

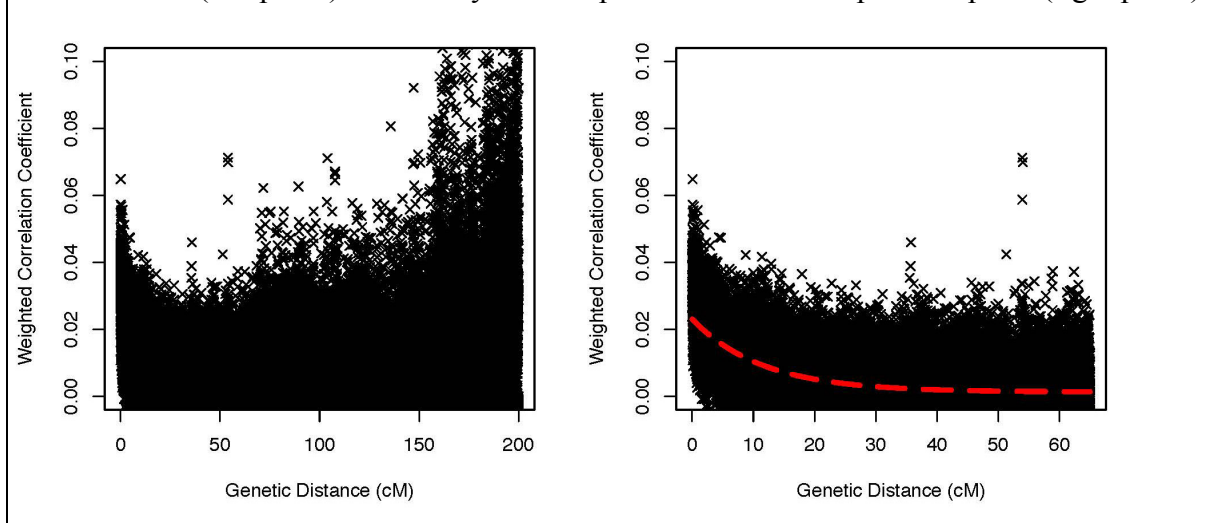
Dating the most recent Neanderthal admixture into the ancestors of Oase 1

Admixture between populations induces correlation in ancestry across the genome of an admixed individual, and the extent of this correlation is informative about the time since mixture²³⁻²⁵. As in⁶ we use this signal to estimate the date of Neanderthal introgression.

Method A – Fitting an exponential decay

We use the Oxford combined genetic map²⁶ and calculate the average covariance over all pairs of SNPs in 0.001 cM bins that carry Neanderthal alleles as defined above. The exponential decay gets lost in noise around 65 cM (Figure S5.1) and thus we fit to this point.

Figure S5.1. Pairwise covariance for SNPs that match the ascertainment scheme in which Neanderthal carries different alleles from Africans. The decay gets lost in noise beyond around 65 cM (left panel) so we only fit an exponential function up to this point (right panel).



Assuming a single pulse of admixture, we fit an exponential function ($y = Ae^{-(n-1)d} + c$), where n = number of generations since Oase 1 had a Neanderthal ancestor and d = genetic distance (in Morgans). Figure S5.1 shows the covariance curves. We compute standard errors using a Weighted Block Jackknife²⁰, removing one chromosome in each run and studying the variability in the estimated dates of mixture. Using this method, we estimate that Oase 1 had Neanderthal ancestors 8.1 ± 5.5 generations back in his family tree.

This estimated date of admixture is an average of both the most recent Neanderthal admixture into the ancestors of Oase 1, and older admixture that was perhaps shared with other non-Africans. We did not succeed in fitting the data as a mixture of two exponential distributions. We therefore turned to alternative methods that focused on dating the most recent mixture.

