 doi:10.20944/preprints201808.0480.v1

Unexplained allele-calling errors may account for apparent Denisovan-Neanderthal F1 genome

David Curtis
d.curtis@ucl.ac.uk
00 44 7973 906 143

UCL Genetics Institute, UCL, Darwin Building, Gower Street, London WC1E 6BT.
Centre for Psychiatry, Barts and the London School of Medicine and Dentistry, Charterhouse Square, London EC1M 6BQ.

Kaywords

Neanderthal, Denisova, genome, allele.

Abstract

The recent report that DNA extracted from ancient bone must have from the offspring of a female Neanderthal and a male Denisovan depends on the inference that the subject has a high level of heterozygosity for Neanderthal and Denisovan alleles across the genome. Here I point out that the relative frequencies of derived transversion polymorphisms vary markedly between the new specimen, Denisova 11, and two high-coverage Neanderthal genomes. In Denisova 11 the AC and CG polymorphisms are much commoner than the others and are almost twice as common as the AT polymorphism. In the high-coverage Neanderthal genomes the four types of transversion are about equally common, with the AT being slightly commoner than the others. These results suggest that allele-calling errors are frequent and that this may provide an alternative explanation for the observed heterozygosity.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2018 doi:10.20944/preprints201808.0480.v1

Unexplained allele-calling errors may account for apparent Denisovan-Neanderthal F1 genome

David Curtis

The conclusion that DNA extracted from a bone found in a cave inhabited by Denisovans and Neanderthals must have come from the offspring of a female Neanderthal and a male Denisovan essentially arises from two observations (Slon *et al.*, 2018). The first is that the specimen, referred to as Denisova 11, has increased heterozygosity compared to the genomes of two Neanderthals and one Denisovan. The second is that observed alleles are consistent with the specimen being mostly heterozygous for Neanderthal and Denisovan alleles across the genome. To address the latter issue the authors examined informative transversion sites at which a Neanderthal was homozygous for one allele (NN) and a Denisovan for the other (DD). For each site they sampled randomly two DNA fragments without replacement from Denisova 11. As they point out, if the genotype is actually ND then the probabilities for the two random fragments at this site to have alleles NN, ND or DD are 0.25, 0.5 and 0.25. If all sites across the genome are heterozygous (i.e. the subject is the F1 offspring of these two parents) then these yield the expected proportions for all calls. By contrast, for an F2 subject the expected proportions are 37.5%, 25% and 37.5%. They report that the observed proportions for Denisova 11 are 27.3%, 43.5% and 29.2% using the Altai Neanderthal genome (Prüfer *et al.*, 2014) and 26.9%, 43.1% and 30.0% using the Vindija 33.19 Neanderthal genome (Prüfer *et al.*, 2017). Taking account of the fact that the ancestral genomes used would not be identical with the actual parents, they conclude that the observed proportions are consistent with Denisova being F1 but not being F2 ($p<10^{-16}$) or the product of some more distantly admixed mating.

A couple of caveats should be noted. With regard to the relative increase in heterozygosity, this might reflect that Denisova 11 had parents who were simply less closely related than was usually the case in Neanderthal and Denisovan populations. Perhaps early hominins tended to live in family groups in which most matings were highly consanguineous. If the parents of Denisova came from two more distantly related groups then she might have significantly increased heterozygosity. Of course, the proposed explanation, that the parents are of two different species, represents an extreme case of this scenario. However it is not obvious that the observed heterozygosity is incompatible with a range of alternative scenarios. The second issue worth mentioning is that alleles described as “Neanderthal” or “Denisovan” are not necessarily specific to either species. The attribution is based on the observation that the alleles were derived, i.e. present in hominids but not non-human primates, and that they were homozygous for different alleles in a single Neanderthal and a single Denisovan genome. If one considers biallelic polymorphisms whose alleles occur at equal frequency (0.5) in both species then an eighth of them are expected to show discrepant genotypes, apparently NN and DD, entirely by chance. So even if Neanderthals and Denisovans had identical genetic profiles taking a single genome from each would produce many variants which would be allocated to one species or the other. Thus we cannot assume that all variants labelled as such do in fact reflect the ancestry of one or other species. This point is illustrated by the report that two random fragments from a third Neanderthal (‘Goyet Q56-1’ (Hajdinjak *et al.*, 2018)) were DD in 7.5% of cases. There is no suggestion that this is supposed to indicate a Denisovan contribution to ancestry and clearly these D alleles cannot be specific to Denisovans.

