000 carriers) about 90% of apparently unrelated men share a Y type that is present at only about 1% in the general population. In other words, the Attenborough surname is essentially equivalent to a genetic marker of identity. For more common names the structure of variation is more complex, reflecting multiple founders for names, or past female marital infidelity. The company FamilyTreeDNA has been at the forefront of exploiting the surname-Y link, running 'surname projects' from its website in which men sharing surnames can enrol and have their DNA tested. In this way large numbers of men can identify with a group in which connections exist via distant paternal ancestors; the genetic link may be an adjunct to traditional paper-based genealogical research, but can also indicate a genealogical connection for which no other evidence exists.

From population genetics to individual genetics

Academic geneticists who are interested in making deductions about the human past almost invariably study populations, rather than individuals. The statistical tools of classical population genetics, developed in the 1930s, are suited to understanding the frequencies of genetic types within groups of people. Likewise, the more recent approaches that use complex mathematical models to describe and analyse genealogies and demography are applied to populations of DNA sequences (Currat and Excoffier 2005). These studies are built around testing hypotheses, and their conclusions come with measures of uncertainty attached to them, or p-values that indicate the strength of statistical support, given the sample size.

By contrast, the non-academics who think about genetic diversity, and sometimes participate in genetic testing themselves, are naturally interested in the individual. When we recruit participants for our studies, our aim is to publish papers based on population data, but we also promise individual-based information in return for participation. We struggle to balance the need for

rigour in the information we provide to an individual about their ancestry against the possibility that it is too disappointingly vague or equivocal to be of any interest at all. Genetic testing companies are, of course, focused on individuals as their customers, and do not generally have such qualms. This has led to frequent over-interpretation of individual-based data, including press reports of some fantastical conclusions; for example, one, in the UK's *Daily Telegraph* newspaper (Cramb 2012), told of a man tested by 'Britains DNA' (sic) who had been informed that he was 'the grandfather of everyone in Britain'. This presentation of nonsensical non-science does not seem helpful to the customer.

The discussion above contrasts academics and non-academics, equating the latter with commercial testing companies. However, the reality is more messy and interesting. Most commercial testing companies, including Britains DNA, 23AndMe, African Ancestry and FamilyTreeDNA, have academic scientists as founders or advisory board members (Bolnick et al. 2007), and shareholders or beneficiaries. The evidence of these companies' websites suggests that some of the scientists manage the balancing act between the commercial individual-focused test result and the academic population-focused research paper better than others.

The degree of general shared ancestry is not widely appreciated. We are all related to each other: a genome-wide analysis of SNPs in an apparently unrelated set of people with French origins (International HapMap Consortium 2007) found average sharing of about 0.34% of their genomes by descent from ancestors in common – equivalent to the expected average for a pair of third cousins. And even when we consider the markedly non-random mating habits of our fellow humans, a simulation-based study suggests that a common ancestor in the family trees of all 7 billion humans on the planet lived only 3400 years ago (Rohde, Olson, and Chang 2004).

The privileging of particular ancestors among the many may be natural enough, but it has no scientific support, and the claims made by some testing companies about the ancestry of individuals has been called 'genetic astrology' (Thomas 2013). In an effort to counter endemic over interpretation, the organisation Sense about Science has published a guide to genetic ancestry testing (Sense about Science 2013). It points out that tests may give the

impression that a customer's results are unique and provide information about their personal history, but (like the predictions based on star-signs) the same history could as well apply to many other people. In addition, a result from a personal DNA test could be matched with ancestral stories other than the one that the testing company provides.

The future

What is the future of genetic ancestry testing? Science and technology seem certain to continue their forward march. Genetic tests will become cheaper and more sophisticated, and genetic databases larger and with better geographical coverage. Ancient DNA studies will contribute the genomes of transported slaves, Vikings, Anglo-Saxons, and Romans, and perhaps even Genghis Khan himself. There will be more fascinating and surprising findings reported in academic journals, underpinned by the unromantic business of statistic analysis and hypothesis testing.

However, it requires considerable optimism to believe that the practice of individual genetic ancestry testing will make similar progress. The tests offered by companies are inherently unreliable, and far from being objective, are driven by established cultural preconceptions. Testing reifies race in the minds of many consumers as a biological phenomenon, and supplies apparent links with existing or extinct populations, or with dead or mythical celebrities, with unwarranted certainty. Thus testing feeds notions of indigeneity and group membership.

