nature ecology & evolution

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41559-018-0735-8

In the format provided by the authors and unedited.

Multiple episodes of interbreeding between Neanderthal and modern humans

Fernando A. Villanea[1,2] and Joshua G. Schraiber[1,2]*

[1]Department of Biology, Temple University, Philadelphia, PA, USA. [2]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, USA. *e-mail: joshua.schraiber@temple.edu

Supplementary Material: Multiple episodes of interbreeding between Neanderthals and modern humans

Fernando A. Villanea and Joshua G. Schraiber

Compiled on September 26, 2018.

1 Analytical theory

1.1 One pulse model

An introgression of intensity f can be modeled as an injection of alleles at frequency f into a population. Each allele represents an introgressed haplotype, which will then undergo genetic drift until the present, at which time it is sampled at some (random) frequency. The Wright-Fisher diffusion model of genetic drift enables us to calculate the probability of sampling k out of n haplotypes as introgressed after drift by computing

$$p_{n,k}(t;f) = \int_0^1 \binom{n}{k} y^k (1-y)^{n-k} \phi(f,y;t) dy$$

where $\phi(f,y;t)$ is the probability that a haplotype has gone from frequency f to frequency y in $2N_e t$ generations. Using well known results [Ewens, 2012], we obtain a differential equation for the frequency dependent part, $\mu_{n,k}(t) \equiv \int_0^1 y^k (1-y)^{n-k} \phi(f,y;t) dy$,

$$\frac{d}{dt}\mu_{n,k} = \frac{k(k-1)}{2}\mu_{n,k-1} - k(n-k)\mu_{n,k} + \frac{(n-k)(n-k-1)}{2}\mu_{n,k+1}.$$

This is a linear system of differential equations and can be solved by matrix exponentiation. Thus,

$$p_{n,k}(t;f) = \binom{n}{k} e^{Qt}\mathbf{f}$$

where Q is the matrix of coefficients of the system of differential equations and $\mathbf{f} = ((1-f)^n, f(1-f)^{n-1}, \ldots, f^n)^T$. Note that Q is an $(n+1)\times(n+1)$ matrix, because it includes all haplotype frequencies from $k=0$ to $k=n$; however, it is only of rank n because $\sum_{k=0}^{n} p_{n,k}(t;f) = 1$. This approach is similar to that used in Kamm et al. [2018] and Jouganous et al. [2017].

1.2 Two pulse model

We can apply a similar logic to the one pulse model, and this time obtain an approximate formula. Working in a similar setting to before, we suppose that an admixture of intensity f_1 occurred, then t_1 generations more recently was followed by a second admixture of intensity f_2, which was t_2 generations more ancient than the present. Then, we want to evaluate the integral

$$\begin{aligned} p_{n,k}(t_1,t_2;f_1,f_2) &= \int_0^1 \int_0^1 \binom{n}{k} y^k (1-y)^{n-k} \phi(f_1,z;t_1)\phi(f_2+(1-f_2)z,y;t_2)dzdy \\ &= \int_0^1 \left(\int_0^1 \binom{n}{k} y^k (1-y)^{n-k} \phi(f_2+(1-f_2)z,y;t_2)dy \right) \phi(f_1,z;t_1)dz \end{aligned}$$

because we need to integrate over all possible allele frequencies at the time of the second pulse of admixture.

The internal integral is identical to the one pulse model, however, the initial allele frequency needs to be adjusted to $f_2+(1-f_2)z$. Thus, the result will be a linear combination of terms that look like $d_{n,k} = (f_2+(1-f_2)z)^k(1-f_2-(1-f_2)z)^{n-k}$ Thus, we need to derive a differential equation for

$$\eta_{n,k}(t) \equiv \int_0^1 (f_2+(1-f_2)z)^k (1-f_2-(1-f_2)z)^{n-k} \phi(f,z;t)dz.$$

Applying the Wright-Fisher generator to $d_{n,k}$, we get

$$\begin{aligned} \mathcal{L}d_{n,k} &= \frac{1}{2}x(1-x)\frac{d^2}{dx^2}d_{n,k} \\ &= \frac{1}{2}(1-f_2)^2 x(1-x)\left(k(k-1)d_{n-2,k-2} - 2k(n-k)d_{n-2,k-1}\right. \\ &\quad \left. +(n-k)(n-k-1)d_{n-2,k}\right). \end{aligned}$$

Now, put $D = k(k-1)d_{n-2,k-2} - 2k(n-k)d_{n-2,k-1} + (n-k)(n-k-1)d_{n-2,k}$, and write

$$\begin{aligned} (1-f_2)^2 x(1-x)D &= (f_2+(1-f_2)x-f)(1-f_2-(1-f_2)x)D \\ &= (f_2+(1-f_2)x)(1-f_2-(1-f_2)x)D - f_2(1-f_2-(1-f_2)x)D. \end{aligned}$$

Now, the first term looks like

$$k(k-1)d_{n,k-1} - 2k(n-k)d_{n,k} + (n-k)(n-k-1)d_{n,k+1},$$

which is the same as the one pulse model. However, the second term will be

$$k(k-1)d_{n-1,k-2} - 2k(n-k)d_{n-1,k-1} + (n-k)(n-k-1)d_{n-1,k}.$$

Note that the second term is *not* the same as the first term with $n \mapsto n-1$. However, we will apply the approximation, that $d_{n,k} \approx d_{n-1,k}$ for large n, Thus, we have

$$\begin{aligned}\frac{d}{dt}\eta_{n,k} = &\frac{k(k-1)}{2}\eta_{n,k-1} - k(n-k)\eta_{n,k} + \frac{(n-k)(n-k-1)}{2}\eta_{n,k+1} \\ &- f_2\left(\frac{k(k-1)}{2}\eta_{n,k-2} - k(n-k)\eta_{n,k-1} + \frac{(n-k)(n-k-1)}{2}\eta_{n,k}\right).\end{aligned}$$

The first line is simply the same differential equation as the one pulse case, while the second line is shifted down one term. Defining the matrix corresponding to that differential equation as Q_m, we see that

$$p_{n,k}(t_1, t_2; f_1, f_2) \approx \binom{n}{k} e^{Qt_2} e^{(Q - f_2 Q_m)t_1} \mathbf{f}_t$$

where $\mathbf{f}_t = ((1-f_2-(1-f_2)f_1)^n, (f_2+(1-f_2)f_1)(1-f_2-(1-f_2)f_1)^{n-1}, \ldots, (f_2+(1-f_2)f_1)^n)^T$.

