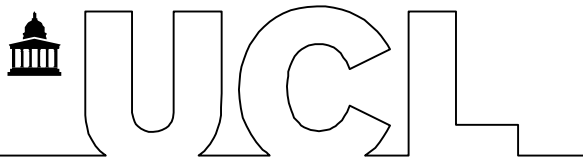
David Balding Professor of Statistical Genetics, Genetics Institute

Mark Thomas Professor of Evolutionary Genetics, Dept of Genetics, Evolution & Environment

Faculty of Life Sciences



[redacted]

Dear Mr Moffat,

cc by e-mail to Dr Jim Wilson and Dr Gianpiero Cavalleri

We write to you concerning your interview on the BBC Radio 4 'Today' Program on the 9th July 2012. We are scientists who are engaged in research on inferring human population history and evolution. We think that your interview grossly misrepresented our research field, containing many false assertions about what can be inferred from genetic testing about an individual's ancestry.

We outline below some of the errors made during your interview:

- "*It has been posited by scientists for about 20 years that Adam and Eve really existed. They may never have met, but they really existed*".

  According to the Bible, Adam and Eve were the first man and woman and they clearly did meet. There can be no scientific evidence for their existence. Probably you are referring to the most recent common ancestor (MRCA) at the mtDNA and at (the NPAR of) the Y chromosome. If so you should have indicated this. It is then obvious that there must be an Adam and an Eve: it is spurious to mention any time such as 20 years for scientists "positing" what is simply a matter of definition, and misleading to assert that they "really existed" as if that were a claim with any content – it misleads people to think you are confirming the existence of the Biblical Adam and Eve. Moreover there is nothing special about those individuals to entitle them to be called "Adam" and "Eve". Humans are expected to have many thousands of different MRCAs across the genome (see e.g. Li, Durbin, *Inference of human population history from individual whole-genome sequences* Nature, 2011, 475(7357): 493—496 for an example of genome-wide TMRCA inference from a pair of haploid genomes). Because of the larger effective population size (Ne), the times since autosomal and X MRCAs are expected to be greater than those at mtDNA and Y, but because there are so many of them, and because of the high variability in the time depth of coalescence processes, they have a very wide range. At some genomic loci we expect to have MRCAs just a few thousand years ago, and at other loci our MRCAs will have lived millions of years ago. We don't expect there to be anything special about the mtDNA and Y MRCAs, and because they are just two among many thousands of MRCAs, those two convey little about an individual's ancestry.

- "*What happened was a kind of genetic bottleneck – around 70 thousand BC an Indonesian volcano called mount Toba blew itself to smithereens and it was almost a species extinction event and so all the other lineages apart from these two, apart from Adam and Eve were destroyed*".

  While the Toba eruption may be uncontroversial, its effects on human population size and on patterns of genetic variation certainly is. More importantly, no matter how devastating its effects were, there is no reason to believe that this event caused all lineages apart from

those of “Adam and Eve” to be destroyed and indeed it is highly unlikely that this was the case.

- “*And what happened with the Britain’s DNA project is that about 4 or 5 weeks ago we discovered a remarkable individual, a Mr Ian Kinnaird from Caithness and he has Eve’s DNA – he’s only two removed from Eve … he carries a marker called L1b which is only two mutations different from what Eve’s marker must have been … he’s Eve’s grandson …*”.

  Mr Kinnaird is not “Eve’s grandson” and he does not carry mtDNA Eve’s DNA any more than any other person alive. Mr Kinnaird may well carry an L1b mtDNA, as do many Africans, but everybody alive today is chronologically equidistant from ‘mtDNA Eve’. Your claim that “*… is only two mutations different from what Eve’s marker must have been”* is highly surprising. Behar et al 2012 (Amer J Hum Genet 90, 675–684) inferred the root haplotype for human mtDNA and then looked at the distribution of mutation differences from that root haplotype to nearly 20,000 contemporary individuals. They found a range of differences from 41 to 77, with a mean of 57.1. It therefore seems extraordinary that an individual only 2 mutation differences from the inferred root haplotype could exist. Are you confusing lineages with mutations? It would appear from the mtDNA phylogeny of those authors (Fig 2) that the MRCA of haplogroup L1b is separated from the global MRCA by two coalescence nodes. If that is what you meant then saying “two mutations different” is wrong.