While it seems possible to argue that there are alternative reasons for increased heterozygosity and that the N and D alleles may not in fact be species-specific, neither of these considerations can

 doi:10.20944/preprints201808.0480.v1

explain the high rate of ND calls. Even if all variants had a frequency of 0.5 in both species, the proportion of heterozygous variants in an individual would be 50%. Then if one selected two DNA fragments at random one would expect to obtain ND on only 25% of occasions because half the time the second fragment would have the same allele as the first by chance. The only way for this proportion to rise would be if there were some systematic bias whereby it was more likely than not that the two fragments should come from different members of the homologous chromosome pair and it is difficult to think of any biological mechanism or lab process which could produce such an artefact. Thus, the key finding that the observed proportion of ND calls is around 43% remains. If the allele calls are accurate it is hard to think of an alternative explanation than that Denisova 11 has a Neanderthal and a Denisovan parent.

However there are grounds to suspect that the allele calls are not, in fact, accurate. The reason for this comes from the data presented in Supplementary Table 5.2 (Slon *et al.*, 2018), the relevant cells from which are reproduced here in Table 1 and Figure 1. These show the frequencies of different types of transversion polymorphism in the regions of the Denisova 11 genome which are said to harbor ancestry from Neanderthals on both chromosomes. (These are the only regions for which polymorphism frequencies are reported broken down by ancestral and derived allele.) The authors comment that the overall frequencies are in general higher in Denisova 11 than in the two high-coverage Neanderthal genomes, which is true. However what is striking and troubling is that in Denisova 11 the AC and CG polymorphisms are much commoner than the others and are almost twice as common as the AT polymorphism. In the high-coverage Neanderthal genomes the four types of transversion are about equally common, with the AT being slightly commoner than the others. There is no plausible biological explanation for there really to be a marked excess of AC and CG polymorphisms and for the relative frequencies to differ so markedly between Denisova 11 and the Neanderthals. It is likewise not plausible that such a marked deviation could have occurred by chance. Thus one can only conclude that some systematic errors are occurring and that the allele calls for Denisova 11 are not accurate. Allele-calling errors would be expected to increase the estimated number of heterozygotes. The effect seems to be large and may quite possibly be sufficient to explain the high proportion of observed ND calls. Unless this matter can be satisfactorily resolved one cannot conclude that Denisova 11 had a Neanderthal mother and a Denisovan father.

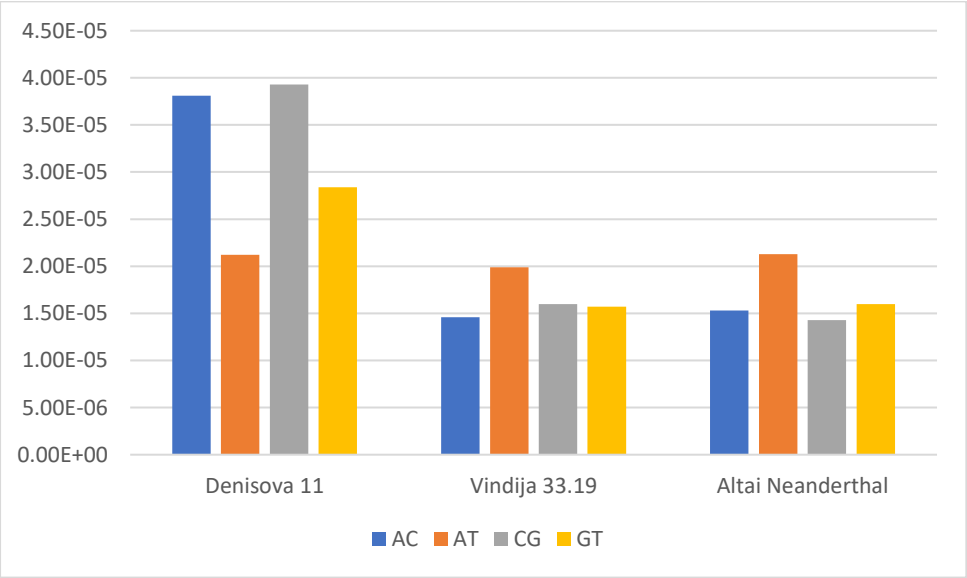
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2018 doi:10.20944/preprints201808.0480.v1

Table 1. Frequencies of derived transversion polymorphisms in regions harbouring Neanderthal ancestry in Denisova 11 within Denisova 11, Vindija 33.19 and the Altai Neanderthal, extracted from Supplementary Table 5.2 of the original publication (Slon *et al.*, 2018).

