J Biosci (2019)44:69
DOI: 10.1007/s12038-019-9890-6

Commentary

On Methodological issues in the Indo-European debate *By* Michel Danino

Keywords. Archaeogenetics; Aryan; Harappa; Harappan; Indo-Aryan; Indo-European

The prehistories of the population and languages of India are scientific questions (Majumder 2018). Unlike questions decided in mathematics or a court of law, scientific questions cannot lead to definitive proof. Rather, new evidence and advances in methods will bring a succession of improved models that account better for wider ranges of more accurately documented evidence. There will always be doubts about aspects of the prevailing model and data that it accounts for imperfectly. If this were not the case, there would be no room for scientific progress. Nonetheless, this does not mean that the prevailing theoretical model is simply wrong, though these loose ends can be selectively pulled together to build a convincing-looking case for non-scientists that it must be wrong. Our argument here is not that Michel Danino's argument lacks learning or points worthy of consideration, but that it is unscientific.

Danino covers, in the erudite and scholarly way one would expect from a heritage enthusiast, a lot of old ground. What he documents is that the subject of Indo-European and Indo-Iranian languages has been debated by some great minds for a long time without clear resolution.

But, as he says, there is now a new line of evidence available due to a dramatic advance in DNA sequencing technology: archaeogenetics. Where Danino's argument falls down is in his attempts to summarise and critique this new body of evidence. After his stress on the importance of careful, critical, scholarly and interdisciplinary research in the areas of textual analysis, linguistics and archaeology, Danino's dismissal of archaeogenetics has produced an oversimplified caricature.

In fact, many archaeogeneticists, like many linguists and archaeologists, are trying very hard to forge interdisciplinarity, and indeed trans-disciplinarity (Vander Linden 2019). The very term 'archaeogenetics' reflects this, and was originally promoted by the eminent archaeologist Colin Renfrew, who is well known for his efforts to bring together archaeology, linguistics and genetics (Renfrew 2000). This kind of endeavour is not easy, because it entails not only colleagues from each discipline working closely together, in multi-disciplinary fashion, but also for them to internalise each other's knowledge and thought processes. Instead, Danino relies on stereotypes of archaeogenetic thinking – and indeed quotations of critics from outside the discipline that are more than a decade old – without engaging with the details of the research he dismisses. Curiously, he fails to cite our research paper on the settlement of South Asia (Silva *et al.* 2017) – which he has nevertheless previously censured in the press (Danino 2017) – and, although he does mention a preprint from the Harvard group (Narasimhan *et al.* 2018), he prefers to target popular works and media gossip concerning unpublished research.

The suggestion that these studies might be politically motivated is flimsy at best. A more careful reading of David Reich's comments (Reich 2018) indicates that his motivations in this context were practical (that is, how to accommodate the preconceptions of colleagues when reporting results), rather than political. Danino's assertion that archaeogenetics assumes the existence of prehistoric migrations from the outset completely misses the mark. The reason that archaeogenetics is so powerful, and now able to contribute to debates such as this, is because it is able to test directly and explicitly whether or not migrations took place at a particular time.

For example, let us take the case of the much-maligned Y-chromosome haplogroup R1a. The reason we argue that R1a lineages are intrusive into the Indian Subcontinent is as follows: South Asian lineages nest phylogenetically within Central Asian, and ultimately Eastern European, lineages, including almost all those seen using ancient DNA analysis of Bronze Age samples – and the clusters restricted to South Asia date to 3000–4000 years ago (Silva *et al.* 2017, fig. 5). This is not circular reasoning – it is based on careful analysis of data.

http://www.ias.ac.in/jbiosci
Published online: 08 July 2019

Danino wonders why the conclusion differs now from those in the past, but he provides the answer himself – more efficient sequencing technologies allow us to obtain better data, yielding more reliable results. The age estimates are now far more robust than in the past, for two reasons: the availability of modern whole-genome sequence data allows for much more accurate and precise molecular-clock estimates; and it is now possible to recover ancient DNA, admittedly, not from South Asia, where almost none has been recovered to date, but from Central Asia and the East European Steppe, and also central and western Europe, where R1a looks likely to have dispersed alongside some Indo-European languages (Baltic, Slavic and Germanic). This new Y-chromosome evidence is strongly consistent with a respected linguistic argument – broadly based on the sound systems, grammars and vocabularies of the languages – that Indo-Iranian, Balto-Slavic and Germanic descend from a single ancestral node on the family tree of Indo-European languages (Ringe *et al.* 2002). (It is also the case that signals of Bronze Age migration are comparatively minor on the mitochondrial lineage, on which earlier studies predominantly focused, so that inferences of a mainly indigenous origin were then correct – but only for the female line of descent. We will return to this sex-bias pattern later.)