1.3 Dilution model

Under this model, an admixture of intensity f_1 occurs, then t_1 generations more recently, an unadmixed group contributes to the population at hand with intensity f_2 t_2 generations in the past. Again, we can write down an integral to solve,

$$\begin{aligned}p_{n,k}(t_1, t_2; f_1, f_2) &= \int_0^1 \int_0^1 \binom{n}{k} y^k (1-y)^{n-k} \phi(f_1, z; t_1)\phi((1-f_2)z, y; t_2) dz dy \\ &= \int_0^1 \left(\int_0^1 \binom{n}{k} y^k (1-y)^{n-k} \phi((1-f_2)z, y; t_2) dy\right) \phi(f_1, z; t_1) dz.\end{aligned}$$

Again, the internal integral is the same as the one pulse model, except that the initial allele frequency is $(1-f_2)z$. Evidently, the result that integral will be a function of terms $c_{n,k} = ((1-f_2)z)^k (1-(1-f_2)z)^{n-k}$, so we need to solve integrals of the form

$$\nu_{n,k} \equiv \int_0^1 ((1-f_2)z)^k (1-(1-f_2)z)^{n-k} \phi(f_1, z; t_1) dz.$$

Applying the generator of the Wright-Fisher diffusion to the function $c_{n,k}$ we see that

$$\begin{aligned}\mathcal{L}c_{n,k} &= \frac{1}{2}x(1-x)\frac{d^2}{dx^2}c_{n,k} \\ &= \frac{1}{2}(1-f)^2 x(1-x)\left((k(k-1)c_{n-2,k-2} - 2k(n-k)c_{n-2,k-1}\right. \\ &\quad \left. + (n-k)(n-k-1)c_{n-2,k}\right).\end{aligned}$$

Let $C = k(k-1)c_{n-2,k-2} - 2k(n-k)c_{n-2,k-1} + (n-k)(n-k-1)c_{n-2,k}$, then note that

$$\begin{aligned}(1-f)^2x(1-x)C &= ((1-f)x)(1-(1-f)x-f)C \\ &= ((1-f)x)(1-(1-f)x)C - f((1-f)x)C.\end{aligned}$$

Multiplying through, we see that the first term looks like

$$k(k-1)c_{n,k-1} - 2k(n-k)c_{n,k} + (n-k)(n-k-1)c_{n,k+1},$$

while the second term will be

$$k(k-1)c_{n-1,k-1} - 2k(n-k)c_{n-1,k} + (n-k)(n-k-1)c_{n-1,k+1},$$

i.e. it is the same except with $n \mapsto n-1$. Making an approximation that $c_{n,k} \approx c_{n-1,k}$ for large n, we can pull out a factor of $(1-f_2)$ and obtain a system of differential equations,

$$\frac{d}{dt}\nu_{n,k} \approx (1-f_2)\left(\frac{k(k-1)}{2}\nu_{n,k-1} - k(n-k)\nu_{n,k} + \frac{(n-k)(n-k-1)}{2}\nu_{n,k+1}\right).$$

Noting that this is essentially the same differential equation as the one pulse model, we have that

$$\begin{aligned}p_{n,k}(t_1,t_2;f_1,f_2) &\approx \binom{n}{k}e^{Qt_2}e^{(1-f_2)Qt_1}\mathbf{f}_d \\ &= \binom{n}{k}e^{Q((1-f_2)t_1+t_2)}\mathbf{f}_d\end{aligned}$$

where now $\mathbf{f}_d = ((1-(1-f_2)f_1)^n, ((1-f_2)f_1)(1-(1-f_2)f_1)^{n-1}, \ldots, ((1-f_2)f_1)^n)^T$.

Note that this surprisingly simple form suggests that a dilution can be understood as an admixture of intensity $(1-f_2)f_1$ occurring $(1-f_2)t_1 + t_2$ generations ago.

2 Error model

2.1 Single population

To incorporate false negative and false positive calls into our model, assume that there are independent false negative and false positive calls with rates ϵ_+ and ϵ_-, respectively. Specifically, every individual that has a fragment is called negative independently with probability ϵ_- and every individual that doesn't is called positive with probability ϵ_+. Define $b(k;N,p)$ to the probability mass function of a binomial random variable with size N and probability p. Also define $f(k;N_1,N_2,p_1,p_2)$ to be the distribution of the difference of two binomial random variables. Then we have that

$$f(k;N_1,N_2,p_1,p_2) = \sum_{i=0}^{N_2} b(k+i;N_1,p_1)b(i;N_2,p_2)$$

by a simple argument. Thus, we have that the probability of observing a fragment at frequency k with error is

$$\tilde{p}_{n,k} = \sum_{i=0}^{n} f(k-i; n-i, i, \epsilon_+, \epsilon_-) p_{n,i}$$

because if we have i introgressed fragments, we independently take false positives out of the $n-i$ non-introgressed fragments and false negatives out of the i introgressed fragments. If the number of false positives minus false negatives is d, then we have $i+d$ total fragments after errors, and thus need $d = k-i$ to end up with exactly k fragments. We then sum over all i.