Method B – Fitting the distribution of large chunks

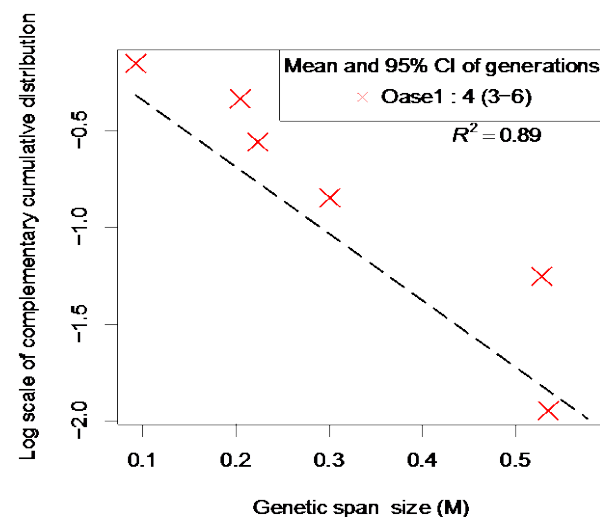
Because the chunks of Neanderthal introgression are so large that we can reasonably infer the positions of the largest chunks, and because we are interested specifically in the admixture that gave rise to the largest chunks, we reasoned that it might be valid to simply determine the sizes of the large chunks, and then to fit a distribution to them. For this analysis, we use the seven chunks identified by the HMM, whose positions are listed in Table S5.1.

We assume that the chunks resulting from the most recent admixture event have a distribution $y = Ae^{-(n-1)d}$, where n = number of generations since Oase 1 had a Neanderthal ancestor and d = genetic distance (in Morgans). The fact that the exponent has the quantity

($n-1$) in it reflects the fact that Neanderthal and modern human chromosomes are only expected to begin to be observed in recombinant form in the second generation after admixture (in the first generation after admixture, each individual has one entirely Neanderthal and one entirely modern human chromosome at each locus).

Taking the natural logarithm of this distribution, we get $\ln(y) = -(n-1)d + \ln(A)$. Figure S5.2 shows the cumulative distribution of the number of chunks, plotted on a log-scale to allow a linear fit. The slope of this plus 1 translates to a date of admixture. The resulting fit is good ($R^2=0.89$), and has a slope corresponding to an admixture date of 4 generations ago (95% confidence interval obtained by linear regression of 3-6 generations ago).

Figure S5.2 Fraction of the Oase 1 genome comprised of the seven largest chunks. The slope of the curve is expected to be ($n-1$), and can be used to estimate the number of generations since Oase 1 had his most recent Neanderthal ancestor.



Method C - Probability of the observed spans of the largest Neanderthal chunks

We examined the expected distribution of the largest chunks of introgression, including the 1 largest, 2 largest, 3 largest, and 4 largest, for different time depths of introgression. We reasoned that the genetic spans of the largest chunks might be robust statistics because:

- We likely have good power to recognize the largest chunks, but may have imperfect power to recognize the smaller chunks. Thus, fitting a model to the largest chunks may be more robust than fitting a model including some smaller chunks. As chunk sizes become smaller, it becomes increasingly possible that we have missed chunks of similar size that are real and hence our measurement of the distribution is less likely to be accurate.
- We are interested here in studying the most recent Neanderthal introgression. When we include shorter chunks in the analysis, we expect to see a size range where older Neanderthal introgression events may be contributing to the observed patterns, and thus our estimate of the date may be higher than that of the most recent mixture.

We carried out a series of simulations that fragmented the genome generation by generation assuming the empirically measured sex-averaged genetic map (using the map from²⁷).

Table S5.2 reports the results of this simulation study, showing the number of generations since mixture that produces a 95% confidence interval of average chunk size for the top k chunks that includes our empirical observation. We observe that for just the top chunk ($k=1$), Oase 1 is consistent with having Neanderthal admixture 4-8 generations in the past. For averages of the top 1-2 ($k=2$), 1-3 ($k=3$), and 1-4 ($k=4$) chunks, Oase 1 is consistent with having Neanderthal admixture 4-6 generations ago. We conclude that the most recent admixture likely occurred 4-6 generations ago.