R1a itself long predates the appearance of Proto-Indo-European languages – it is first seen in the Russian and Ukrainian Mesolithic, and probably arose in the Black Sea region in the last ice age (although Iran has been suggested on the basis of modern distributions: Underhill *et al.* 2015) – no one claims it was only ever carried by Indo-European speakers. But the case that it dispersed from the Steppe into Central Asia with Indo-Iranian speakers, and thence into South Asia in the Bronze Age, is quite compelling from the phylogeographic pattern. That is, we do not 'assume migrations' – we *infer* them.

Similarly, archaeogeneticists have never claimed that R1a was the only Y-chromosome lineage carried by proto-Indo-European-speakers. Another haplogroup, R1b, which is by far the most common nowadays in western/central Europe, is the main lineage associated with putative westward Indo-European dispersals. Again, the clustering of ancient and modern sequences in the tree provides directionality of movement, and, like R1a, although R1b was already around long before the Metal Ages (Fu *et al.* 2016; Karmin *et al.* 2015), the more ancient lineages are extremely uncommon today (Myres *et al.* 2011) and recall a totally different history from that of the younger, Indo-European-associated starburst (Kivisild 2017; Olalde *et al.* 2018, 2019; Silva *et al.* 2019). Incidentally, parallels to the sex-bias genetic pattern mentioned above, which Danino dismisses as discredited old 'rants on the "powerful" Aryans', have also been recently proposed for Iberia (Olalde *et al.* 2019) – where we find, as in the Subcontinent, non-Indo-European languages, of which Basque still survives today – indicating that the suggested social implications of Bronze Age movements were far from unique to South Asia.

Archaeogeneticists would never be so naive as to equate a particular genetic lineage one-to-one with a language or culture. For example, the non-Indo-European-speaking Basques carry the same R1b lineages as their neighbouring Indo-European-speaking populations. Nor do archaeogeneticists argue that people only ever disperse in one direction. Indeed, it is quite clear that South Asian lineages dispersed both north-westwards and north-eastwards over time (Silva *et al.* 2017). Again, this is straightforward to establish with, for example, tree-based analyses of modern lineages – and even more so with ancient DNA, when sampled in sufficient numbers. We therefore have no need to dispute the historical evidence for such migrations, as Danino supposes.

Furthermore, why is it even an issue? Surely he does not mean to suggest that Indo-European (or Indo-Iranian) languages spread from the Subcontinent to Central Asia and Europe? That would be untenable on historical-linguistic grounds, whether there was genetic evidence for substantial outgoing dispersals at that time or not.

Sanskrit, especially Vedic Sanskrit, is the most archaic of the fully attested IE languages, but the discovery of fragments of more archaic languages during the 20th century showed that Sanskrit and PIE were nevertheless very different, and that Sanskrit was derived from a more ancient language that arose nearer the middle of the Indo-European world, north and west of India. For many points of grammar and vocabulary, the classical language of Hinduism reflects an earlier state of affairs than Ancient Greek or Latin. For this reason, the idea that Sanskrit and Proto-Indo-European were one and the same arose amongst European philologists in the 19th century. But in the West, due to 20th century advances including the discovery of the Hittite and Tocharian languages, linguistic science moved on to better models. The fact that Indo-European studies in 20th century Europe inflamed cultural sensitivity and was misused tragically should not obscure the weight of scientific evidence that Indo-Aryan first entered the Subcontinent from the northwest during the Bronze Age.

Danino is hardly the first person to criticise small sample sizes. It is an easy target – India is a big place. But then, what is striking about India is the distribution of two major language families – Dravidian and Indo-European, one more or less autochthonous and the other with an extraordinary Eurasian distribution – and, historically, of a number of extremely widespread archaeological horizons. Is it so unreasonable, therefore, to propose continent-wide explanations for these phenomena based on continent-wide genetic datasets, without sampling every single region and ethnic group in detail? It is important to recognise that small sample sizes do not automatically imply bias.

Danino appears unaware of research on the dispersal of the Neolithic into central and western Europe by pioneer colonisation, which is one of the best-evidenced conclusions to date in archaeogenetics (Silva and Vander Linden 2017; Lazaridis 2018). Archaeogeneticists are scientists: we work by trying to find the best explanations for the data – even if they contradict assumptions and conclusions of other lines of research, recognising when they might have been inconclusive or even mistaken. And, by the way, reality may rarely be parsimonious, but equally, explanations are rarely possibly without some invocation of parsimony either (and incidentally, parsimony remains an indispensable criterion for reconstructing mitochondrial and Y-chromosome trees). There is a disagreeable tendency amongst some scholars to describe any novel piece of work opposing a model that they dislike as 'refuting' that model, whether the evidence or reasoning is substantial or not. When we deplore a particular perspective, we all like to think that any new scholarly condemnation has administered the final *coup de grâce*. But it is not always so, and sometimes we have to challenge our own assumptions as well as those of others.

