The American Journal of Human Genetics

Supplemental Data

Selection and Reduced Population Size Cannot Explain Higher Amounts of Neandertal Ancestry in East Asian than in European Human Populations

Bernard Y. Kim and Kirk E. Lohmueller

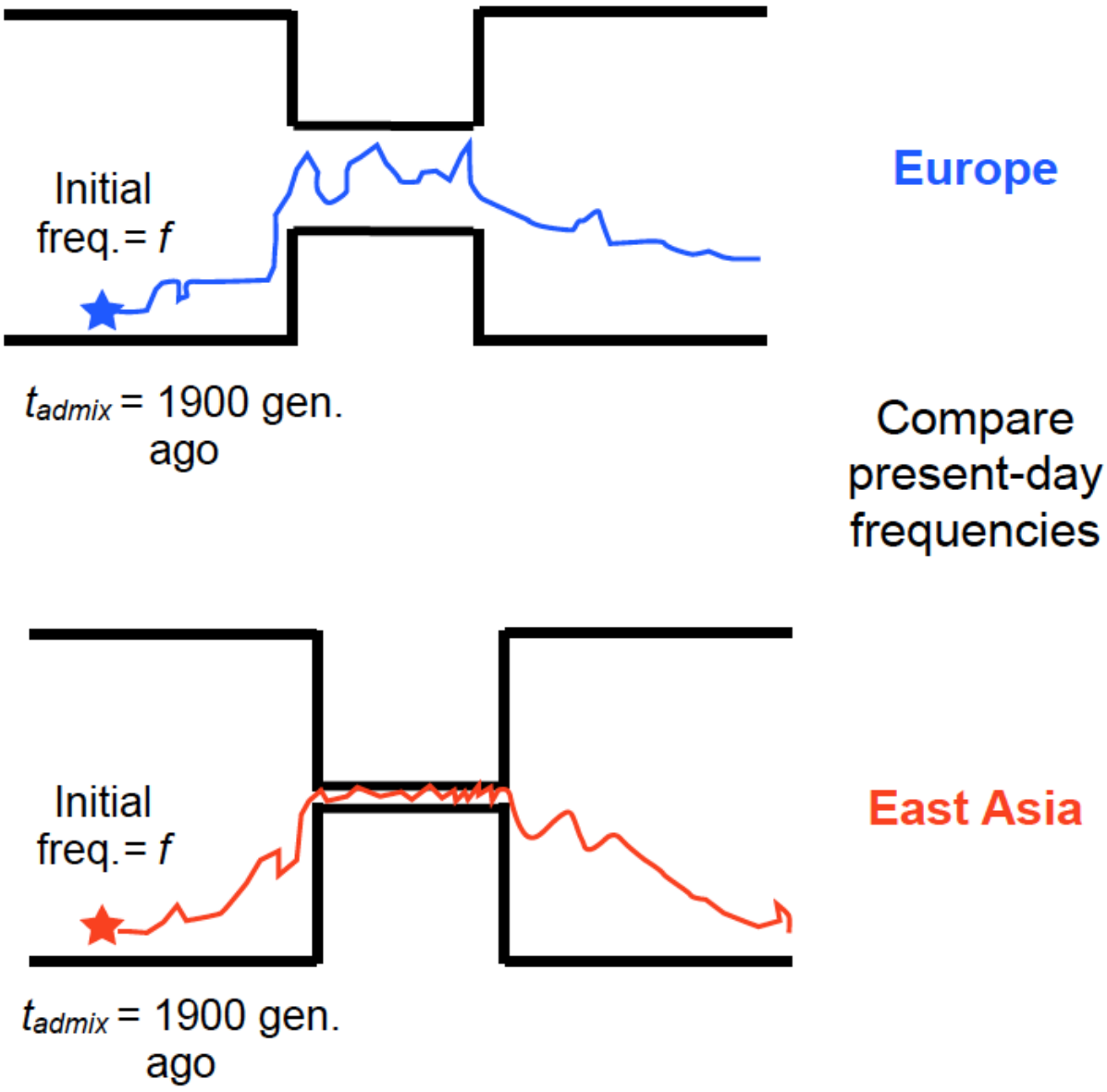

Figure S1: Simulation scenario. At time t_{admix}=1900 generations ago, a Neanderthal allele starts at frequency *f* and changes frequency each generation via selection and drift. Note the difference in bottleneck severity between the European and East Asian populations. See Tables S1 and S2 for a description of the parameters used.