Table S5.2. Average size of the k largest chunks for different numbers of generations since mixture. We carried out 10,000 simulations for each parameter combination (number of generations since mixture, and number k of the largest chunks being averaged), and give the 95% credible interval. Gray boxes include the empirical value.

Generation	k=1	k=2	k=3	k=4
1	270-271	263-264	248-249	237-238
2	128-271	122-231	115-210	111-196
3	72-206	67-172	63-152	59-138
4	41-159	37-127	34-109	31-97
5	22-122	19-95	16-80	15-70
6	9-95	8-71	7-59	7-51
7	2-75	3-54	4-44	5-39
8	1-60	2-43	3-36	4-32
9	0-49	1-35	2-30	3-27
10	0-42	1-31	2-26	3-24
Empirical	58	56	55	49

We can compare the estimated date of Neanderthal mixture obtained from the chunk size distribution to that needed to produce the observed genome-wide proportion of Neanderthal ancestry if all the Neanderthal admixture occurred recently. The average proportions of Neanderthal ancestry expected to be contributed by ancestors at different time depths of admixture are given in Table S5.3. Assuming that all the extra Neanderthal ancestry in Oase 1 was contributed by a single recent ancestor, it would be expected from Table S5.3 that this ancestor lived 3-5 generations back. This overlaps with the estimates from chunk span estimates: 3-6 generations back by Method B and 4-6 generations back by Method C.

Table S5.3. Average proportion of Neanderthal ancestry expected to be contributed by a Neanderthal ancestor different numbers of generation ago

Generations back	Ancestor	Proportion Neanderthal
1	Parent	50%
2	Grandparent	25%
3	Great-grandparent	12.5%
4	Great-great-grandparent	6.25%
5	Great-great-great-grandparent	3.13%
6	Great-great-great-great-grandparent	1.56%

We conclude that it is likely that Neanderthal admixture happened 4-6 generations ago. In other words, Oase 1 possibly had a Neanderthal great-great-grandparent and certainly had a Neanderthal great-great-great-great-grandparent.

Supplementary References

- 1 Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2223-2227 (2013).
- 2 Renaud, G., Kircher, M., Stenzel, U. & Kelso, J. freeIbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* **29**, 1208-1209 (2013).
- 3 Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-406 (2014).
- 4 Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Current biology : CB* **23**, 553-559 (2013).
- 5 Shapiro, B. *et al.* A Bayesian phylogenetic method to estimate unknown sequence ages. *Molecular biology and evolution* **28**, 879-887 (2011).
- 6 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449 (2014).
- 7 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574 (2003).
- 8 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226 (2012).
- 9 Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416-426 (2008).
- 10 Kloss-Brandstatter, A. *et al.* HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Human mutation* **32**, 25-32 (2011).
- 11 Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708-713 (2000).
- 12 Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318-321 (2009).
- 13 Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818 (1998).
- 14 Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214 (2007).
- 15 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093 (2012).
- 16 Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413 (2014).
- 17 Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* (2015).
- 18 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722 (2010).
- 19 Trinkaus, E. *et al.* An early modern human from the Pestera cu Oase, Romania. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11231-11236 (2003).
- 20 Kunsch, H. R. The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics* **17**, 1217-1241 (1989).
- 21 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49 (2014).

- 22 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060 (2010).
- 23 Loh, P. R. *et al.* Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233-1254 (2013).
- 24 Moorjani P, P. N., Hirschhorn JN, Keinan A, Hao L, et al. The History of African Gene Flow into Southern Europeans, Levantines, and Jews. . *PLoS Genet* **7**, e1001373 (2011).
- 25 Hellenthal, G. *et al.* A Genetic Atlas of Human Admixture History. *Science* **343**, 747-751 (2014).
- 26 International HapMap, C. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861 (2007).
- 27 Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature genetics* **31**, 241-247 (2002).