Multiple episodes of interbreeding between Neanderthal and modern humans

Fernando A. Villanea and Joshua G. Schraiber

Neanderthals and anatomically modern humans overlapped geographically for a period of over 30,000 years following human migration out of Africa. During this period, Neanderthals and humans interbred, as evidenced by Neanderthal portions of the genome carried by non-African individuals today. A key observation is that the proportion of Neanderthal ancestry is ~12–20% higher in East Asian individuals relative to European individuals. Here, we explore various demographic models that could explain this observation. These include distinguishing between a single admixture event and multiple Neanderthal contributions to either population, and the hypothesis that reduced Neanderthal ancestry in modern Europeans resulted from more recent admixture with a ghost population that lacked a Neanderthal ancestry component (the ‘dilution’ hypothesis).

To summarize the asymmetric pattern of Neanderthal allele frequencies, we compiled the joint fragment frequency spectrum of European and East Asian Neanderthal fragments and compared it with both analytical theory and data simulated under various models of admixture. Using maximum-likelihood and machine learning, we found that a simple model of a single admixture did not fit the empirical data, and instead favour a model of multiple episodes of gene flow into both European and East Asian populations. These findings indicate a longer-term, more complex interaction between humans and Neanderthals than was previously appreciated.

When anatomically modern humans dispersed out of Africa, they encountered and hybridized with Neanderthals. The Neanderthal component of the modern human genome is ubiquitous in non-African populations, and yet it is quantitatively small, representing on average only ~2% of these genomes. This pattern of Neanderthal ancestry in modern human genomes was initially interpreted as evidence of a single period of admixture, occurring shortly after the out-of-Africa bottleneck. However, subsequent research showed that Neanderthal ancestry is higher by ~12–20% in modern East Asian individuals relative to European individuals.

Neanderthals occupied a vast area of Asia and Europe at the time anatomically modern humans dispersed outside of Africa (~75,000 years BP), and later Europe and Asia (~47,000–55,000 years BP) before the diversification of East Asian and European lineages. The genome of Ust’-Ishim—an ancient individual of equidistant relation to modern East Asians and Europeans—has similar levels of Neanderthal ancestry to modern European individuals relative to modern European individuals.

Neanderthals occupied a vast area of Asia and Europe at the time anatomically modern humans dispersed outside of Africa (~75,000 years BP), and later Europe and Asia (~47,000–55,000 years BP) before the diversification of East Asian and European lineages. The genome of Ust’-Ishim—an ancient individual of equidistant relation to modern East Asians and Europeans—has similar levels of Neanderthal ancestry to modern European individuals. Yet these are found in longer haplotypes, consistent with an admixture episode occurring ~52,000–58,000 years BP. Given the extensive support for a single, shared admixture among Eurasians, there is extensive debate surrounding the observation of increased Neanderthal ancestry in East Asians.

There are several hypotheses that may explain the discrepancy in Neanderthal ancestry between Europeans and East Asians. It is possible that admixture occurred in a singular episode, or ‘pulse’, of gene flow, but demographic and/or selective forces shifted the remaining Neanderthal alleles into the frequencies we see in modern populations. Among these explanations are differential strengths of purifying selection across Eurasia and that modern Europeans lost part of their Neanderthal ancestry through ‘dilution’ by a ghost population that was unadmixed. It is also possible that admixture occurred multiple times; the first pulse of Neanderthal gene flow into the population ancestral to East Asians and Europeans was supplemented by additional pulses after both populations had diverged.

Sankararaman et al. proposed that differences in the level of Neanderthal ancestry in East Asian individuals could be explained by their lower ancestral effective population size relative to Europeans, which would reduce the efficacy of purifying selection against deleterious Neanderthal alleles. However, Kim and Lohmueller found that differences in the strength of purifying selection and population size are unlikely to explain the enrichment of Neanderthal ancestry in East Asian individuals. This conclusion was further strengthened by Juric et al.

Another hypothesis consistent with a single episode of gene flow is that Neanderthal ancestry in modern Europeans was diluted by one of the populations that mixed to create modern Europeans. This population, dubbed ‘basal Eurasian’, possibly migrated out of Africa separately from the population receiving the pulse of Neanderthal gene flow, and thus had little to no Neanderthal ancestry.