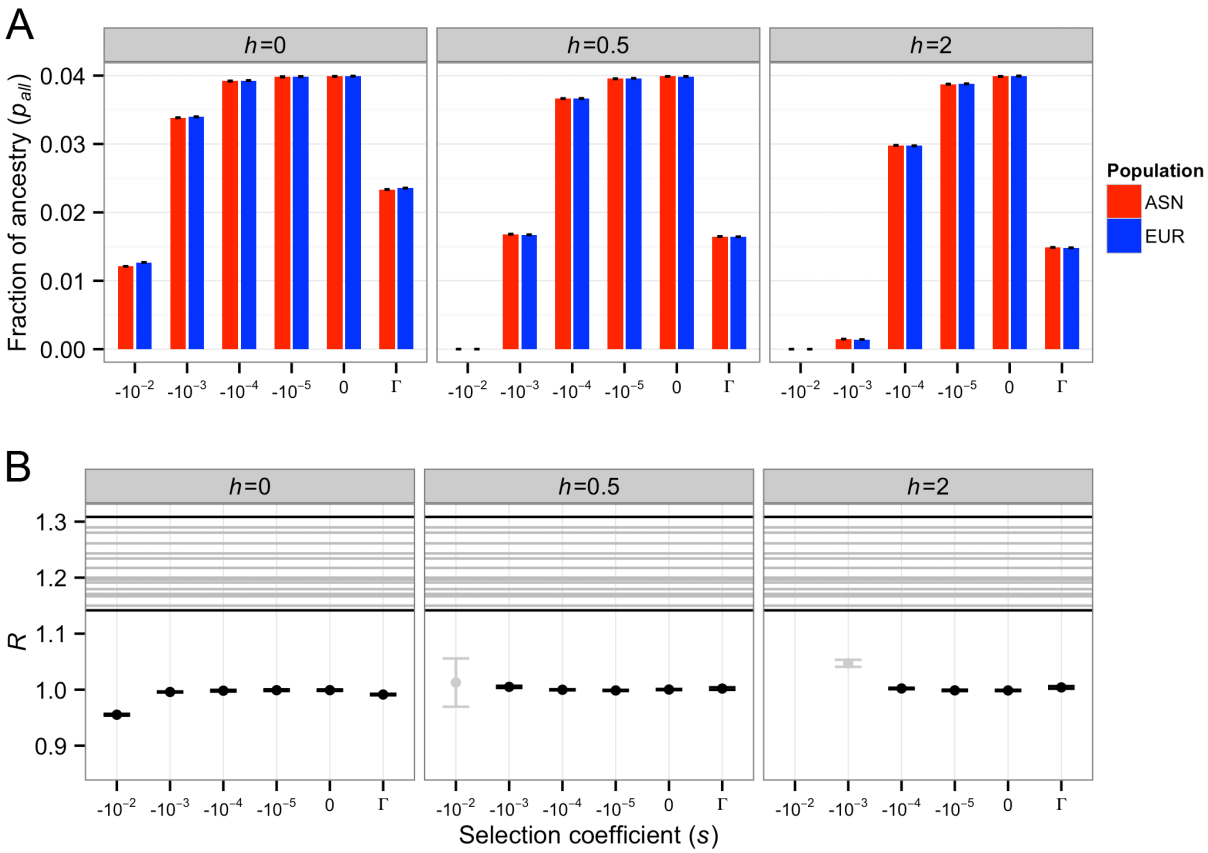

Figure S2: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model when f=4%. Each column depicts results for a different dominance coefficient (h). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (R). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

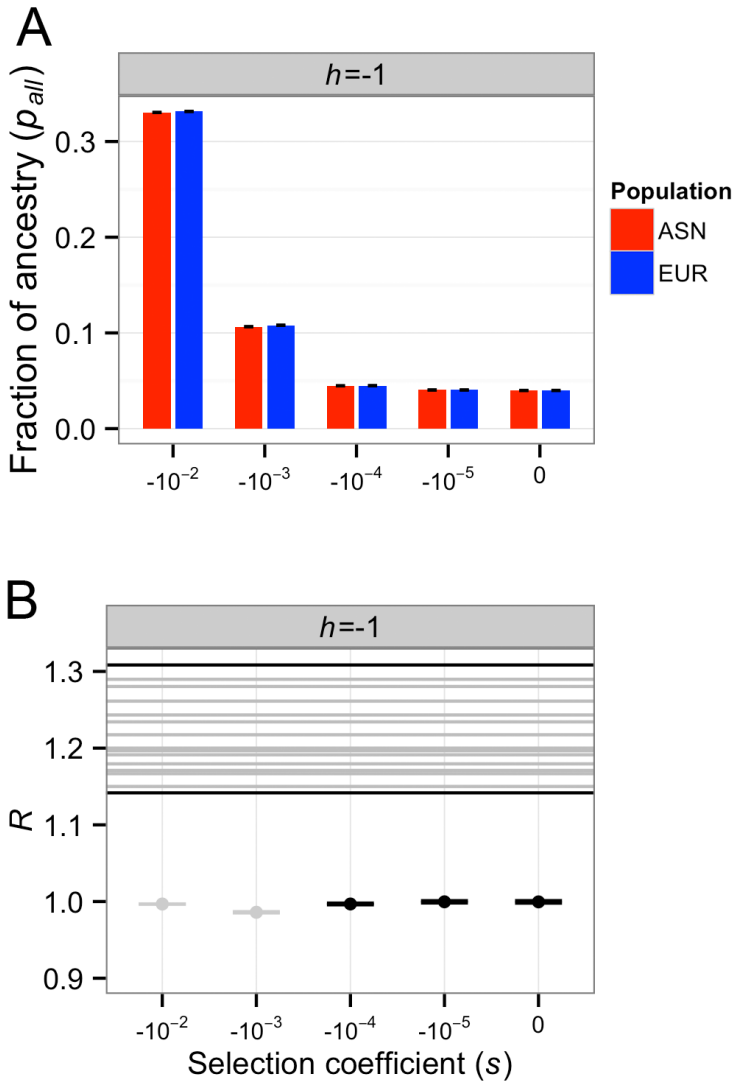

Figure S3: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model when f=4% with overdominance. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (R). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