The companies that apply genetic methods in individual testing are failing their customers in a number of ways: they fail to explain exactly what the tests are, and how they work; they fail to describe the databases with which they compare their customers' results; they fail to give any measure of the robustness of their conclusions; and in tests of uniparentally inherited lineages, they fail to explain how small a proportion of a customer's ancestry their test reflects. Customers are increasingly scientifically literate, as internet sites such as that of the International Society of Genetic Genealogy (www.isogg.org) demonstrate, so the argument cannot be made that the science is too complicated to explain. Formal regulation seems too heavy handed an approach, but an ethical code

(Shriver and Kittles 2008) committing testing companies to clear explanation of their tests and conclusions would go a long way to increasing confidence in what they do. The presentation of nonsense in the guise of science is not only misleading the paying customer, but runs the risk of damaging science itself.

Acknowledgements

All three authors acknowledge support from the Leverhulme Trust, grant number F/00 212/AM. MAJ was also supported by a Wellcome Trust Senior Fellowship, grant number 087576.

References

1000 Genomes Project Consortium. 2012. "An integrated map of genetic variation from 1,092 human genomes." *Nature* 491 (7422): 56-65. doi: 10.1038/nature11632.

Anstead, N. "Transcript: BBC Question Time, 22 October 2009." http://www.scribd.com/doc/34711539/BBCQT-TV-Transcript.

Appenzeller, T. 2013. "Neanderthal culture: old masters." *Nature* 497 (7449): 302-4.

Arthur, C. 1997. "The family link that reaches back 300 generations to a Cheddar cave." In *The Independent*. London: Independent Print Limited (Saturday 08 March 1997).

Barbujani, G., A. Magagni, E. Minch, and L. L. Cavalli-Sforza. 1997. "An apportionment of human DNA diversity." *Proc. Natl. Acad. Sci. USA* 94 (9): 4516-9.

Behar, D. M., E. Metspalu, T. Kivisild, S. Rosset, S. Tzur, Y. Hadid, G. Yudkovsky, et al. 2008. "Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora." *PLoS ONE* 3 (4): e2062. doi: 10.1371/journal.pone.0002062.

Bolnick, D. A. 2008. "Individual ancestry inference and the reification of race as a biological phenomenon." In *Revisiting Race in a Genomic Age*, edited by B.A. Koenig, S.-J.L. Lee and S.S. Richardson, 70-88. New Brunswick: Rutgers University Press.

Bolnick, D. A., D. Fullwiley, T. Duster, R. S. Cooper, J. H. Fujimura, J. Kahn, J. S. Kaufman, et al. 2007. "Genetics. The science and business of genetic ancestry testing." *Science* 318 (5849): 399-400. doi: 10.1126/science.1150098.

Bowes, T. 2014. "How to find your genetic homeland." In *Family Tree (February 2014)*, 18-22.

Brooks-Crozier, J. 2011. "The nature and nurture of violence: Early intervention services for the families of MAOA-low children as a means to reduce violent crime and the costs of violent crime." *Connecticut Law Rev* 44 (2): 531-72.

Bryc, K., E. Y. Durand, J. M. Macpherson, D. Reich, and J. L. Mountain. 2015. "The genetic ancestry of African Americans, Latinos, and European Americans across the United States." *Am J Hum Genet* 96 (1): 37-53. doi: 10.1016/j.ajhg.2014.11.010.

Burton, M.L., C.C. Moore, J.W.M. Whiting, A.K. Romney, D.F. Aberle, J.A. Barcelo, M.M. Dow, et al. 1996. "Regions based on social structure." *Curr. Anthropol.* 37: 87-123.

Caspi, A., J. McClay, T. E. Moffitt, J. Mill, J. Martin, I. W. Craig, A. Taylor, and R. Poulton. 2002. "Role of genotype in the cycle of violence in maltreated children." *Science* 297 (5582): 851-4. doi: 10.1126/science.1072290.

Chimpanzee Sequencing and Analysis Consortium. 2005. "Initial sequence of the chimpanzee genome and comparison with the human genome." *Nature* 437 (7055): 69-87. doi: 10.1038/nature04072.