To quantify the impact of errors in calling fragments, we generated an expected FFS under a 1 pulse model with $f = 0.02$ and $t = 0.1$ diffusion time units. We then computed the Kullback-Leibler divergence between the true FFS and the an FFS with a given false positive and false negative rate. Supplementary Figure 6 shows that, for low amounts of admixture as we simulated here, the impact of false positives is much larger than that of false negatives, due largely to the fact that most of the genome is a true negative. Nonetheless, even with relatively high false negative and false positive rates, such as $\epsilon_- = 0.1$ and $\epsilon_+ = 0.01$ (far higher than the rates seen in simulations from Steinrücken et al. [2018]), the Kullback-Leibler divergence is only ~ 0.005, indicating that false positives and false negatives do not have a substantial effect on the FFS.

2.2 Two populations

We extended the error model to joint fragment frequency spectra by applying it to each row and column of the JFFS. Specifically, letting p_{n_1,n_2,k_1,k_2} represent the entry of the JFFS corresponding to frequency k_1 out of n_1 in population 1 and k_2 out of n_2 frequency in population 2, we first compute

$$\hat{p}_{n_1,n_2,k_1,k_2} = \sum_{i=0}^{n_1} f(k_1-i; n_1-i, i, \epsilon_+, \epsilon_-) p_{n_1,n_2,i,k_2}$$

by modeling error along one axis, and then compute the probability of observing a fragment with error

$$\tilde{p}_{n_1,n_2,k_1,k_2} = \sum_{j=0}^{n_2} f(k_2-j; n_2-j, j, \epsilon_+, \epsilon_-) \hat{p}_{n_1,n_2,k_1,j}.$$

This formula can easily be adjusted to have population specific error rates.

3 Implementation of maximum likelihood method

3.1 Numerical aspects

We implemented Python software to calculate and optimize the likelihoods. Specifically, we have functions to compute the likelihood under the one and two pulse models and include the error model. We used `scipy.sparse.linalg.expm_multiply` to compute the sparse matrix exponentials, and optimized the likelihood using `scipy.optimize.fmin_l_bfgs_b`. Code for implementing the model is available at `https://github.com/Villanea/Neanderthal_admix/blob/master/sym_stat_theory.py`

3.2 Validation of maximum likelihood method

We evaluated the performance of the maximum likelihood method by subsampling 200 simulations per model from the simulations done to train the fully connected neural network. We then computed the maximum likelihood parameter estimates under both the one pulse and the two pulse model. We then performed a likelihood ratio test with 2 degrees of freedom to see if we could reject the one pulse model at the 5% level. The results are summarized in Supplementary Table 1.

Model	Pop	Fraction rejecting one pulse
One pulse	ASN	0.027
One pulse	EUR	0.071
Two pulse	ASN	**0.17**
Two pulse	EUR	0.071
Three pulse	ASN	**0.29**
Three pulse	EUR	**0.23**
Dilution	ASN	0.031
Dilution	EUR	0.031
All pulse	ASN	**0.21**
All pulse	EUR	**0.19**

Supplementary Table 1: False positive rates and power of the maximum likelihood method. The first column indicates the model used for simulation, the second column the population analyzed, and the third column shows the fraction of simulations that rejected the one pulse model at the 5% level. Bolded numbers indicate situations where we expect the one pulse model to be rejected in favor of the two pulse model.

From these results, we conclude that we achieve roughly the expected 5% false positive rate at the 5% level and find that we have ∼20% power. However, we note that the simulations cover an extremely wide range of parameters, including many cases where

Supplementary Table 2: Parameter estimates from the Asian data. Each row corresponds to a different probability cutoff for calling fragments from the Steinrücken dataset. The columns are as follows: f_ij indicates the intensity of admixture pulse i under a model with j pulses, t_ij indicates the time (in diffusion units) after pulse i before the next event in a model with j pulses, FPR_i indicates the inferred false positive rate under a model with i pulses, FNR_i indicates the inferred false negative rate under a model with i pulses, lnL_i indicates the negative log likelihood under a model with i pulses, and lambda indicates the likelihood ratio statistic.

Supplementary Table 3: Parameter estimates from the Europe data. Columns and rows are the same as in Table 2

the second pulse is very small. Thus, we believe that power is much higher for realistic parameter values.

4 Results from maximum likelihood fitting

Supplementary Table 2 and 3 shows the parameter estimates from maximum likelihood fitting of the Asian and European data, respectively, across a variety of cutoffs from the Steinrücken data.

5 Admixture constraints

Given European and East Asian mixture proportions, f_{EUR} and f_{ASN}, respectively, we constrain mixture proportions by constraining

$$a = \frac{f_{\text{ASN}} + f_{\text{EUR}}}{2}$$

and

$$d = f_{\text{ASN}} - f_{\text{EUR}}$$

we then express f_{ASN} and f_{EUR} in terms of the model parameters, and solve for model parameters that will adhere to the constraints.

In the one pulse model, f_1 is a single pulse of Neandertal introgression into the ancestral Eurasian population. So,

$$f_{\text{ASN}} = f_1$$

and

$$f_{\text{EUR}} = f_1.$$

Thus, we see that

$$f_1 = a$$

In the two pulse model, f_1 is a first pulse of Neandertal introgression into the ancestral Eurasian population, and f_2 is a second pulse of introgression into the East Asian population. This results in

$$f_{\text{ASN}} = f_2 + (1 - f_2)f_1$$

and

$$f_{\text{EUR}} = f_1,$$

which yields

$$f_1 = a - d/2$$

and

$$f_2 = \frac{d}{1 + d/2 - a}.$$

For the three pulse model, f_1 is a first pulse of Neandertal introgression into the ancestral Eurasian population, f_2 is a second pulse of introgression into the East Asian population, and f_3 is a second pulse of introgression into the European population. Thus,

$$f_{\text{ASN}} = f_2 + (1 - f_2)f_1$$

and

$$f_{\text{EUR}} = f_3 + (1 - f_3)f_1,$$

Note that in this model, we have more free parameters than constraints, so we sample f_1 from a uniform distribution between 0 and $a - d/2$, and then solve to obtain