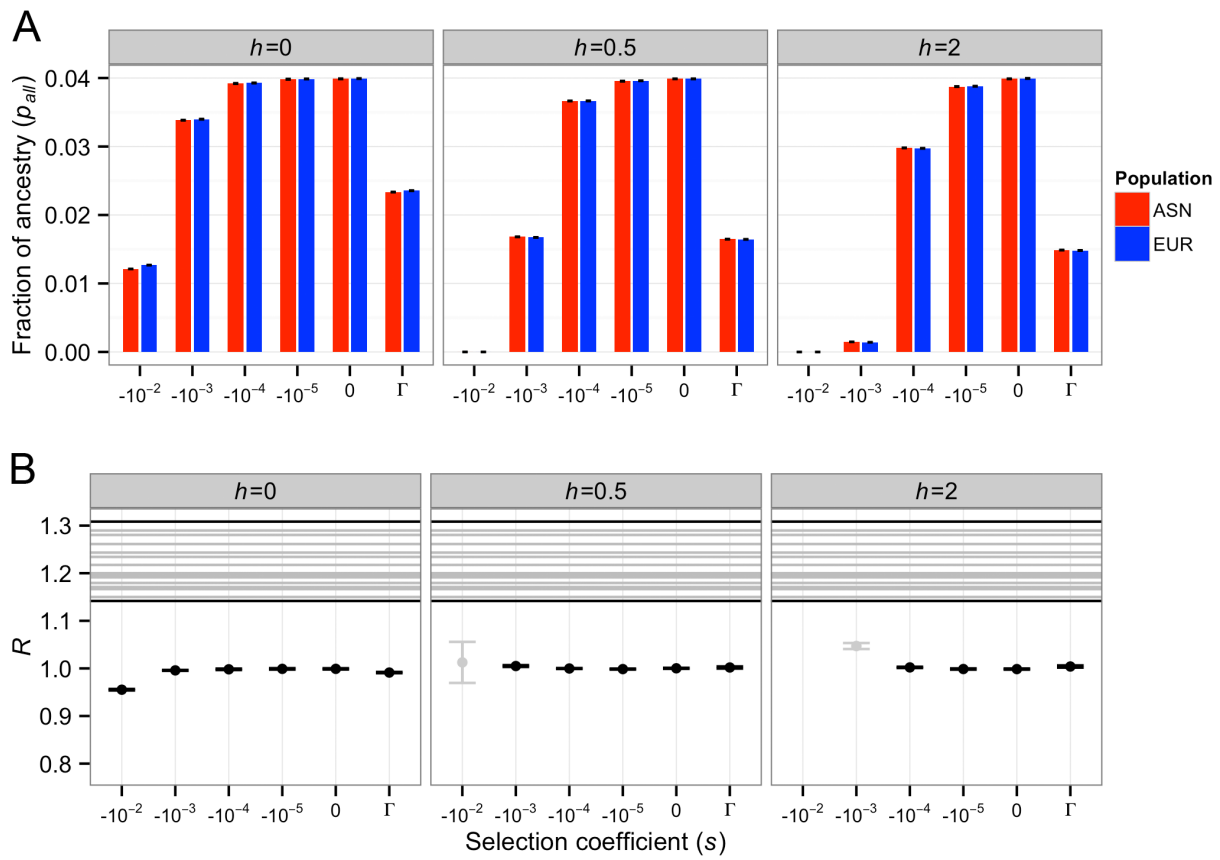

Figure S4: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model with a shorter bottleneck (t_{Blen}=50 generations*)*. The overall severity of the bottleneck (*F*) was equal to that estimated in Keinan et al. Here *f*=4%. Each column depicts results for a different dominance coefficient (*h*). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (*R*). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