In contrast, admixture may have occurred multiple times; the first pulse of Neanderthal gene flow into the population ancestral to East Asians and Europeans was supplemented by additional episodes after both populations had diverged. The finding of an individual from Peștera cu Oase, Romania with a recent Neanderthal ancestor provides direct evidence of additional episodes of interbreeding, although this individual is unlikely to have contributed to modern-day diversity. However, Neanderthal ancestry has remained relatively constant across tens of thousands of years of Eurasian history, suggesting that any additional admixture events must have been smaller in scale than the initial episodes of interbreeding.

Here, we study the asymmetry in the pattern of Neanderthal introgression in modern human genomes between individuals of...
East Asian and European ancestry. We summarize the asymmetric distribution of Neanderthal ancestry tracts in the East Asian and European individuals in the 1000 Genomes Project panel in a joint fragment frequency spectrum (FFS) matrix. First, we fit analytical models using maximum likelihood to explain the distribution of fragments in European and Asian individuals marginally. Then, we compare the joint FFS with the output of genomic data simulated under specific models of admixture between Neanderthals and anatomically modern humans to achieve a higher-resolution picture of the interplay of different demographic forces. Our results support a complex model of admixture, with early admixture occurring before the diversification of European and East Asian lineages, and secondary episodes of gene flow into both populations independently.

Results

We constructed the joint FFS by analysing published datasets of Neanderthal fragment calls in 1000 Genomes Project individuals[10,19]. To avoid complications due to partially overlapping fragments and difficulties calling the edges of fragments, we computed fragment frequencies by sampling a single site every 100 kilobases (kb) and asking how many haplotypes were introgressed with confidence above a certain cut-off (Fig. 1). Our main results make use of fragments called with a posterior probability of 0.45 in the Steinrücken et al. dataset[19], although we verified robustness across a range of cut-offs and between datasets (Supplementary Material). The observed average proportions of Neanderthal ancestry in European and East Asian individuals were 0.0137 and 0.0164, respectively, corresponding to an average enrichment of 19.6% in East Asian individuals (Supplementary Fig. 5 shows how this quantity changes across cut-offs).

We first developed analytical theory to understand what the FFS would look like in each population separately under different demographic models. To our surprise, we found that when looking only at the marginal distribution of introgressed fragment frequencies, the one-pulse model and dilution model were not statistically identifiable (Supplementary Material). In contrast, the two-pulse model was identifiable. Moreover, the analytical theory revealed that population size history only impacted the FFS within each population as a function of the effective population size. Intuitively, this arises because once fragments enter the population, their frequency dynamics only depends on the effective population size, rather than the specifics of the population size history. With this in mind, we developed a maximum-likelihood procedure to fit one- and two-pulse models to the European and East Asian marginal spectra (Methods), and found strong support for the two-pulse model in both cases (likelihood ratio statistic $\Lambda = 193.91$ in East Asians, nominal $P=7\times10^{-43}$; Fig. 2a,b). A subsequent goodness-of-fit test strongly rejected the fit of the one-pulse model ($P=2\times10^{-26}$ in East Asians, $P=0$ in Europeans; $\chi^2$ goodness-of-fit test), but could not reject the fit of the two-pulse model in either population ($P=1$ in East Asians; $P=0.95$ in Europe; $\chi^2$ goodness-of-fit test; see also Supplementary Fig. 1, which shows the residuals of each fit). Thus, we concluded from analysing each population in isolation that the history of admixture was complex, and involved multiple matings with Neanderthals.

Nonetheless, looking at each population individually, we did not have the power to estimate the relative contribution of dilution and multiple admixtures in shaping the patterns of Neanderthal fragments seen between Europe and Asia. To gain a more global picture of the history of human–Neanderthal interbreeding, we developed a supervised machine-learning approach that we applied to the joint FFS between Europe and Asia (Fig. 2c,d). A difficulty when simulating Neanderthal admixture is the large number of free parameters associated with modelling multiple populations from which we have incomplete demographic information. Supervised machine learning applied to genomic datasets is becoming a popular solution for inference (for examples, see refs 20–22). Of particular interest to this study, supervised machine learning has demonstrated the capacity for optimizing the predictive accuracy of an algorithm in datasets that cannot be adequately modelled with a reasonable number of parameters[23]. In practice, this results in the ability to describe natural processes even when they are based on incomplete or imprecise models[23]. Supervised machine learning implementing hidden layers, or deep learning, is particularly effective in population genetic inference and learning informative features of data[24]. A definitive advantage of deep learning is that it makes full use of datasets to learn the mapping of data to parameters, allowing inference from sparse datasets[21]. Comparable likelihood-free inference methods, such as ABC, typically use a rejection algorithm, resulting in most simulations being thrown away. This necessitates a very large number of simulations for accurate inference[25,24]. Deep-learning methods also have the potential to generalize in non-local ways, allowing them to make predictions for data not covered by the training set[25,24].