Danino's assumptions that history is somehow less rich and complex if migrations were involved, and that archaeogeneticists do not allow for other concurrent processes, such as the establishment and operation of socioeconomic networks, seem inherently implausible. We happily support his insistence on a proper appreciation of the historical context of research and more sophisticated, subtle and complex models for prehistoric processes, as well as his desire for more interdisciplinary understanding fostered by collaboration amongst archaeogeneticists, archaeologists and linguists. But we are disappointed that his erudite critique, which includes more than 50 references, cites only 3 research papers by human geneticists.

Despite his calls for interdisciplinary cooperation, Danino seems reluctant to take the evidence of either archaeogenetics or historical linguistics seriously and is, instead, content to fall back on rehearsing well-worn arguments from ages past. We feel that he should grant new lines of evidence the respect they deserve.

For the question of the origins of the Indo-European languages, India occupies a special and unique position on two counts: as home to more speakers of Indo-European languages than any other country and as the only country where the belief system of an ancient Indo-European-speaking group has continued as the mainstream religion of millions. There is therefore much to be gained if researchers into this important chapter of the human past – working inside and outside India – can maintain a safe space to continue this dialogue.

References

Danino M 2017 The problematics of genetics and the Aryan issue, *The Hindu* 29 June, *https://www.thehindu.com/opinion/op-ed/the-problematics-of-genetics-and-the-aryan-issue/article19165320.ece*

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, Furtwängler A, Haak W, *et al.* 2016 The genetic history of Ice Age Europe. *Nature* **534** 200–205

Karmin M, Saag L, Vicente M, Wilson Sayres MA, Järve M, Talas UG, Rootsi S, Ilumäe A-M, *et al.* 2015 A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* **25** 459–466

Kivisild T 2017 The study of human Y chromosome variation through ancient DNA. *Hum. Genet.* **136** 529–546

Lazaridis I 2018 The evolutionary history of human populations in Europe. *Curr. Opin. Genet. Dev.* **53** 21–27

Majumder PP 2018 Understanding the Aryan debate: population genetic concepts and frameworks. *Curr. Sci.* **114** 971–975

Myres NM, Rootsi S, Lin AA, Järve M, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, *et al.* 2011 A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur. J. Hum. Genet.* **19** 95–101

Narasimhan VM, Patterson NJ, Moorjani P, Lazaridis I, Lipson M, Mallick S, Rohland N, Bernardos R, *et al.* 2018 The genomic formation of South and Central Asia. *bioRxiv* 292581

Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Booth T, Rohland N, Mallick S, *et al.* 2018 The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555** 190–196

Olalde I, Mallick S, Patterson N, Rohland N, Villalba-Mouco V, Silva M, Dulias K, Edwards CJ, *et al.* 2019 The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363** 1230–1234

Reich D 2018 *Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past* (New York: Pantheon Books)

Renfrew C 2000 Archaeogenetics: Towards a population prehistory of Europe; in *Archaeogenetics: DNA and the Population Prehistory of Europe* (eds) Renfrew C and Boyle KV (Cambridge: McDonald Institute for Archaeological Research) pp 3–13

Ringe D, Warnow T and Taylor A 2002 Indo-European and computational cladistics. *Trans. Philol. Soc.* **100** 59–129

Silva F and Vander Linden M 2017, Amplitude of travelling front as inferred from 14C predicts levels of genetic admixture among European early farmers. *Sci. Rep.* **7** 11985

Silva M, Oliveira M, Vieira D, Brandão A, Rito T, Pereira JB, Fraser RM, Hudson B, *et al.* 2017 A genetic chronology for the Indian Subcontinent points to heavily sex-biased dispersals. *BMC Evol. Biol.* **17** 88.

Silva M, Dulias K, Oteo-Garcia G, Gandini F, Edwards C, Pala M, Soares P, Wilson JF, *et al.* 2019 Once upon a time in the West: The archaeogenetics of Celtic origins; in *Exploring Celtic Origins* (eds) B Cunliffe and JT Koch (Oxford: Oxbow Books) pp 153–191

Underhill PA, Poznik GD, Rootsi S, Järve M, Lin AA, Wang J, Passarelli B, Kanbar J, *et al.* 2015 The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur. J. Hum. Genet*. **23** 124–131

Vander Linden M 2019 Towards a clearer view into human prehistory. *Science* **363** 1153–1154

Marina Silva[1], John T Koch[2], Maria Pala[1], Ceiridwen J Edwards[1], Pedro Soares[3,4] and Martin B Richards[1]*

[1]*Department of Biological and Geographical Sciences, School of Applied Sciences, University of Huddersfield, Huddersfield, UK*

[2]*University of Wales Centre for Advanced Welsh and Celtic Studies, National Library of Wales, Wales, UK*

[3]*CBMA (Centre of Molecular and Environmental Biology), Department of Biology, University of Minho, Campus de Gualtar, Braga, Portugal*

[4]*Institute of Science and Innovation for Bio-Sustainability (IB-S), University of Minho, Campus de Gualtar, Braga, Portugal*

**Corresponding author (Email, M.B.Richards@hud.ac.uk)*