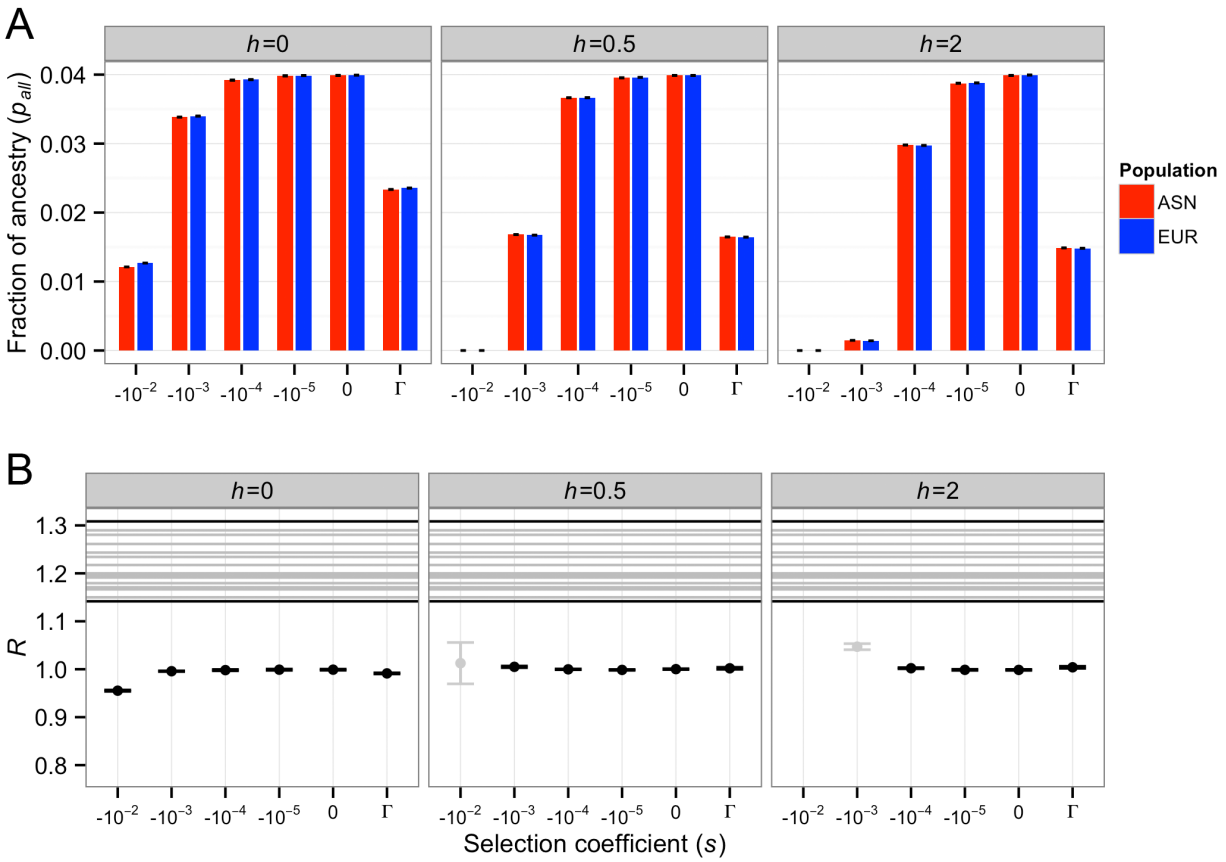

Figure S5: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model with a longer bottleneck (t_{Blen}=200 generations*)*. The overall severity of the bottleneck (*F*) was equal to that estimated in Keinan et al. Here *f*=4%. Each column depicts results for a different dominance coefficient (*h*). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (*R*). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

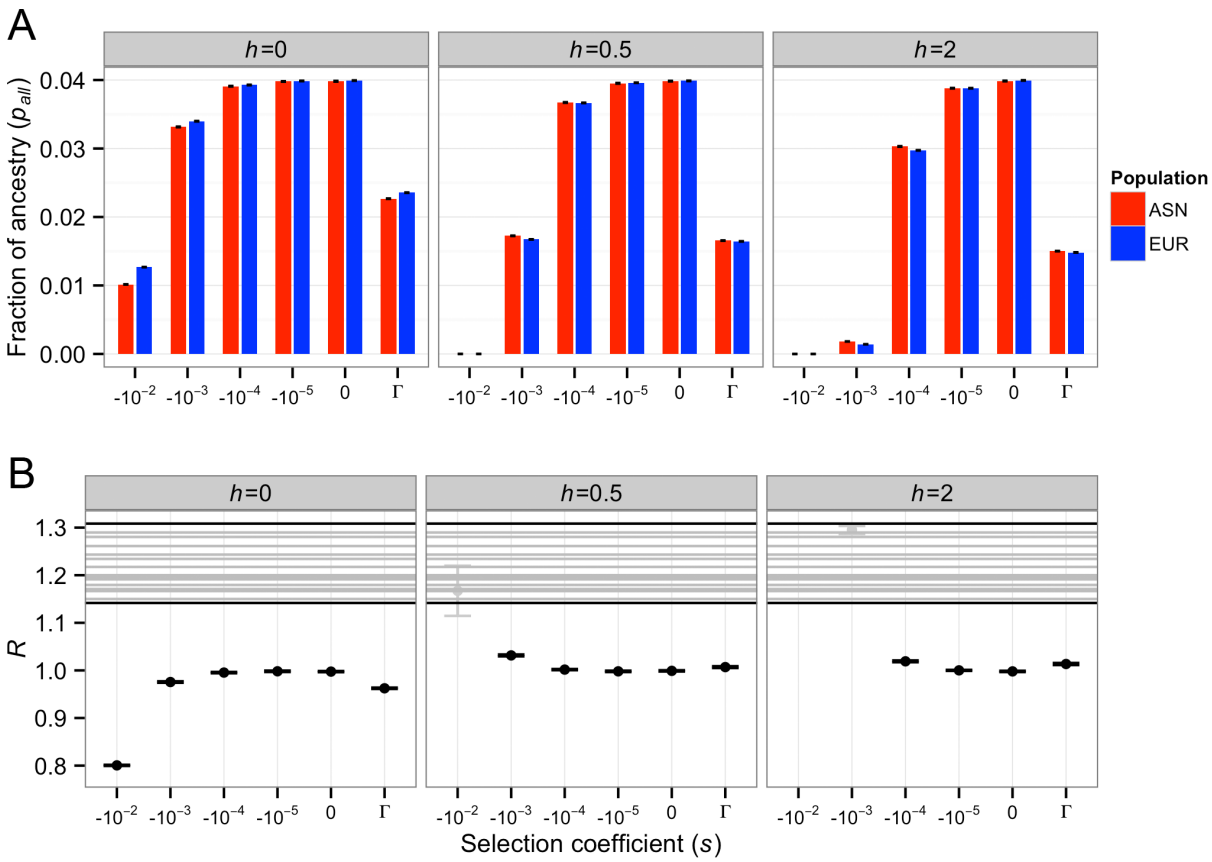

Figure S6: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model where the bottleneck in ASN was 2-times more severe than that estimated by Keinan et al. The severity of the EUR bottleneck was as estimated by Keinan et al. Here f=4%. Each column depicts results for a different dominance coefficient (h). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (R). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