- “*..but we found Sheban DNA – a marker called HV – which we didn’t expect to find, and I say we’ve got nine people who …*”.

  There is no scientific justification for claiming that haplogroup HV corresponds to “Sheban DNA”, or that the Queen of Sheba carried this haplogroup, or that people who carry one of the many HV lineages are descended from the Q of S. We know very little about the Sheban people, and we don’t know whether the Q of S had a different ancestry from the majority of the population of that region. Moreover, we know little about the migrations that have occurred in and around that region since her time, and therefore we cannot know the relationship between current genetic types in that region and the genetic types that prevailed 3K years ago. Haplogroup HV is today widely dispersed in the Middle East, the Caucasus and other regions. It is unsurprising that some British people carry the haplogroup, and this does not imply any connection with the Red Sea region let along one specific individual who is thought to have lived there 3K years ago. Incidentally, HV, like L1b and other haplogroups can legitimately be called lineages or haplogroups, but it is not usual to refer to them as “markers”.

- “*Men with the name Cohen, 97%, a huge number, share the same marker*…”.

  One of us (MGT) was the first author on the original paper to estimate a common ancestry date for the majority of Cohanim (Thomas MG, Skorecki K, et al (1998) Origins of Old Testament priests. Nature 394:138-140). Only around 45% of Ashkenazi Cohanim and 56% of Sephardic Cohanim carry the Cohen Modal haplotype (CMH) according to that study. Others carry Y chromosomes that are only a small number of mutation steps away from the modal haplotype and so may be considered as belonging to this cluster, but even with these ‘cluster’ Y chromosomes the figure is far from 97%. Subsequent research by others has split the high-frequency cluster in Cohanim into at least two lineages. The above paper and subsequent studies have shown that the CMH is not specific to Cohens but carried by many other individuals, both Jewish and non-Jewish.

- *…Aaron, the first high priest, and so as I say, the Bible, through the Britain's DNA project and other research is really beginning to come alive…*".

  The association of the Cohen Y MRCA with Aaron remains unproven – we can only say that it is an intriguing possibility. The estimated age of 'the Cohen lineage' is dependent on which haplotypes are assigned to it. More generally, what is the evidence for your repeated claim that through the Britain's DNA project the Bible is coming alive? We do not believe there is any.

- "*It's the marker that finds it way back into the past*".

  We suspect from this and similar statements, as well from www.britainsdna.com/demo.html, that you are using an approach that can be labelled interpretive phylogeography. While such an approach has been used in a number of papers in scientific journals, you should be aware that:
  1. The majority of population geneticists do not accept this as a legitimate scientific approach.
  2. Most interpretative phylogeographic analysis is unsystematic and so susceptible to the subjective biases of the researcher.
  3. No scientific paper has ever demonstrated that this inference approach works.
  4. One of the very few systematic incarnations of this approach (Nested Clade Phylogeographic Analysis) has been demonstrated not to work (see Neilsen and Beaumont, 2009, Molecular Ecology 18, 1034–1047, and references therein).

- *"…everyone in Britain is an immigrant the only issue is when did you arrive"*

  Because we have very many ancestral lineages, typically any individual's ancestors arrived at many different times.

- "*We have found that 33% of men are closely associated to the founding lineages of Britain …*".

  No you have not because you do not know what the "*founding lineages of Britain*" are. Nobody does and for some time this is likely to remain the case. Presumable you are referring to what Wilson et al 2001 (Genetic evidence for different male and female roles during cultural transitions in the British Isles. Proc Natl Acad Sci U S A 98:5078-5083) called the Atlantic Modal Haplotype (first identified by MGT, an author on the above paper). There is considerable controversy on the origins of this lineage, with some arguing for its spread being driven by Neolithic expansions (see Balaresque, P. et al. 2010 *A predominantly Neolithic origin for European paternal lineages*. PLoS Biol. **8**, e1000285), and others arguing for an origin in Palaeolithic Iberian refugial populations. However, both James Wilson and MGT were co-authors on a recent paper (Busby et al 2012, Proc. R. Soc. B 279, 884-892) that states: "As a consequence, the existing data and tools are insufficient to make credible estimates for the age of this *haplogroup*, and conclusions about the timing of its origin and dispersal should be viewed with a large degree of caution". This is wholly inconsistent with the claim that it is "*a founding lineage of Britain*".

