

No evidence for unknown archaic ancestry in South Asia

To the Editor — Genomic studies have documented a contribution of archaic Neanderthals and Denisovans to non-African ancestry^{1,2}. Recently, Mondal et al.³ published a major dataset—the largest whole-genome sequencing study of diverse South Asians thus far—including 60 mainland groups and 10 indigenous Andamanese groups. They reported analyses claiming that nearly all South Asians harbor ancestry from an unknown archaic human

population that is neither Neanderthal nor Denisovan. However, the statistics cited in support of this conclusion do not replicate in other datasets, and in fact contradict the conclusion.

The main evidence cited by Mondal et al.³ is statistics suggesting that South Asians (indigenous Andamanese and mainland Indian groups) share fewer derived alleles with sub-Saharan Africans than Europeans and East Asians do; such statistics have previously

been reported in Australo-Melanesians, for whom these signals represent key evidence of Denisovan admixture². To document their signal, Mondal et al.³ computed *D* statistics^{1,4} of the form

$$D(\text{ancestral allele, African; East Asian, } X)$$

These statistics test the hypothesis of an equal rate of derived allele sharing of East Asians and group *X* with Africans.

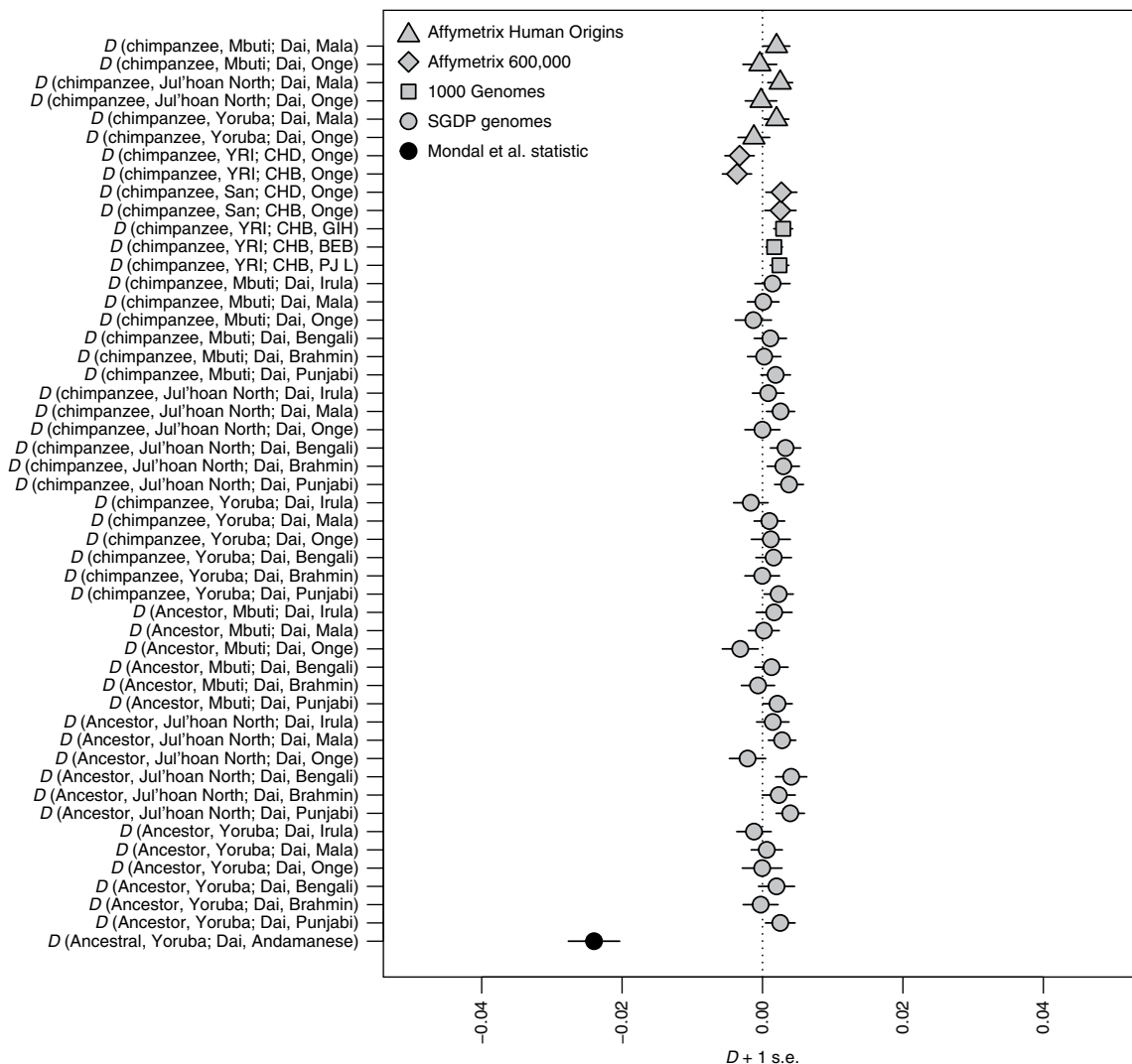


Fig. 1 | The key statistic used to support the claim of unknown archaic ancestry by Mondal et al. in Andamanese and mainland Indians is inconsistent with all previously published datasets. Evidence for unknown archaic ancestry in Andamanese and mainland Indians does not replicate in four previously published datasets. Error bars show one weighted block jackknife standard error (s.e.) on each side. All statistics except the one reported by Mondal et al.³ are consistent with no excess archaic admixture in South Asians ($|Z| < 2$). SGDP, Simons Genome Diversity Panel.