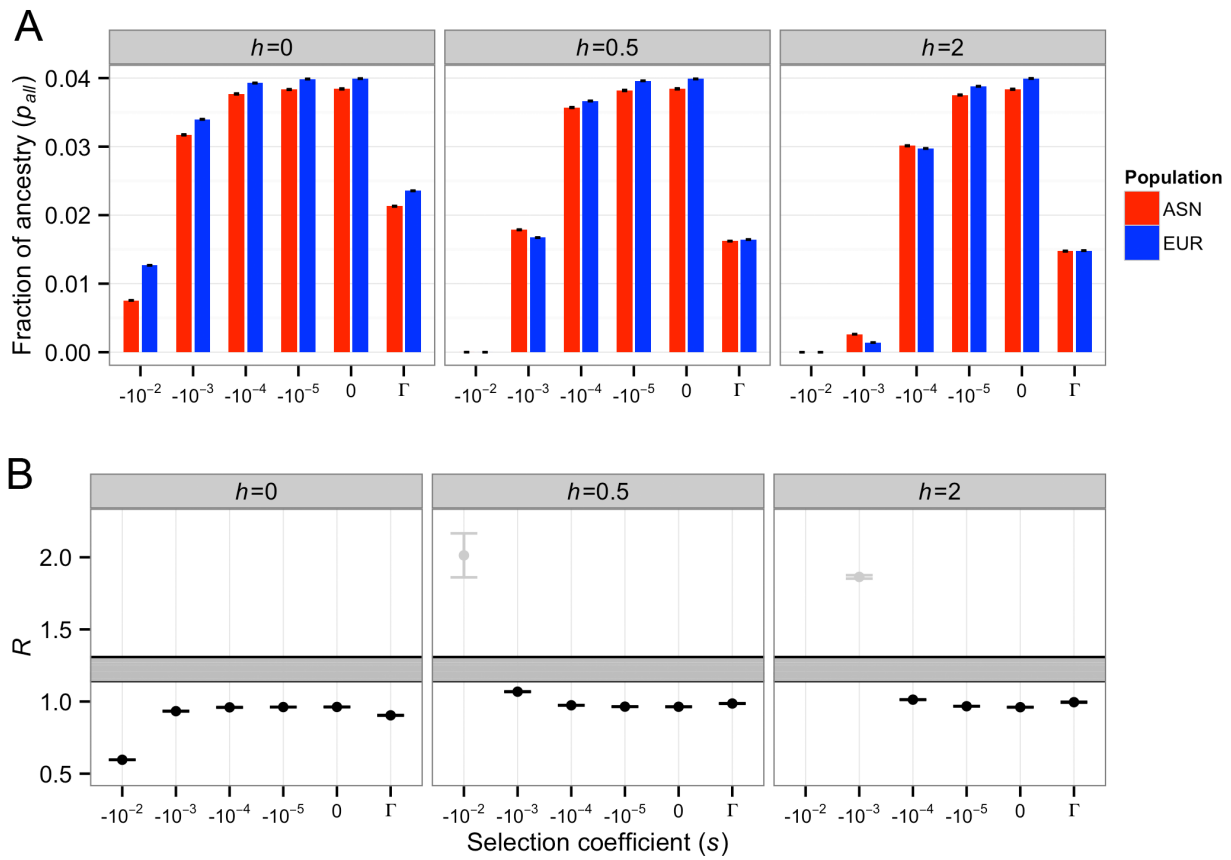

Figure S7: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model where the bottleneck in ASN was 5-times more severe than that estimated by Keinan et al. The severity of the EUR bottleneck was as estimated by Keinan et al. Here f=4%. Each column depicts results for a different dominance coefficient (h). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (R). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