$$f_2 = \frac{a + d/2 - f_1}{1 - f_1}$$

and

$$f_3 = \frac{a - d/2 - f_1}{1 - f_1}.$$

For the dilution model, f_1 is a single pulse of Neandertal introgression into the ancestral Eurasian population, and f_4 represents the dilution from the Basal Eurasian population into the European population. This yields admixture proportions

$$f_{\text{ASN}} = f_1$$

and

$$f_{\text{EUR}} = (1 - f_4)f_1,$$

resulting in

$$f_1 = a + d/2$$

and

$$f_4 = \frac{d}{a + d/2}$$

In the model with 3 pulses of Neandertal admixture and dilution, f_1 is a first pulse of Neandertal introgression into the ancestral Eurasian population, f_2 is a second pulse of introgression into the East Asian population, and f_3 is a second pulse of introgression into the European population, while f_4 represents the dilution from the Basal Eurasian population into the European population. Further, we always assume that dilution is more recent than the second pulse in Europe. In this case, admixture proportions are

$$f_{\text{ASN}} = f_2 + (1 - f_2)f_1$$

and

$$f_{\text{EUR}} = (f_3 + (1 - f_3)f_1)(1 - f_4)$$

Again, we have more free parameters than constraints so we first draw f_1 from a uniform distribution between 0 and $a - d/2$ and f_4 from a uniform distribution between 0 and 0.5. Then, we set

$$f_2 = \frac{a + d/2 - f_1}{1 - f_1}$$

and

$$f3 = \frac{a - d/2 + f_1(1 - f_4)}{(1 - f_1)(1 - f_4)}$$

6 Neural Network Weights

In order to quantify the impact of different frequency spectrum categories to the inferences of the FCNN, we computed the matrix product of the weights across the fully connected layers. Note that we flatten our input FFS matrix to a single vector of length $m_1 = 64 \times 64 = 4096$ initially. Then, we compute the weighted sum of the weights leading to the nodes in the subsequent layers of the FCNN. Specifically, if layer i has m_i nodes, and w_i is the $m_i \times m_{i+1}$ matrix where $w_{i,j,k}$ provides the weights from node j in layer i to node k in layer $i + 1$, then we compute the matrix product

$$M = w_i w_{i+1} \cdots w_n$$

when we have n layers. The resulting matrix M will be 4096×5, with each of the 5 columns corresponding to one of the different models. Thus, we map each of the columns back into the original 64×64 matrix, resulting in the panels shown in Supplementary Figure 7.

This figure shows the signals that the FCNN extracts from the data to distinguish the models from the "average" simulated dataset. For instance, compared to all the other models which have an excess of Neandertal ancestry in East Asia, the FCNN identifies

the 1 pulse model by those that have relatively less Neandertal ancestry in East Asia (as can be seen by the red blob along the y axis) and relatively more Neandertal ancestry in Europe (as can be seen by the yellow blob on the x axis). Similarly, models with additional pulses can be identified by relatively less shared low frequency fragments and relatively more private moderate frequency fragments.

7 Robustness of FCNN results

The Neandertal fragment calls from Steinrücken et al. [2018] are based on the posterior probability of introgression at each position estimated using diCal-admix. Because each position has a different probability of introgression, defining a global cut-off is necessary to obtain a consensus across the genome, such that a higher cut-off results in a higher certainty of the calls (i.e. higher precision), but more false negatives (i.e. lower recall). For the analysis presented in the main text, we used a cut-off of 0.45, recommended in Steinrücken et al. [2018] as it provides the best balance across performance metrics based on their precision-recall curves. However, to test the robustness of the selected cut-off, we generated FFS based on a range of cut-offs and analyzed them using the fully-connected neural network. Supplementary Figure 4 shows that our results are consistent across the entire range of cutoffs. Likewise, we explored the relation between higher false positive rate (i.e. lower posterior probability cutoffs in the Steinrucken data) and the reported fraction of introgression for each population. We find that the relative enrichment in East Asia is constant across a wide range of false positive rates (Supplementary Figure 5).

In addition, we had access to the introgressed fragments calls from Sankararaman et al. [2014], which were ascertained independently using a conditional random field method. We converted these fragment calls into introgressed site calls by looking at the same positions every 100kb used for the Steinrucken data, and counting how many individuals presented an introgressed fragment which overlapped with that site. We used these fragment calls as independent confirmation of all results (Supplementary Figure 8).

8 Error model implemented on FCNN

False positive error on the introgressed Neandertal fragment calls could mimic the signal for secondary pulses of admixture, in particular because most of the genome is a true negative, false positives are likely to be low frequency, mimicking more recent introgression. In order to test if false positive errors resulted in the signal of secondary gene flow we observe, we extended the error model into FFS 2D matrix. We then trained the FCNN classifier using data with errors. Specifically, we trained two different situations: one in which there are symmetric errors with a false positive rate of 0.14% and a false negative rate of 1% in both Europe and Asia, as well as an asymmetrical model in which false positives only occur in East Asia. For each model, we provided it with training data consisting of half data with

no errors and half data with errors.

We found that our symmetric error classifier performed well on both data with errors and data without errors (Supplementary Figure 9a). To test the robustness of our symmetric error classifier to false positive error that it wasn't trained with, we provided it with test data including 0%, 0.1%, 0.2%, and 1% false positive error rates into both Europe and East Asia populations based on FFS created under model 1 (equal true Neandertal ancestry). Figure 9b shows that while adding false positive errors to FFS simulated under the 1 pulse model had the predicted effect of mimicking the signal of the 3 pulse and all pulse models, the effect is not noticeable at 0.1% or 0.2% error rates, and only becomes problematic at 1%. Note that 1% false positive errors represents an extreme case, considering the average Neandertal introgression fraction in our data is around $\sim$1.5% (Supplementary Figure 9c). A 1% false positive rate would indicate that the vast majority of the calls are false positives, which we believe is unlikely given the concordance between the Steinrucken and Sankaraman datasets. We then used the FCNN classifier trained with false positive error to classify the Steinrucken empirical data, finding similar results to those when we used the FCNN trained without error (Supplementary Figure 9d).