Mondal et al.³ also reported statistics with a European population in place of East Asians, but it is already known that Asian populations have a greater amount of Neanderthal ancestry than Europeans^{5,6}, so we focus here on the new claim of Mondal et al.³ of more archaic ancestry in South Asian populations than in East Asians. In the computation of Mondal et al.³, these statistics showed excess derived allele sharing between Africans and East Asians when X was any Indian group or Andamanese, a result they interpret as evidence of more archaic ancestry in South Asians than in East Asians. As they found no evidence of excess allele sharing with Neanderthals or Denisovans, they argue that the contribution is from an unsampled archaic lineage.

We sought to replicate these statistics in four different datasets: two SNP datasets of ~600,000 SNPs each^{4,7,8}, publicly available genotypes for whole-genome sequence data from phase 3 of the 1000 Genomes Project (~78 million SNPs)⁹ and diploid genotypes called for high-coverage genomes in the Simons Genome Diversity Panel¹⁰ (~34 million SNPs using the recommended 'level 1' filters; available publicly or for academic research on population history). We found that there was no evidence for excess archaic ancestry in South Asians in any of these four datasets (Fig. 1), and in fact the values reported by Mondal et al.³ ($D = -0.024 \pm 0.004$; Supplementary Table 13 of their study) are inconsistent with those in each of these other datasets (all $P < 1 \times 10^{-5}$ by a one-tailed test).

Mondal et al.³ also reported statistics suggesting more archaic ancestry in indigenous Australians than in indigenous Papuans, as reflected in D statistics that were far more skewed from zero when X was

Australian than when it was Papuan³. They interpret this as evidence of a greater extent of unknown archaic ancestry in Australians than in Papuans. However, we did not replicate this excess when recomputing this statistic using high-coverage genome data from these populations in the Simons Genome Diversity Panel: $D(\text{chimpanzee}, \text{Yoruba}; \text{Dai}, \text{Australian}) = -0.031 \pm 0.003$ and $D(\text{chimpanzee}, \text{Yoruba}; \text{Dai}, \text{Papuan}) = -0.029 \pm 0.003$. In addition, a direct comparison between Australians and Papuans provided no evidence for a difference: $D(\text{chimpanzee}, \text{Yoruba}; \text{Australian}, \text{Papuan})$ was only $Z = 0.6$ standard errors from zero. These findings support the notion that Papuans and Australians descend from a homogeneous ancestral population and are inconsistent with the suggestion that Australians harbor much more archaic ancestry than Papuans.

In fact, some of the statistics computed by Mondal et al.³ directly contradict their proposed model of unknown archaic ancestry specific to Indians and Andamanese. Figure 1 of Mondal et al.³ suggests that the Riang (RIA)—a Tibeto-Burman-speaking group from the northeast of India for which sequencing data are newly reported in the study—derive almost all of their ancestry from the same East Asian lineages as populations such as the Dai and Han Chinese, which in Fig. 2 of Mondal et al.³ have no evidence of unknown archaic ancestry. Under the authors' hypothesis of more archaic ancestry in lineages that are unique to South Asia than in lineages shared with East Asians, one would not expect a significant statistic in the Riang, but in fact the signal is just as strong as it is for the Andamanese Onge, Andamanese Jarawa, mainland Irula and mainland Birhor, the great majority of whose ancestry is

inferred to derive from lineages unique to South Asia.

One possible explanation for the skew that the authors observe is batch artifacts, reflecting differences in laboratory or computer processing between the data newly reported by Mondal et al.³ and the data from non-Indians used for comparison⁶. Whatever the explanation, our analyses contradict the claim of unknown archaic ancestry in South Asians. □

Pontus Skoglund^{1,2,4*}, Swapan Mallick^{1,2,3}, Nick Patterson² and David Reich^{1,2,3*}

¹Department of Genetics, Harvard Medical School, Boston, MA, USA. ²Broad Institute of Harvard and MIT, Cambridge, MA, USA. ³Howard Hughes Medical Institute, Boston, MA, USA. ⁴Present address: The Francis Crick Institute, London, UK. *e-mail: pontus.skoglund@crick.ac.uk; reich@genetics.med.harvard.edu

Published online: 26 April 2018

<https://doi.org/10.1038/s41588-018-0097-9>

References

- Green, R. E. et al. *Science* **328**, 710–722 (2010).
- Reich, D. et al. *Nature* **468**, 1053–1060 (2010).
- Mondal, M. et al. *Nat. Genet.* **48**, 1066–1070 (2016).
- Patterson, N. et al. *Genetics* **192**, 1065–1093 (2012).
- Skoglund, P. & Jakobsson, M. *Proc. Natl. Acad. Sci. USA* **108**, 18301–18306 (2011).
- Meyer, M. et al. *Science* **338**, 222–226 (2012).
- Reich, D. et al. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
- Lazaridis, I. et al. *Nature* **513**, 409–413 (2014).
- 1000 Genomes Project Consortium. *Nature* **526**, 68–74 (2015).
- Mallick, S. et al. *Nature* **538**, 201–206 (2016).

Acknowledgements

P.S. was supported by the Swedish Research Council (VR grant 2014-453). N.P. and D.R. were supported by NIH grant GM100233, and D.R. is a Howard Hughes Medical Institute investigator.

Competing interests

The authors declare no competing interests.