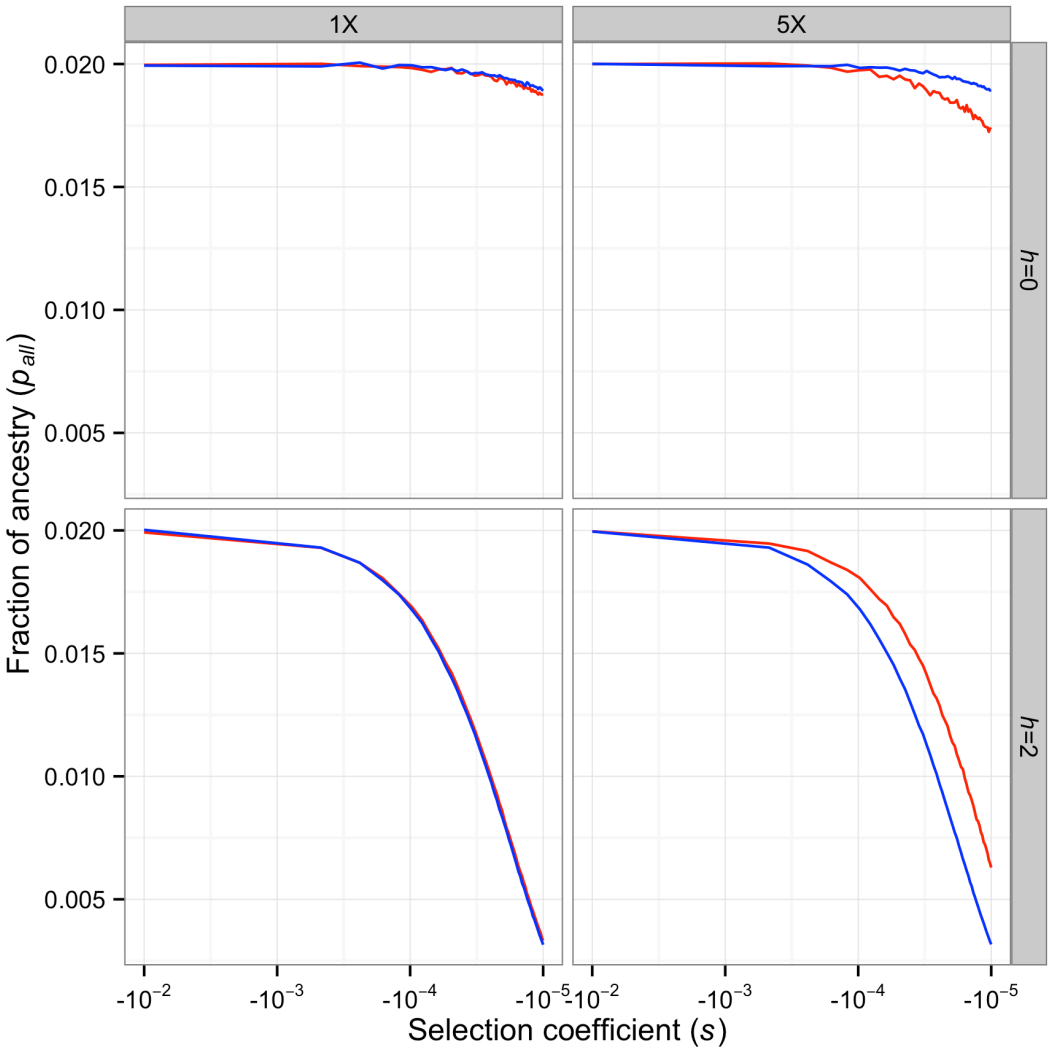

Figure S8: Predicted mean Neanderthal allele frequency at the end of the population bottlenecks in East Asia (ASN) and Europe (EUR) for the recessive and underdominant cases (h=0 and 2, respectively). (Left) Population sizes were set to those inferred in Keinan et al.[11] (Right) Population size in ASN was assumed to be 5-fold smaller than that estimated in Keinan et al.[11] In all cases, constant sized populations were simulated for 100 generations. Here f=2%.

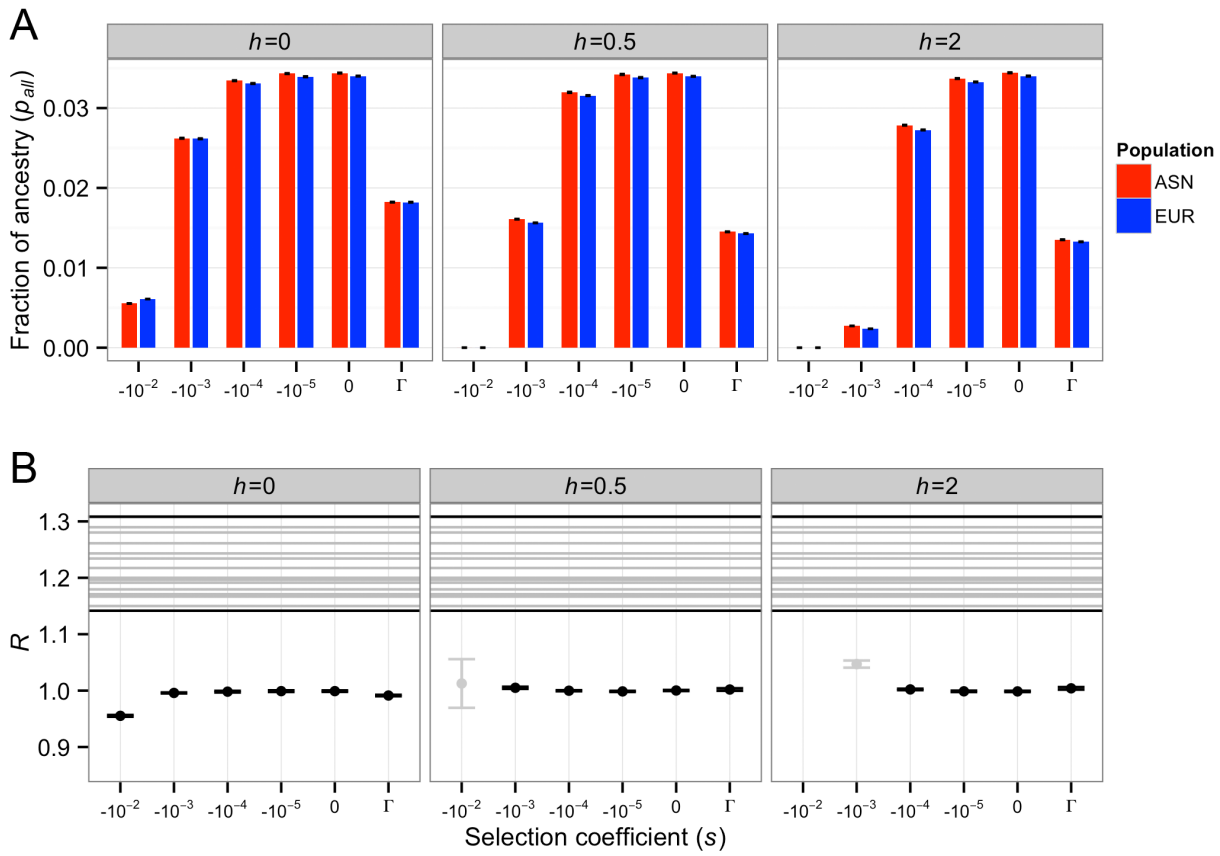

Figure S9: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Gravel et al.[23] complex demographic model when *f*=4%. Each column depicts results for a different dominance coefficient (*h*). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (*R*). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