- *"..we found people that have got Berber and Tuareg ancestry from the Saharan nomads"*

  You cannot know this to be true from DNA data. Perhaps you mean that you found people carrying a haplotype that is today present in the Berber and Tuareg, and no doubt also in other populations. One can then estimate the TMRCA but, because of the ubiquity of

migration, we can only make informed guesses about where that ancestor was or what ethnic label might reasonably be applied (perhaps an ancient Scottish migrant to the Sahara enjoyed reproductive success? More likely the modern-day Scots sharing a haplotype with Berber/Tuareg descend from a common ancestor in some other part of the world).

In summary, human ancestry is much more complex than you indicate, and reliable inferences from polymorphism data at one or two loci are much more limited than you claim. The numbers of our ancestors approximately doubles with each generation back into the past, reaching up towards 1,000 distinct ancestors just 10 generations ago (approximately 250 years), and many thousands – possibly hundreds of thousands – 500 years ago. Often, those ancestors will have lived in many diverse locations. Although you may be able to estimate the TMRCA at mtDNA and/or Y, you cannot say where that ancestor lived, and it implies very little about the individual's ancestry overall, which is a complex network of many lineages.

Some valid inferences can be made about the demographic history of populations using samples from multiple individuals, more so than for the ancestry of individuals. But this remains a challenging task – not least because of the stochastic natures of inheritance and mutation processes. Many scientists are engaged in human population history research, and most endeavour to cautiously apply robust statistical methods in making inferences, and to communicate an appropriate degree of uncertainty. It is damaging to them, and to science in general, if exaggerated, false and misleading claims are made in a high-profile and usually-reliable medium such as the Today program. We understand that you may have little expertise in genetics but we see that your business partners include geneticists who have published papers in this field and will appreciate the points we have made and so we feel you have no excuse for appearing so poorly-informed. You have a responsible position as Rector of St Andrew's University, and we hope you will use it in future to support and not undermine science – the real discoveries that are being made about human history are no less exciting, if more complex and uncertain, than the stories you recounted.

Many DNA-based ancestry-testing organisations make similarly exaggerated claims about what genetic tests can say of a person's ancestry. We are concerned that claims made by this industry mislead the public about the nature of human ancestry in general, by conforming to, and profiting from, out-dated and scientifically unsupported views of the human past, race and ethnic identity, instead of attempting to overturn them in a scientifically informed manner.

The public has a great appetite for stories about its ancestry from DNA data, and this provides a very strong commercial incentive to cut corners, exaggerate beyond the evidence, and ignore uncertainty. Such misleading practices are aided by the fact that, since there is very little that can be proven about an individual's ancestry from, say, mtDNA or Y polymorphism data, it is equally difficult to prove that any specific claim is false. However, we have had direct experience of disappointed customers who, once it is explained to them that the genetic ancestry results for which they have paid a substantial sum cannot be confirmed scientifically, become disillusioned with science more generally; they previously believed that scientists could be trusted.

We are further concerned that you appear to have used the medium of the BBC to promote a commercial enterprise under the guise of a scientific study. Indeed you took the opportunity to direct listeners to your website www.britainsdna.com which appears to describe a commercial service rather than a research study as the interview suggested. What are the aims and objectives of your research study? Do you have permission from an ethics committee for research on human subjects; if so please give details. You claimed that the service was ``massively subsidised" but this seems implausible given that the cost for mtDNA or Y chromosome testing is £170, which should be ample to cover all costs with a substantial profit besides. Please explain where the "massive subsidy" comes from and what it subsidises.

Yours sincerely

David Balding and Mark Thomas
Darwin Building
Gower Street London WC1E 6BT