We simulated Neanderthal admixture by specifying five demographic models with different numbers of admixture events (Fig. 3), and produced joint FFS under a wide range of parameters. We used the simulated FFS to train a fully connected neural network (FCNN). The trained network classified models successfully ~58% of the time (well above the 20% expected by chance) and was not over-fit to the training data (Supplementary Fig. 2). We then examined how the precision of the prediction changed when we required different levels of support for the chosen model (Fig. 4a). Crucially,
we saw that when the classifier had high confidence in a prediction, it was very often correct, and that multiple pulse models were not often confused with the dilution model (Supplementary Fig. 3).

Finally, we applied the trained FCNN to our empirical joint FFS (Figs. 2c,d and 4b). Strikingly, we found that the FCNN supported our two most complicated demographic models, favouring a model with three pulses of admixture (posterior probability ~0.55), and with a lower probability, a model with three pulses of admixture and dilution (posterior probability ~0.44). These results were consistent across a range of cut-offs for calling introgressed fragments (Supplementary Fig. 4), were robust to errors in fragment calling (Supplementary Figs. 9 and 10) and dovetailed with our maximum-likelihood results, showing that the best-fit model must include multiple episodes of human–Neanderthal interbreeding.

### Discussion

Despite initial indications of a simple history of admixture between humans and Neanderthals, more detailed analyses suggest that there might be additional, population-specific episodes of admixture. By analysing the joint FFS of introgressed Neanderthal haplotypes in modern Europeans and Asians, we found strong support for a model of multiple admixture events. Specifically, our results support a model in which the original pulse of introgression into the ancestral Eurasian population is supplemented with additional pulses to both European and East Asian populations after those populations diverge, resulting in elevated Neanderthal ancestry in East Asians relative to Europeans. This is similar to a model recently proposed by Vernot et al.\textsuperscript{13} for explaining differential levels of Neanderthal ancestry across Europe, Asia and Melanesia. Importantly, our results exclude a demographic model where the difference in Neanderthal
ancestry between Europeans and East Asians is driven primarily through dilution of Neanderthal ancestry in Europe due to recent admixture with basal Eurasians—a population lacking Neanderthal ancestry. Nonetheless, we cannot exclude dilution as playing a role in the differences in Neanderthal ancestry between Europe and East Asia; a model that includes multiple pulses of Neanderthal introgression and dilution through basal Eurasians was the second likeliest model in the five-model comparison. Given the evidence that basal Eurasians contributed to the modern European gene pool\textsuperscript{16}, we suspect that dilution plays a role in shaping the pattern of Neanderthal ancestry across Eurasia. However, a large amount of dilution would be necessary if it were the only factor explaining the \(\sim 19.6\%\) difference in Neanderthal ancestry between Europe and East Asia, in contrast with recent work that inferred a smaller (\(\sim 9.4\%\)) contribution of basal Eurasians to modern European individuals\textsuperscript{27,28}.

Several confounding factors could impact our inference. Although it is unlikely that differential purifying selection is responsible for the discrepancy between European and East Asian Neanderthal ancestry\textsuperscript{22,23,26}, some Neanderthal ancestry was probably deleterious\textsuperscript{15,16} and our models assume neutrality. However, the strength of selection against introgressed fragments is probably small compared with the demographic forces at work; moreover, there is relatively little evidence of strong differences in the strength of selection between different non-African populations\textsuperscript{26,27}. To explore the impact of selection, we obtained the FFS from simulations of deleterious Neanderthal ancestry by Petr et al.\textsuperscript{18} and asked whether we classified their scenarios with selection as a two-pulse model using maximum likelihood. We found that we rejected a one-pulse model at the 5% level in only 1 out of 15 different simulations with selection, suggesting that we were unlikely to misclassify selection against Neanderthal ancestry as a two-pulse model.