Cramb, A. 2012. "Scottish lecturer found to be 'grandfather of everyone in Britain'." In *Daily Telegraph*. London: Telegraph Media Group (29 Jun 2012).
Currat, M., and L. Excoffier. 2005. "The effect of the Neolithic expansion on European molecular diversity." *Proc. Biol. Sci.* 272 (1564): 679-88.
Daniel, G.R. 2001. *More than Black? Multiracial Identity and the New Racial Order* Philadelphia; p.3: Temple University Press.
Dennis, C., R. Gallagher, and P. Campbell. 2001. "Everyone's genome." *Nature* 409 (6822): 813. doi: 10.1038/35057264.
Duster, T. 2003. *Backdoor to Eugenics* New York: Routledge.
Elhaik, E., T. Tatarinova, D. Chebotarev, I. S. Piras, C. Maria Calo, A. De Montis, M. Atzori, et al. 2014. "Geographic population structure analysis of worldwide human populations infers their biogeographical origins." *Nat Commun* 5: 3513. doi: 10.1038/ncomms4513.
Fenner, J. N. 2005. "Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies." *Am. J. Phys. Anthropol.* 128 (2): 415-23.
Fujimura, J. H., and R. Rajagopalan. 2011. "Different differences: the use of 'genetic ancestry' versus race in biomedical human genetic research." *Soc Stud Sci* 41 (1): 5-30.
Fullwiley, D. 2007. "The molecularization of race: institutionalizing human difference in pharmacogenetics practice." *Science as Culture* 16 (1): 1–30.
———. 2008. "The biologistical construction of race: 'admixture' technology and the new genetic medicine." *Soc Stud Sci* 38 (5): 695-735.
Gill, P., P. L. Ivanov, C. Kimpton, R. Piercy, N. Benson, G. Tully, I. Evett, E. Hagelberg, and K. Sullivan. 1994. "Identification of the remains of the Romanov family by DNA analysis." *Nat. Genet.* 6: 130-5.
Graur, D., Y. Zheng, N. Price, R. B. Azevedo, R. A. Zufall, and E. Elhaik. 2013. "On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE." *Genome Biol Evol* 5 (3): 578-90. doi: 10.1093/gbe/evt028.
Greely, H.T. 2008. "Genetic genealogy: genetics meets the marketplace." In *Revisiting Race in a Genomic Age*, edited by B.A. Koenig, S.-J.L. Lee and S.S. Richardson, 215-34. New Brunswick: Rutgers University Press.
Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, et al. 2010. "A draft sequence of the Neandertal genome." *Science* 328 (5979): 710-22. doi: 10.1126/science.1188021.
Hacking, I. 2005. "Why Race Still Matters." *Daedalus* 134 (1): 102–16.
Hammer, M. F., D. M. Behar, T. M. Karafet, F. L. Mendez, B. Hallmark, T. Erez, L. A. Zhivotovsky, S. Rosset, and K. Skorecki. 2009. "Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood." *Hum Genet* 126 (5): 707-17. doi: 10.1007/s00439-009-0727-5.
Hodgson, C. 2014. "Dead Famous DNA: 5 things we've learnt from the show, featuring Elvis, Napoleon and Adolf Hitler's wife Eva Braun." In *Daily Mirror*. London (9 April).
International HapMap Consortium. 2007. "A second generation human haplotype map of over 3.1 million SNPs." *Nature* 449 (7164): 851-61. doi: nature06258 [pii]