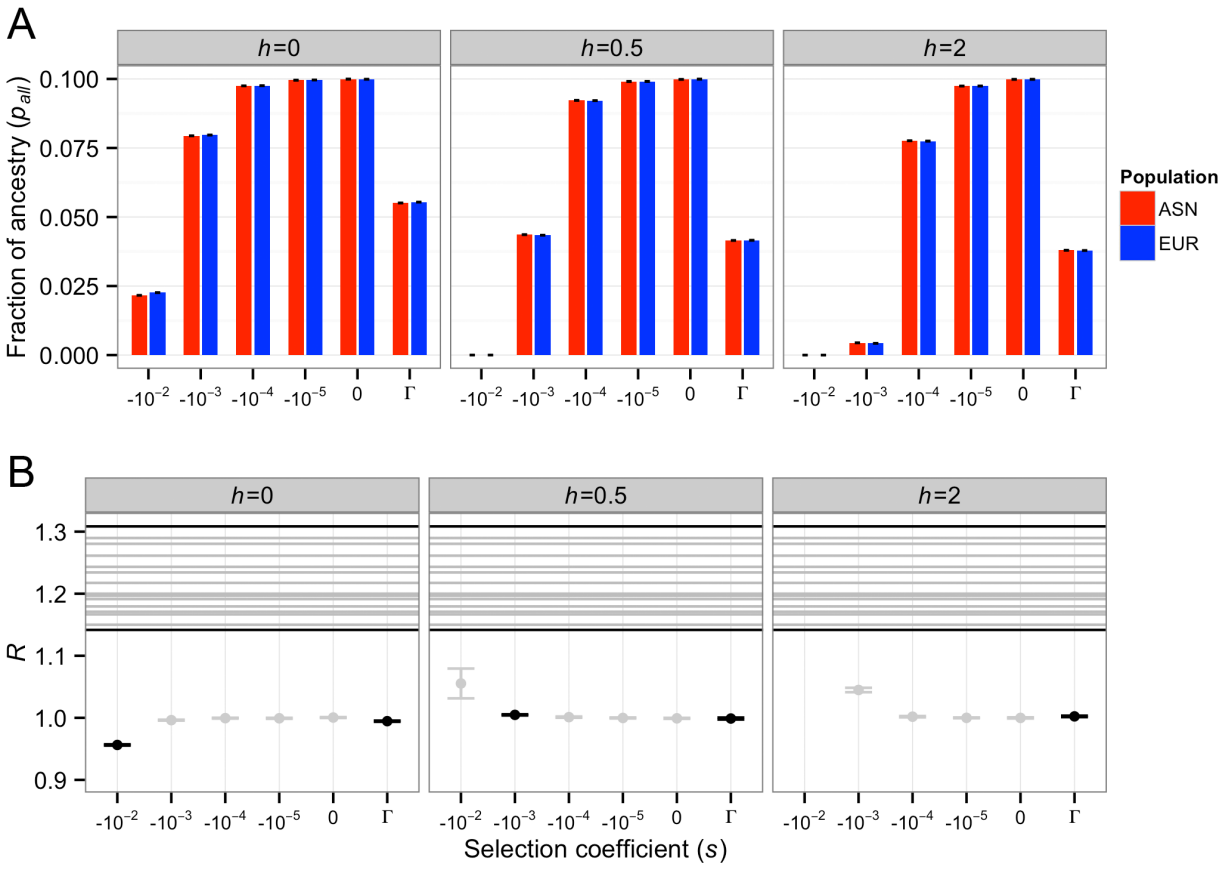

Figure S10: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model when f=10%. Each column depicts results for a different dominance coefficient (h). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (R). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