An unlikely, but possible scenario is that the enrichment of Neandertal ancestry observed in East Asian individuals is entirely the result of false positive errors found only on this population. We tested this using our FCNN trained with only false positive errors in East Asia. The newly trained FCNN performs comparably to the previous error model (Supplementary Figure 10a). We then again explored robustness to error rates that weren't used to train the classifier by incorporating 0%, 0.1%, 0.2%, and 1% false positive error rates into the East Asia population only, in FFS created under model 1 (equal true Neandertal ancestry). Adding false positive error to the East Asia population had the predicted effect of mimicking the signal of the 2 pulse model (Supplementary Figure 10b). However, similar to the previous symmetric error classifier, the effect is not noticeable at 0.1% or 0.2% error rates (corresponding approximately to the elevation in Neandertal ancestry in East Asia), and only becomes problematic at 1%, which represents the extreme case (Supplementary Figure 10c). Once more, we used the FCNN classifier trained with false positive error to classify the Steinrucken empirical data, finding similar results as in the error-free and the symmetric error models (Supplementary Figure 10d).

References

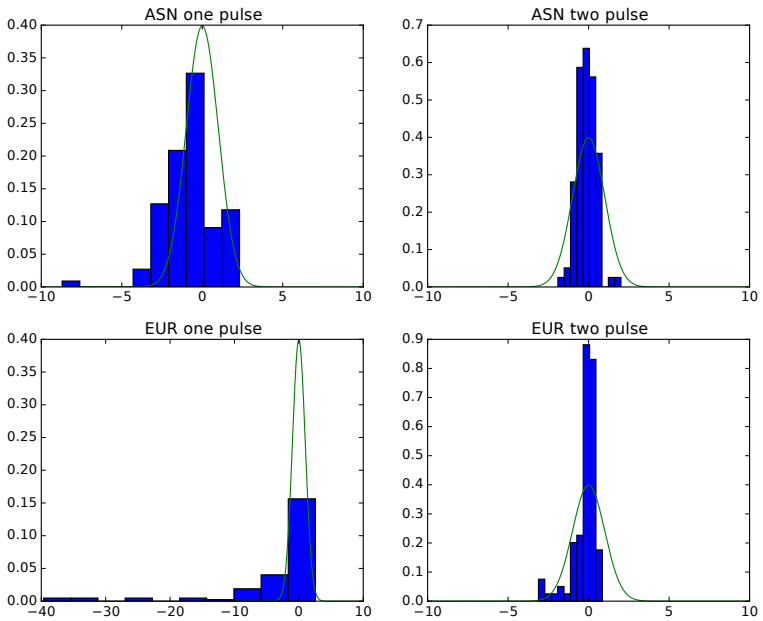
Warren J Ewens. *Mathematical population genetics 1: theoretical introduction*, volume 27. Springer Science & Business Media, 2012.

Julien Jouganous, Will Long, Aaron P Ragsdale, and Simon Gravel. Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. *Genetics*, pages genetics–117, 2017.

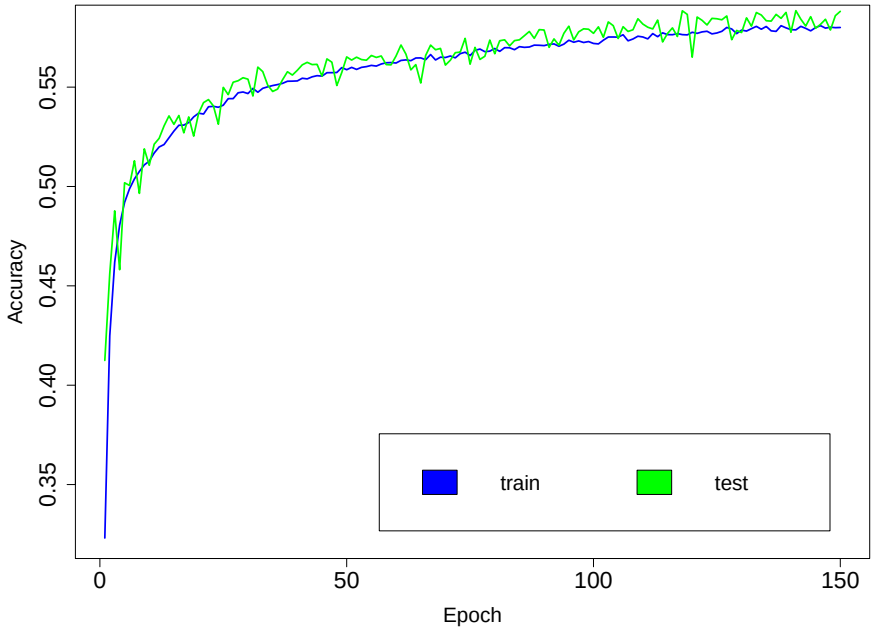
John A. Kamm, Jonathan Terhorst, Richard Durbin, and Yun S. Song. Efficiently inferring the demographic history of many populations with allele count data. *bioRxiv*, 2018. doi: 10.1101/287268. URL `https://www.biorxiv.org/content/early/2018/03/23/287268`.

Sriram Sankararaman, Swapan Mallick, Michael Dannemann, Kay Prüfer, Janet Kelso, Svante Pääbo, Nick Patterson, and David Reich. The genomic landscape of neanderthal ancestry in present-day humans. *Nature*, 507(7492):354, 2014.