Of additional concern was the power to detect fragments in each population. To address this, we implemented a model of fragment-calling errors (Supplementary Material). Based on simulations done by Steinrücken et al.\textsuperscript{19}, we expected false positive rates of approximately 0.1%, and false negative rates of approximately 1%; such rates do not cause substantial shifts in the FFS (Supplementary Fig. 6). Moreover, after extensive simulations, we found that the neural network trained with errors was robust to false positive fragment calls at a rate of 0.2%, and produced consistent results when applied to the real data (Supplementary Material). Finally, in an attempt to see inside the ‘black box’ of the FCNN, we examined how the weights propagated from each entry of the JFFS to the final assignments (Supplementary Material). In doing so, we found that moderate frequency haplotypes were most important in distinguishing between models (Supplementary Fig. 7), whereas errors in calling fragments were most likely to impact low-frequency haplotypes. This, combined with the fact that our results were robust across two different datasets and a range of cut-offs for determining archaic ancestry, convinced us that our results were robust to errors in fragment calling.

In addition, it is possible that some of the Neanderthal ancestry in East Asia has been misclassified, and in fact originated from Denisovan introgression. The misclassified archaic fragments could then mimic the signal of additional pulses of Neanderthal introgression. To address this concern, we inferred the position of Denisovan fragments based on data from Browning et al.\textsuperscript{28} and masked 1.6% of the positions across the genome (Methods). The masking resulted in 0.49% of sites called as Neanderthal introgression in the Steinrücken et al.\textsuperscript{15} data being removed from the introgression data we used in all analyses. Although we do not believe we removed all misclassified Denisova ancestry, we think it is unlikely that a substantial enough proportion remained to mimic the signal of additional Neanderthal pulses. These problems are likely to be further resolved as our ability to make accurate introgression calls for the various ancient human populations improves in the future.

Our work provides additional evidence for the ubiquity of archaic admixture in recent human history, consistent with recent work showing that humans interbred with Denisovans multiple times\textsuperscript{29}. Although we found that additional pulses of admixture in both East Asians and Europeans are necessary to explain the distribution of Neanderthal ancestry in Eurasia, we were unable to settle why East Asians have elevated Neanderthal ancestry. Interestingly, in contrast with Denisovans, there does not seem to be evidence of Neanderthal population structure within introgressed fragments\textsuperscript{30}. Combined with our results, this indicates that the Neanderthal population or populations that admixed with Eurasians must have been relatively closely related. This is consistent with the established inference of a long-term small effective size across Neanderthals\textsuperscript{29,31}, which has held up to scrutiny despite some claims of a larger Neanderthal effective size\textsuperscript{30,32}. Thus, we believe that a probable explanation for our results is that gene flow between humans and Neandertals was intermittent and ongoing, but in a somewhat geographically restricted region. Differential levels of admixture between different
Eurasian groups may primarily reflect how long those populations coexisted with Neanderthals in that region.

Methods

Data. We obtained the joint FFS by first downloading publicly available Neanderthal introgression calls from two sources \(^{1,9}\). The Sankararaman data \(^{10}\) consisted of the location of introgressed fragments along the genome in each phased haplotype for the 100 Genomes Project populations. The Steinrücken et al. data \(^{11}\) consisted of the probability of Neanderthal origin in 500-basepair windows of the genome across each phased haplotype for the Central European, Han Chinese and Southern Han Chinese individuals from the 100 Genomes Project.

We computed fragment frequencies by sampling a single site every 100 kb from both sources of data. To compute the joint FFS, we counted how many haplotypes were introgressed at each position. For the Steinrücken et al. data \(^{11}\), we called a site introgressed if it had a posterior probability of being introgressed above 45%. We then applied the 1000 Genomes accessibility mask (downloaded from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/accessible_genome_masks/20141020.pilot_mask_whole_genome.bed). In the Supplement, we show that our results are robust to the cut-off and consistent between both datasets (Supplementary Figs 4 and 8). We also masked Denisova fragments that were falsely called as Neanderthal by downloading the S’ fragment calls from Browning et al. \(^{8}\) and masking any fragment that matched Denisova > 85% and Neanderthal < 12%, resulting in removal of 1.6% of the genome overall and 0.49% of Neanderthal fragments. Given that the Denisova and Neanderthal populations diverged relatively early following the divergence of the Neanderthal/Denisova lineage and the modern human lineage, it is unsurprising that a relatively small fraction of Denisova haplotypes were falsely assigned as Neanderthal introgression. Finally, we masked the (0, 0), (0, 1), (1, 0) and (1, 1) position of the FFS matrix to reduce the impact of false negative and false positive fragment calls. For use in the FCNN, we projected the joint FFS down to 64 × 64 individuals (Fig. 2c,d).