Polymorphism	Denisova 11	Vindija 33.19	Altai Neanderthal
AC	3.81E-05	1.46E-05	1.53E-05
AT	2.12E-05	1.99E-05	2.13E-05
CG	3.93E-05	1.60E-05	1.43E-05
GT	2.84E-05	1.57E-05	1.60E-05

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2018 doi:10.20944/preprints201808.0480.v1

Figure 1. Frequencies of derived transversion polymorphisms in regions harbouring Neanderthal ancestry in Denisova 11 within Denisova 11, Vindija 33.19 and the Altai Neanderthal, extracted from Supplementary Table 5.2 of the original publication (Slon *et al.*, 2018).



References

Hajdinjak, M., Fu, Q., Hübner, A., Petr, M., Mafessoni, F., Grote, S., Skoglund, P., Narasimham, V., Rougier, H., Crevecoeur, I., Semal, P., Soressi, M., Talamo, S., Hublin, J.-J., Gušić, I., Kućan, Ž., Rudan, P., Golovanova, L. V., Doronichev, V. B., Posth, C., Krause, J., Korlević, P., Nagel, S., Nickel, B., Slatkin, M., Patterson, N., Reich, D., Prüfer, K., Meyer, M., Pääbo, S. and Kelso, J. (2018) 'Reconstructing the genetic history of late Neanderthals', *Nature*. Nature Publishing Group, 555(7698), pp. 652–656. doi: 10.1038/nature26151.

Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., Reher, D., Hopfe, C., Nagel, S., Maricic, T., Fu, Q., Theunert, C., Rogers, R., Skoglund, P., Chintalapati, M., Dannemann, M., Nelson, B. J., Key, F. M., Rudan, P., Kućan, Ž., Gušić, I., Golovanova, L. V, Doronichev, V. B., Patterson, N., Reich, D., Eichler, E. E., Slatkin, M., Schierup, M. H., Andrés, A. M., Kelso, J., Meyer, M. and Pääbo, S. (2017) 'A high-coverage Neandertal genome from Vindija Cave in Croatia.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 358(6363), pp. 655–658. doi: 10.1126/science.aao1887.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P. H., de Filippo, C., Li, H., Mallick, S., Dannemann, M., Fu, Q., Kircher, M., Kuhlwilm, M., Lachmann, M., Meyer, M., Ongyerth, M., Siebauer, M., Theunert, C., Tandon, A., Moorjani, P., Pickrell, J., Mullikin, J. C., Vohr, S. H., Green, R. E., Hellmann, I., Johnson, P. L. F., Blanche, H., Cann, H., Kitzman, J. O., Shendure, J., Eichler, E. E., Lein, E. S., Bakken, T. E., Golovanova, L. V., Doronichev, V. B., Shunkov, M. V., Derevianko, A. P., Viola, B., Slatkin, M., Reich, D., Kelso, J. and Pääbo, S. (2014) 'The complete genome sequence of a Neanderthal from the Altai Mountains', *Nature*. Nature Publishing Group, 505(7481), pp. 43–49. doi: 10.1038/nature12886.

Slon, V., Mafessoni, F., Vernot, B., de Filippo, C., Grote, S., Viola, B., Hajdinjak, M., Peyrégne, S., Nagel, S., Brown, S., Douka, K., Higham, T., Kozlikin, M. B., Shunkov, M. V., Derevianko, A. P., Kelso, J., Meyer, M., Prüfer, K. and Pääbo, S. (2018) 'The genome of the offspring of a Neanderthal mother and a Denisovan father', *Nature*. Nature Publishing Group, p. 1. doi: 10.1038/s41586-018-0455-x.