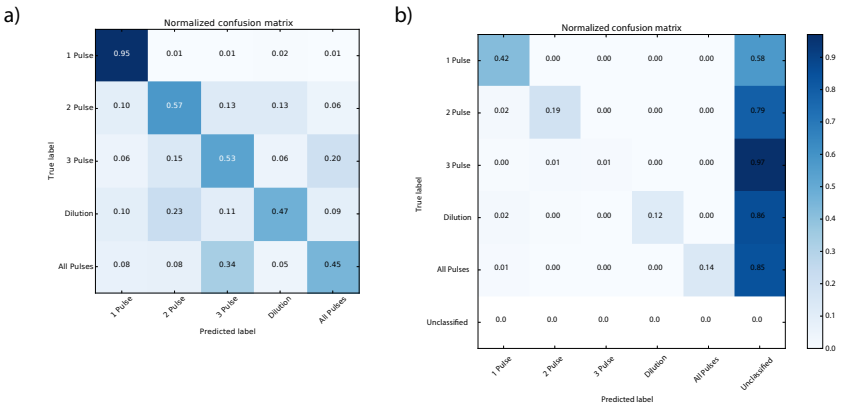
Matthias Steinrücken, Jeffrey P. Spence, John A. Kamm, Emilia Wieczorek, and Yun S. Song. Modelbased detection and analysis of introgressed neanderthal ancestry in modern humans. *Molecular Ecology*, 2018. doi: 10.1111/mec.14565.



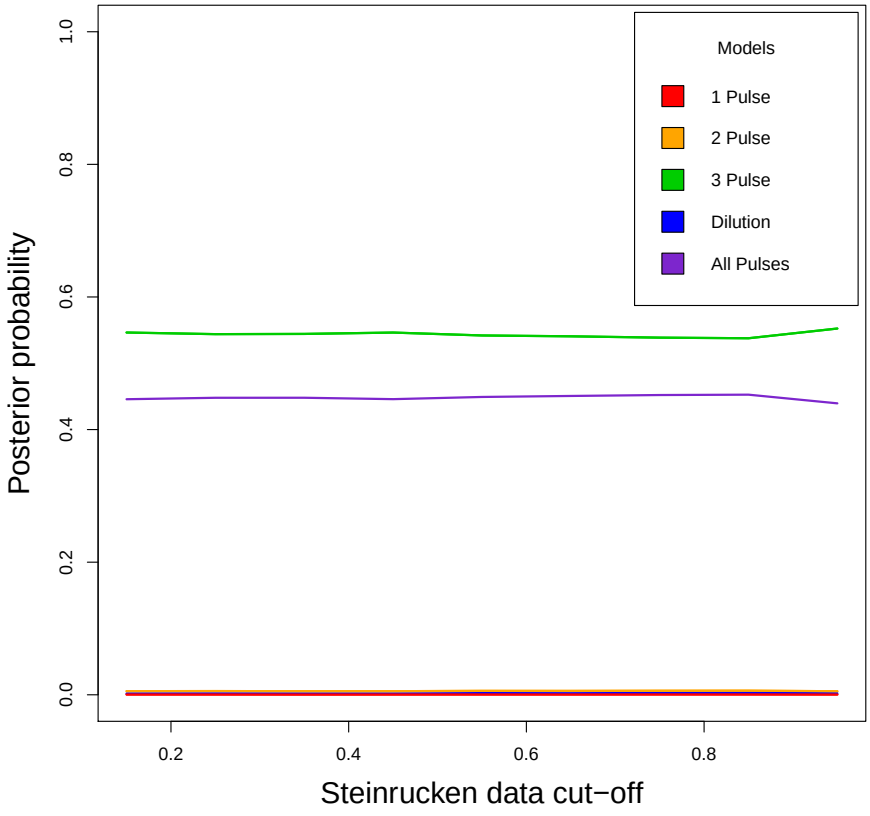
Supplementary Figure 1: Residuals from fitting the maximum likelihood model. Each panel corresponds to the indicated model, while the histogram shows the empirical residuals. The line shows a Normal(0,1) distribution.



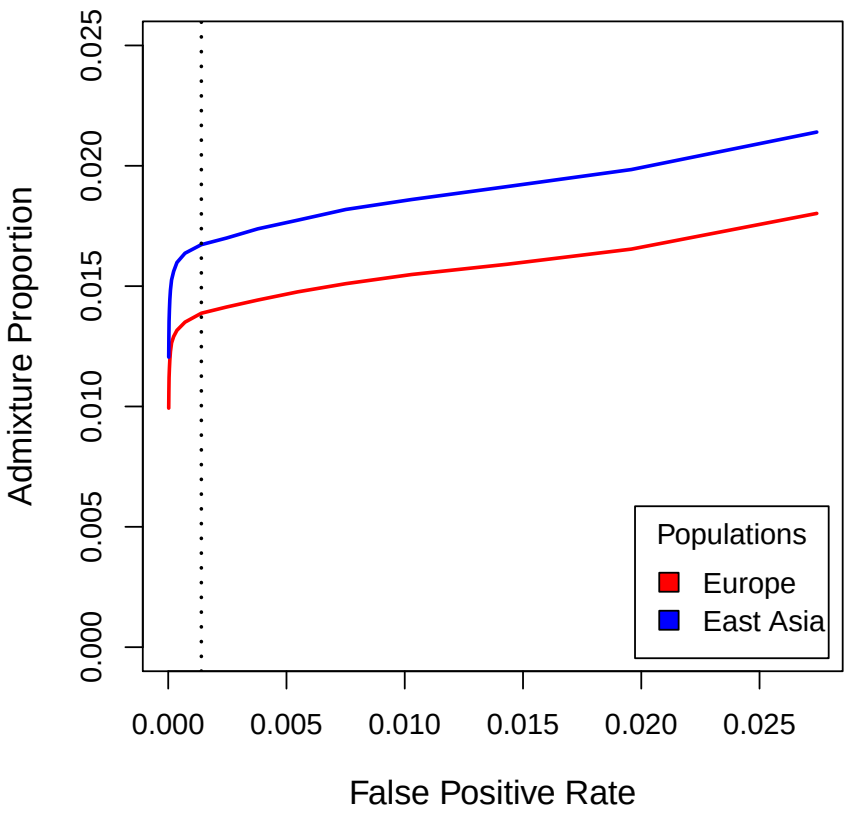
Supplementary Figure 2: Training and validation accuracy of FCNN as it trained over 150 epochs. The x axis indicates the training epoch (i.e. one pass through the whole dataset) while the y-axis shows the accuracy on either training data (blue) or validation data (green).