Analytical model. We modelled introgression of the intensity \(f\) as injection of alleles at frequency \(q\) into the population at the time of introgression. In the one-pulse model, this resulted in an exact expression for the expected FFS under the Wright–Fisher diffusion model (Supplementary Material). Multiple pulse models could be solved analytically using a dynamic programming algorithm as in Kamm et al. \(^{12}\), but we instead approximated the expected frequency spectrum by making the approximation that the probability of sampling \(k\) introgressed haplotypes in a sample of size \(n + 1\) was the same as sampling \(k\) haplotypes in a sample of size \(n\) for large \(n\) (see Jouganous et al. \(^{33}\)). This resulted in closed-form expressions for the expected frequency spectrum under both the two-pulse and the dilution model (Supplementary Material). With an expected frequency spectrum given parameters \(\theta\) and model \(M\), \(p_{\theta, M}(\theta, M) = \sum_{k=0}^{n} \binom{n}{k} q^k (1-q)^{n-k} \), we computed the likelihood:

\[ L(\theta, M) = \prod_{n=1}^{N} x_n \log[p_{\theta, M}(\theta, M)] \]

where \(x_n\) is the number of fragments found in \(k\) out of \(n\) individuals. We optimized the likelihood using scipy, and compared models using the likelihood ratio statistic:

\[ A = 2(\log L(\theta_1, M_1) - \log L(\theta_2, M_2)) \]

where \(\theta_1\) and \(M_1\) correspond to the r-pulse model. Under the null, \(A\) should be \(\chi^2\) distributed with 2 d.f. (since there are two additional parameters in the two-pulse model). Simulations in the Supplementary Material suggest that \(P\) values under this model are well calibrated, despite the impact of linkage; thus, we opted to report nominal \(P\) values from our likelihood ratio test.

Simulations. We used msprime \(^{13}\) to simulate Neanderthal introgression into two modern populations with multiple potential admixture episodes and dilution from basal Eurasians (Fig. 3). For each replicate, we simulated the complete genomes for 170 European individuals and 394 East Asian individuals, matching the sampling available from the 1000 Genomes Project panel. We used the human recombination map (downloaded from http://www.well.ox.ac.uk/~amjali/AMap/; ref. \(^{34}\}). In each simulation, we mimicked our sampling scheme on the real data by sampling 1 site every 100 kb and calling a Neanderthal fragment by asking which individuals coalesced with the Neanderthal sample more recently than the human–Neanderthal population split time.

For each simulation, we drew demographic parameters, including effective population sizes and divergence times, from uniform distributions. For effective population sizes, we used 5,000–50,000 individuals for Neanderthals, 5,000–50,000 for Eurasians, and 5,000–100,000 for the European and East Asian populations. For divergence times, we used 12,000–26,000 generations for Neanderthals and humans, and 1,300–2,000 generations for the Eurasian split. The divergence between basal Eurasians and Eurasians was fixed at 3,000 generations. Lastly, we drew introgression times between 1,500 and 3,000 generations for gene flow into Eurasians, and between 800 and 2,000 generations for gene flow into the European and East Asian populations. The time for the introgression event between basal Eurasians and Europeans (dilution) was drawn from a uniform distribution of 200–2,000 generations.

To ensure that our simulations focused on the correct parameter space, we constrained the resulting amount of Neanderthal introgression in the modern European and East Asian genomes. The average Neanderthal ancestry \(a\) was drawn from a uniform distribution between 0.01 and 0.03, and the difference in ancestry \(d\) between the East Asian and European populations was drawn from a uniform distribution between 0 and 0.01. We then determined the introgression intensity given \(a\) and \(d\) (Supplementary Material).