10.1038/nature06258.
International Human Genome Sequencing Consortium. 2001. "Initial sequencing and analysis of the human genome." *Nature* 409 (6822): 860-921.
Jehaes, E., H. Pfeiffer, K. Toprak, R. Decorte, B. Brinkmann, and J. J. Cassiman. 2001. "Mitochondrial DNA analysis of the putative heart of Louis XVII, son of Louis XVI and Marie-Antoinette." *Eur J Hum Genet* 9 (3): 185-90. doi: 10.1038/sj.ejhg.5200602.
Jobling, M. A., E. J. Hollox, M. E. Hurles, T. Kivisild, and C. Tyler-Smith. 2013. *Human Evolutionary Genetics* New York and London: Garland Science.
Kemp, A. 2013. *Four Flags: The indigenous people of Great Britain: DNA, History and the Right to Existence of the Native Inhabitants of the British Isles*: CreateSpace Independent Publishing Platform.
King, T. E., and M. A. Jobling. 2009a. "What's in a name? Y chromosomes, surnames and the genetic genealogy revolution." *Trends Genet* 25 (8): 351-60. doi: 10.1016/j.tig.2009.06.003.
King, T.E., and M. A. Jobling. 2009b. "Founders, drift and infidelity: the relationship between Y chromosome diversity and patrilineal surnames." *Mol. Biol. Evol.* 26: 1093-102. doi: 10.1093/molbev/msp022.
Lind, J. M., H. B. Hutcheson-Dilks, S. M. Williams, J. H. Moore, M. Essex, E. Ruiz-Pesini, D. C. Wallace, S. A. Tishkoff, S. J. O'Brien, and M. W. Smith. 2007. "Elevated male European and female African contributions to the genomes of African American individuals." *Hum Genet* 120 (5): 713-22. doi: 10.1007/s00439-006-0261-7.
Magnier, M. 2004. "Identity Issues in Mongolia." In *Los Angeles Times*. Los Angeles, CA: Tribune Publishing.
McEvoy, B., and D. G. Bradley. 2006. "Y-chromosomes and the extent of patrilineal ancestry in Irish surnames." *Hum. Genet.* 119 (1-2): 212-9.
Moore, L. T., B. McEvoy, E. Cape, K. Simms, and D. G. Bradley. 2006. "A Y-chromosome signature of hegemony in Gaelic ireland." *Am. J. Hum. Genet.* 78 (2): 334-8.
Morning, A. 2014. "And you thought we had moved beyond all that: biological race returns to the social sciences." *Ethnic Racial Stud.* 37 (10): 1676–85. doi: 10.1080/01419870.2014.931992.
Nelson, A. 2008. "Bio science: genetic genealogy testing and the pursuit of African ancestry." *Soc Stud Sci* 38 (5): 759-83.
Novembre, J., T. Johnson, K. Bryc, Z. Kutalik, A. R. Boyko, A. Auton, A. Indap, et al. 2008. "Genes mirror geography within Europe." *Nature* 456 (7218): 98-101. doi: 10.1038/nature07331.
O'Dushlaine, C., R. McQuillan, M. E. Weale, D. J. Crouch, A. Johansson, Y. Aulchenko, C. S. Franklin, et al. 2010. "Genes predict village of origin in rural Europe." *Eur J Hum Genet* 18 (11): 1269-70. doi: 10.1038/ejhg.2010.92.
Oppenheimer, S. 2003. *Out of Eden: the peopling of the world* London: Constable.
Ostrer, H., and K. Skorecki. 2013. "The population genetics of the Jewish people." *Hum Genet* 132 (2): 119-27. doi: 10.1007/s00439-012-1235-6.
Phelan, J.C., B.G. Link, S. Zelner, and L.H. Yang. 2014. "Direct-to-consumer racial admixture tests and beliefs about essential racial differences." *Soc Psychol Q* 77 (3): 296–318.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. "Inference of population structure using multilocus genotype data." *Genetics* 155: 945-59.
Prosapia Genetics. "Prosapia Genetics." https://www.prosapiagenetics.com/index.php.
Ramachandran, S., O. Deshpande, C. C. Roseman, N. A. Rosenberg, M. W. Feldman, and L. L. Cavalli-Sforza. 2005. "Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa." *Proc. Natl. Acad. Sci. USA* 102 (44): 15942-7.
Richards, M., V. Macaulay, E. Hickey, E. Vega, B. Sykes, V. Guida, C. Rengo, et al. 2000. "Tracing European founder lineages in the near eastern mtDNA pool." *Am. J. Hum. Genet.* 67 (5): 1251-76.
Rocca, R. A., G. Magoon, D. F. Reynolds, T. Krahn, V. O. Tilroe, P. M. Op den Velde Boots, and A. J. Grierson. 2012. "Discovery of Western European R1b1a2 Y chromosome variants in 1000 genomes project data: an online community approach." *PLoS ONE* 7 (7): e41634. doi: 10.1371/journal.pone.0041634.
Rohde, D. L., S. Olson, and J. T. Chang. 2004. "Modelling the recent common ancestry of all living humans." *Nature* 431 (7008): 562-6. doi: 10.1038/nature02842.
Rosenberg, N.A., J.K. Pritchard, J.L. Weber, H. M. Cann, K.K. Kidd, L.A. Zhivotovsky, and M. W. Feldman. 2002. "Genetic structure of human populations." *Science* 298: 2381-5.
Royal, C. D., J. Novembre, S. M. Fullerton, D. B. Goldstein, J. C. Long, M. J. Bamshad, and A. G. Clark. 2010. "Inferring genetic ancestry: opportunities, challenges, and implications." *Am J Hum Genet* 86 (5): 661-73. doi: 10.1016/j.ajhg.2010.03.011.
Sabol, S. Z., S. Hu, and D. Hamer. 1998. "A functional polymorphism in the monoamine oxidase A gene promoter." *Hum Genet* 103 (3): 273-9.
Sense about Science. 2013. "Sense about genetic ancestry testing (http://www.senseaboutscience.org/resources.php/119/sense-about-genetic-ancestry-testing)."
Sherwin, A. 2014. "Did Adolf Hitler marry a Jewish woman? DNA tests 'show Eva Braun associated with Ashkenazi Jews'." In *Independent*. London.
Shimona, G. 2003. *Community and conscience: the Jews in apartheid South Africa* Lebanon, NH, USA: Brandeis University Press.
Shriver, M. D., and R. A. Kittles. 2008. "Genetic ancestry and the search for personalized genetic histories." In *Revisiting Race in a Genomic Age*, edited by B.A. Koenig, S.-J.L. Lee and S.S. Richardson, 201-14. New Brunswick: Rutgers University Press.
Shriver, M.D., M.W. Smith, L. Jin, A. Marcini, J.M. Akey, R. Deka, and R.E. Ferrell. 1997. "Ethnic-affiliation estimation by use of population-specific DNA markers." *Am. J. Hum. Genet.* 60: 957-64.
Soodyall, H. 2013. "Lemba origins revisited: tracing the ancestry of Y chromosomes in South African and Zimbabwean Lemba." *S Afr Med J* 103 (12 Suppl 1): 1009-13. doi: 10.7196/samj.7297.
Stone, A. C., J. E. Starrs, and M. Stoneking. 2001. "Mitochondrial DNA analysis of the presumptive remains of Jesse James." *J Forensic Sci* 46 (1): 173-6.
Sykes, B. 2001. *The Seven Daughters of Eve* London: Bantam Press.