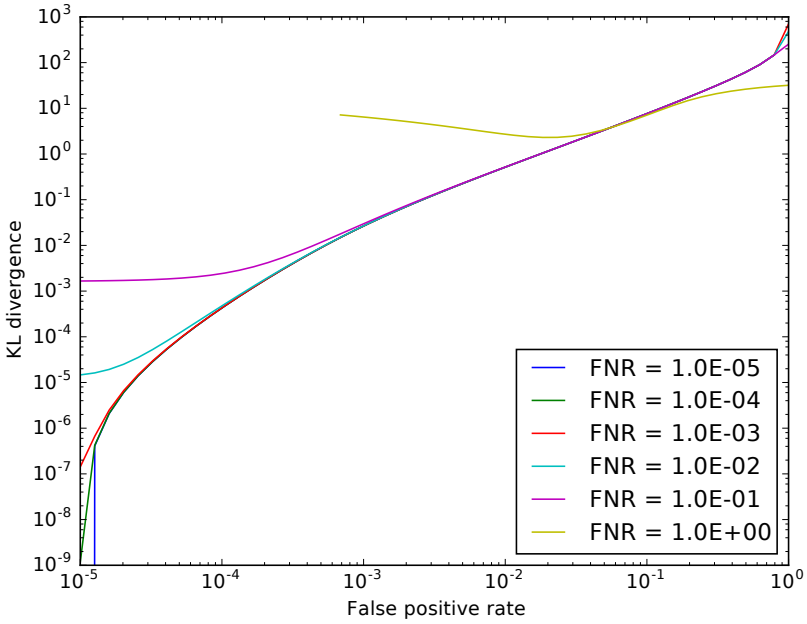
Supplementary Figure 3: a) Confusion matrix of simulated data categorized by the FCNN. b) Confusion matrix of simulated data categorized by the FCNN when accepting results with a probability cut-off of 0.8.



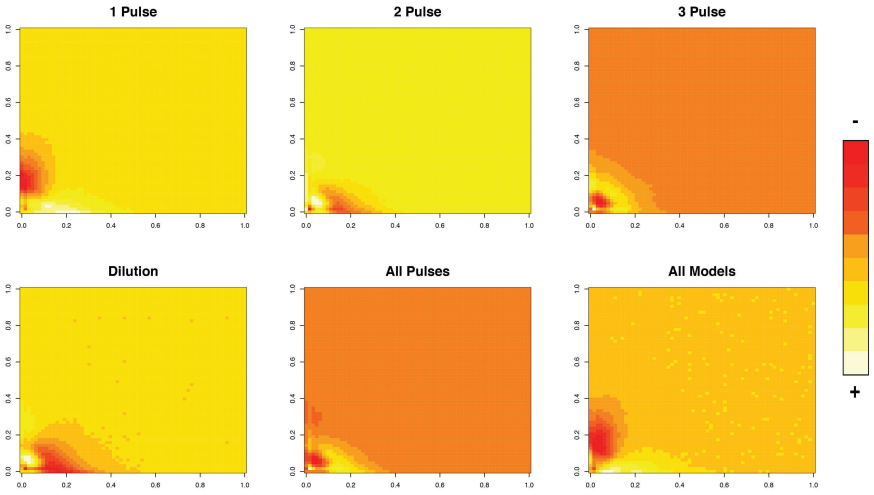
Supplementary Figure 4: Posterior probability of the empirical introgression data from the FCNN classifier under different cut-offs of the posterior probability of introgression in the Steinrücken et al. [2018] data. The x axis indicates the posterior probability cutoff, and the y axis the model probability according to the FCNN. Each line corresponds to a different model.



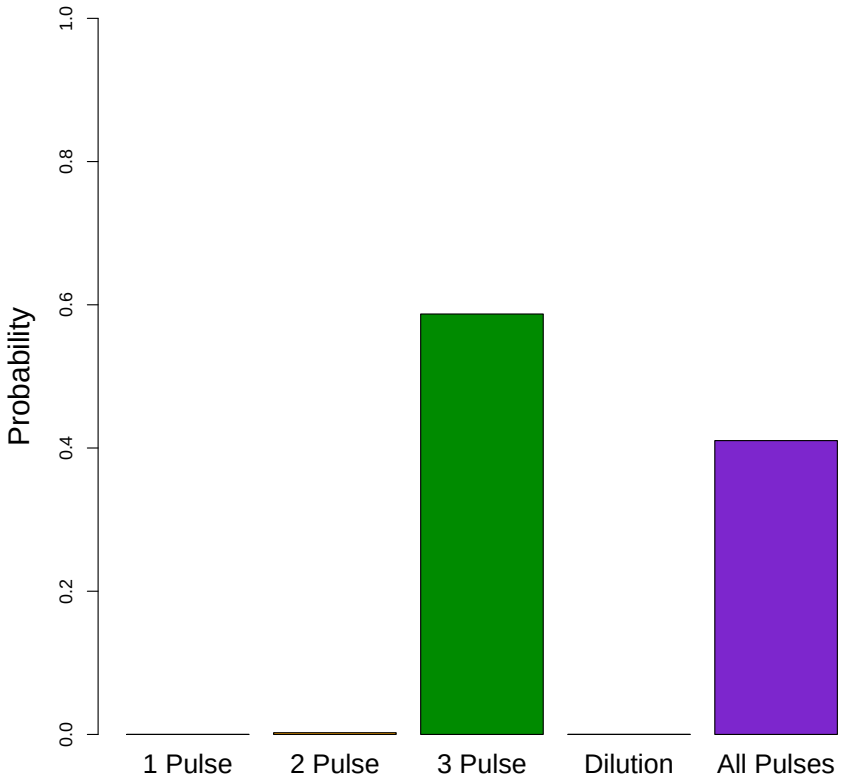
Supplementary Figure 5: Neandertal admixture proportion in East Asia and Europe relative to the false positive rate for the Steinrucken data, obtained from the precision-recall curves provided in Steinrücken et al. [2018]. The dotted line marks the expected false positive rate at the cut-off used in this study (Posterior probability >0.45, corresponding to FPR = 0.0014).