Machine learning (FCNN). Using the resulting joint FFS, we trained a simple FCNN to categorize a joint FFS into one of five demographic models. The network was implemented in Keras (https://github.com/keras-team/keras) using a TensorFlow back end. The network used a simple architecture of three Dense layers (from 1,024 nodes to 512 nodes to 64 nodes), each followed by a dropout layer (0.20).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability. Our simulation pipeline is available at https://github.com/Villanea/Neanderthal_admix/blob/master/n_admix_10.py, which generates the simulated genomes, identifies Neanderthal introgression in 100-kb windows and applies the 1000 Genomes accessibility mask, and outputs an FFS and accompanying parameters for each replicate. The code used to implement the FCNN can be found at https://github.com/Villanea/Neanderthal_admix/blob/master/Fully_connected_network.py. Code for performing maximum-likelihood estimation of FFS parameters can be found at https://github.com/Villanea/Neanderthal_admix/blob/master/sym_stat_theory.py.

Data availability

No novel datasets were generated or analysed during the current study.

Received: 8 June 2018; Accepted: 18 October 2018; Published online: 26 November 2018

References

1. Green, R. et al. A draft sequence of the Neandertal genome. Science 328, 710–722 (2010).
2. Prüfer, K. et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science 358, 655–658 (2017).
3. Sankararaman, S., Patterson, N., Li, H., Pääbo, S. & Reich, D. The date of interbreeding between Neandertals and modern humans. PLoS Genet. 8, e1002947 (2012).
4. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovian individual. Science 338, 222–226 (2012).
5. Wall, J. D. et al. Higher levels of Neandertal ancestry in East Asians than in Europeans. Genetics 194, 199–209 (2013).
6. Karmin, M. et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. Genome Res. 25, 459–466 (2015).
7. Pržuljnik, G. D. et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat. Genet. 49, 593–599 (2016).
8. Skoglund, P. & Mathieson, I. Ancient genomics of modern humans: the first decade. Annu. Rev. Genom. Hum. Genet. 19, 381–404 (2018).
9. Fu, Q. et al. An early modern human from Romania with a recent Neandertal ancestor. Nature 524, 216–219 (2015).
10. Sankararaman, S. et al. The genomic landscape of Neandertal ancestry in present-day humans. Nature 507, 354–357 (2014).
11. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. Nature 536, 419–424 (2016).
12. Vernot, B. & Akey, J. M. Complex history of admixture between modern humans and Neandertals. Am. J. Hum. Genet. 96, 448–453 (2015).
13. Vernot, B. et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science 352, 235–239 (2016).
14. Harris, K. & Nielsen, R. The genetic cost of Neandertal introgression. Genetics 203, 881–891 (2016).
15. Kim, B. Y. & Lohmueller, K. E. Selection and reduced population size cannot explain higher amounts of Neandertal ancestry in East Asian than in European human populations. Am. J. Hum. Genet. 96, 454–461 (2015).
16. Juric, I., Aeschbacher, S. & Coop, G. The strength of selection against Neandertal introgression. PLoS Genet. 12, e1006340 (2016).
17. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513, 409–413 (2014).
18. Petz, M., Pääbo, S., Kelso, J. & Vernot, B. The limits of long-term selection against Neandertal introgression. Preprint at https://www.biorxiv.org/content/early/2018/07/04/362566 (2018).
19. Steinrücken, M., Spence, J. P., Kamm, J. A., Wieczorek, E. & Song, Y. S. Model-based detection and analysis of introgressed Neanderthal ancestry in modern humans. *Mol. Ecol.* **27**, 3873–3888 (2018).

20. Ronen, R., Udpa, N., Halperin, E. & Bafna, V. Learning natural selection from the site frequency spectrum. *Genetics* **195**, 181–193 (2013).

21. Schrider, D. R. & Kern, A. D. S/HIC: robust identification of soft and hard sweeps using machine learning. *PLoS Genet.* **12**, e1005928 (2016).

22. Sheehan, S. & Song, Y. S. Deep learning for population genetic inference. *PLoS Comput. Biol.* **12**, e1004845 (2016).

23. Schrider, D. R. & Kern, A. D. Supervised machine learning for population genetics: a new paradigm. *Trends Genet.* **34**, 301–312 (2018).

24. Bengio, Y. et al. in *Large-Scale Kernel Machines* (eds Bottou, L., Chapelle, O., DeCoste, D. & Weston, J.) 321–360 (MIT Press, Cambridge, 2007).

25. Kamm, J. A., Terhorst, J., Durbin, R. & Song, Y. S. Efficiently inferring the demographic history of many populations with allele count data. Preprint at https://www.biorxiv.org/content/early/2018/03/23/287268 (2018).