Thomas, M. G., T. Parfitt, D. A. Weiss, K. Skorecki, J. F. Wilson, M. le Roux, N. Bradman, and D. B. Goldstein. 2000. "Y chromosomes traveling south: the cohen modal haplotype and the origins of the Lemba--the "Black Jews of Southern Africa"." *Am J Hum Genet* 66 (2): 674-86. doi: 10.1086/302749.

Thomas, M.G. 2013. "To claim someone has 'Viking ancestors' is no better than astrology." In *The Guardian*. London: Guardian News and Media.

Thomas, M.G., K. Skorecki, H. Ben-Ami, T. Parfitt, N. Bradman, and D.B. Goldstein. 1998. "Origins of Old Testament priests." *Nature* 394: 138-40.

Underhill, P. A., and T. Kivisild. 2007. "Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations." *Annu. Rev. Genet.* 41: 539-64. doi: 10.1146/annurev.genet.41.110306.130407.

Vickers, S. "BBC News Report: Lost Jewish tribe 'found in Zimbabwe'." http://news.bbc.co.uk/1/hi/8550614.stm.

Wade, N. 2014. *A Troublesome Inheritance: Genes, Race and Human History* New York: The Penguin Press.

Y Chromosome Consortium. 2002. "A nomenclature system for the tree of human Y-chromosomal binary haplogroups." *Genome Res.* 12: 339-48.

Yoshiura, K., A. Kinoshita, T. Ishida, A. Ninokata, T. Ishikawa, T. Kaname, M. Bannai, et al. 2006. "A SNP in the ABCC11 gene is the determinant of human earwax type." *Nat. Genet.* 38 (3): 324-30. doi: 10.1038/ng1733.

Zerjal, T., Y. Xue, G. Bertorelle, R.S. Wells, W. Bao, S. Zhu, R. Qamar, et al. 2003. "The genetic legacy of the Mongols." *Am. J. Hum. Genet.* 72: 717-21.