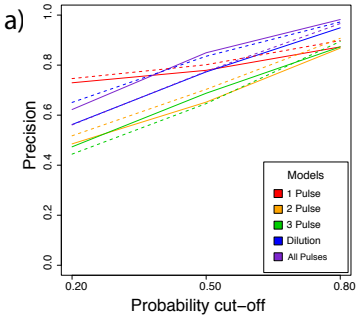
Supplementary Figure 6: The impact of false positive and false negative fragment calls on the FFS. The x axis shows the false positive rate, and the y axis the Kullback-Liebler divergence of the observed FFS to the true FFS (larger values indicate more difference). Each line corresponds to a different false negative rate.



Supplementary Figure 7: Weights projected across layers into the final dense layer, representing the relative importance of each position along the FFS matrix when classifying a FFS into one of the five final categories.

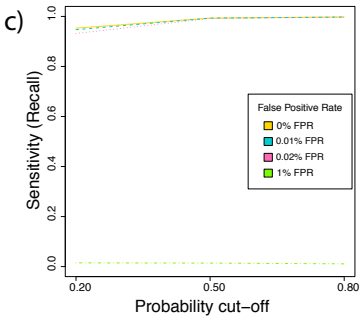


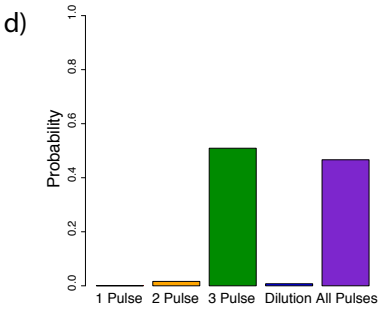
Supplementary Figure 8: Posterior probability of the empirical introgression data from Sankararaman et al., 2014 matching each of the five demographic models, determined by the FCNN classifier.



b)

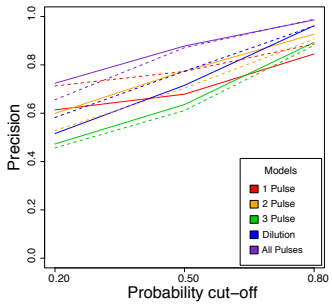
F. P. E.	TRUE MODEL	INCORRECT MODELS			
RATE	1 PULSE	2 PULSE	3 PULSE	DILUTION	ALL PULSES
0%	953	18	9	11	9
0.1%	947	16	13	10	14
0.2%	932	18	19	7	24
1%	15	0	557	0	428





Supplementary Figure 9: a-d) Results of training the FCNN classifier using FFS including false positive error into both the East Asia and European populations. a) Posterior probability that the chosen model is correct (precision), for all models under different levels of support for the chosen model. The x axis shows the probability cutoff that we used to classify models, and the y axis shows the precision. Each line corresponds to a different model. Solid lines correspond to FSS simulated with no error, and dashed lines correspond to FFS simulated with false positive error. b) Results of the test to determine the false positive error would be confused with a signal of secondary admixture. c) Sensitivity of the FCNN classifier when incorporating various rates of false positive error into both populations. This shows the probability that the one pulse model is chosen given that the data was simulated under the one pulse model. d) Posterior probability of the Steinrucken et al. 2018 empirical introgression data matching each of the five demographic models, determined by the FCNN classifier.

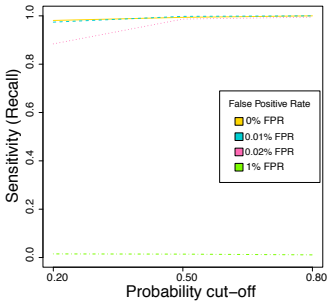
a)



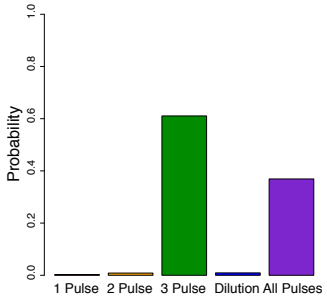
b)

F. P. E.	TRUE MODEL	INCORRECT MODELS			
RATE	1 PULSE	2 PULSE	3 PULSE	DILUTION	ALL PULSES
0%	981	4	1	6	8
0.1%	974	6	2	12	6
0.2%	884	48	31	19	18
1%	0	615	238	5	142

c)



d)



Supplementary Figure 10: a-d) Results of training the FCNN classifier using FFS including false positive error into only the East Asia population. a) Posterior probability that the chosen model is correct (precision), for all models under different levels of support for the chosen model. The x axis shows the probability cutoff that we used to classify models, and the y axis shows the precision. Each line corresponds to a different model. Solid lines correspond to FSS simulated with no error, and dashed lines correspond to FFS simulated with false positive error. b) Results of the test to determine the false positive error would be confused with a signal of secondary admixture. c) Sensitivity of the FCNN classifier when incorporating various rates of false positive error into the East Asia population. This shows the probability that the one pulse model is chosen given that the data was simulated under the one pulse model. d) Posterior probability of the Steinrucken et al. 2018 empirical introgression data matching each of the five demographic models, determined by the FCNN classifier.