26. Do, R. et al. No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat. Genet.* **47**, 126–131 (2015).

27. Simons, Y. B., Turchin, M. C., Pritchard, J. K. & Sella, G. The deleterious mutation load is insensitive to recent population history. *Nat. Genet.* **46**, 220–224 (2014).

28. Browning, S. R., Browning, B. L., Zhou, Y., Tucci, S. & Akey, J. M. Analysis of human sequence data reveals two pulses of archaic Denisovan admixture. *Cell* **173**, 53–61 (2018).

29. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).

30. Mafessoni, F. & Prüfer, K. Better support for a small effective population size of Neandertals and a long shared history of Neandertals and Denisovans. *Proc. Natl Acad. Sci. USA* **114**, E10256–E10257 (2017).

31. Rogers, A. R., Bohlender, R. J. & Huff, C. D. Early history of Neandertals and Denisovans. *Proc. Natl Acad. Sci. USA* **114**, 9859–9863 (2017).

32. Rogers, A. R., Bohlender, R. J. & Huff, C. D. Reply to Mafessoni and Prüfer: Inferences with and without singleton site patterns. *Proc. Natl Acad. Sci. USA* **114**, E10258–E10260 (2017).

33. Jouganous, J., Long, W., Ragsdale, A. P. & Gravel, S. Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. *Genetics* **206**, 1549–1567 (2017).

34. Kelleher, J., Etheridge, A. M. & McVean, G. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput. Biol.* **12**, 1–22 (2016).

35. Hinch, A. G. et al. The landscape of recombination in African Americans. *Nature* **476**, 170–175 (2011).

**Acknowledgements**

We are grateful to S. Mathieson and J. Spence for several useful discussions about neural network architecture and appropriate methods for training neural networks. We also thank J. Spence and M. Steinrücken for extensive discussions on errors in fragment calling. I. Mathieson, S. Mathieson and J. Spence provided invaluable feedback on an early draft of this manuscript that helped improve its clarity. K. Harris provided invaluable discussions during the conception and work of this manuscript. We are grateful to M. Petr and B. Vernot for sharing processed simulation data with us, and for discussions about the impact of selection on Neanderthal ancestry.

J.G.S. and F.A.V. were supported by NIH grant R35 GM124745. This research was supported in part by the National Science Foundation through major research instrumentation grant number 1625061 for the Owl’s Nest high-performance cluster at Temple University.

**Author contributions**

F.A.V. and J.G.S. designed the study, analysed the data, performed the simulations and wrote the manuscript.

**Competition interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-018-0735-8.

Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.G.S.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2018
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted. Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
- Clearly defined error bars
- State explicitly what error bars represent (e.g. SD, SE, CI)

Software and code

Policy information about availability of computer code

Data collection

Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.

Data analysis

We used msprime (https://msprime.readthedocs.io/en/stable/) to generate coalescent data and keras (https://keras.io/) to perform neural network analyses. Custom code is available at www.github.com/villanea/Neandertal_admix/

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

No novel datasets were generated or analyzed during the current study.
Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences  ☐ Behavioural & social sciences  ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description
We used published call sets of Neandertal haplotypes to generate fragment frequency spectra between Europe and Asia. We used both novel maximum likelihood methodology and deep learning to make inferences based on the distribution of fragment frequencies.

Research sample
We used Neandertal fragment calls that were made on the 1000 genomes CEU and CHB/CHS individuals.

Sampling strategy
We used publicly available data, and hence our sample size of individuals was out of our control. We chose to analyze one site every 100 kilo base pairs to minimize the effects of linkage between introgressed fragments.

Data collection
Callsets were downloaded from the appropriate repositories described in the associated manuscripts.

Timing and spatial scale
n/a

Data exclusions
n/a

Reproducibility
We repeated the entire analysis using two different, semi-independent call sets of Neandertal fragments.

Randomization
n/a

Blinding
n/a

Did the study involve field work?  ☐ Yes  ☒ No

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a  Involved in the study

☒ Unique biological materials
☒ Antibodies
☒ Eukaryotic cell lines
☒ Palaeontology
☒ Animals and other organisms
☒ Human research participants

Methods

n/a  Involved in the study

☒ ChiP-seq
☒ Flow cytometry
☒ MRI-based neuroimaging