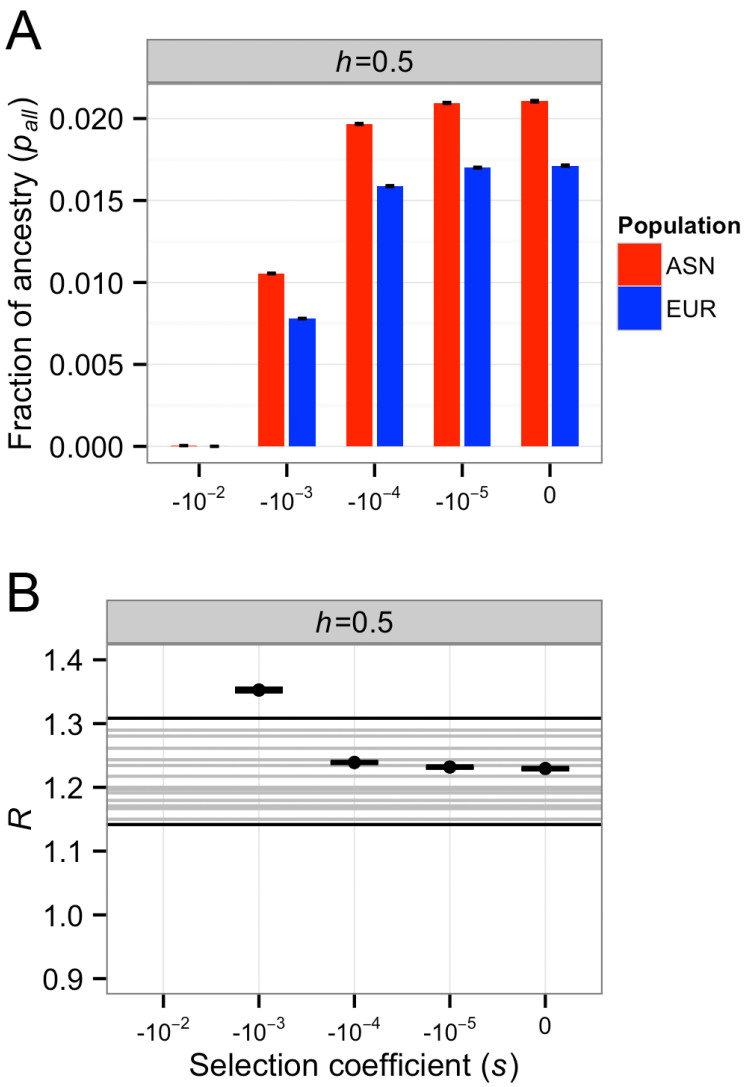

Figure S11: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Gravel et al.[23] complex demographic model when f=2% with a second pulse of Neanderthal admixture into East Asia. Specifically, 920 generations ago, the amount of Neanderthal ancestry at each site in East Asia was increased by 15% of the initial value of f (i.e. here 0.003 was added to the frequency of Neanderthal alleles in the East Asian population). Vernot and Akey[6,25] have estimated that that the second pulse of Neanderthal admixture into East Asia was about 15% of the initial admixture proportion, concordant with our present simulation. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (R). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Note that a broad range of selection coefficients provide values of R compatible with the observed ratio. The model where s=-0.01 predicts R=16. This point was omitted for plotting purposes.

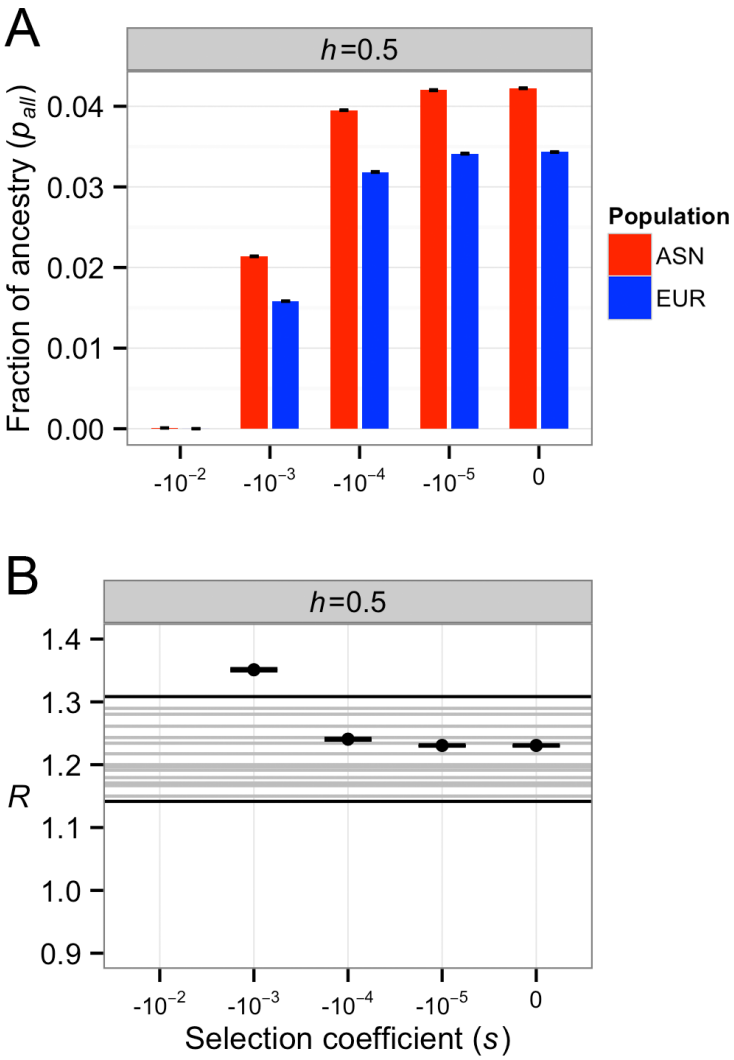

Figure S12: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Gravel et al.[23] complex demographic model when f=4% with a second pulse of Neanderthal admixture into East Asia. Specifically, 920 generations ago, the amount of Neanderthal ancestry at each site in East Asia was increased by 15% of the initial value of f (i.e. here 0.006 was added to the frequency of Neanderthal alleles in the East Asian population). Vernot and Akey[6,25] have estimated that that the second pulse of Neanderthal admixture into East Asia was about 15% of the initial admixture proportion, concordant with our present simulation. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (R). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Note that a broad range of selection coefficients provide values of R compatible with the observed ratio. The model where s=-0.01 predicts R=16. This point was omitted for plotting purposes.

Table S1. Parameters used for the Keinan et al. bottleneck model

Population	N	F	t_B	t_{Blen}	N_B
Parameters inferred in Keinan et al.					
ASN	10063	0.123	720	100	407
EUR	10085	0.091	640	100	549
Shorter bottleneck					
ASN	10063	0.123	720	50	204
EUR	10085	0.091	640	50	275
Longer bottleneck					
ASN	10063	0.123	720	200	814
EUR	10085	0.091	640	200	1098
2-fold more severe bottleneck					
ASN	10063	0.246	720	100	275
EUR	10085	0.091	640	100	549
5-fold more severe bottleneck					
ASN	10063	0.615	720	100	110
EUR	10085	0.091	640	100	549

Table S2. Parameters used for the Gravel et al. model

Parameter	Value
t_1	980
t_2	920
N_{AFR}	14474
N_b	1861
N_{ASN0}	550
N_{EUR0}	1032
r_{ASN}	0.0048
r_{EUR}	0.0038
$m_{ASN\ AFR}$	0.78e-5
$m_{EUR\ AFR}$	2.5e-5
$m_{EUR\ ASN}$	3.11e-5

Table S3: Summary of simulation results [This is in the Excel spreadsheet]

Table S4: Expected *D* statistics under realistic models of human history assuming 0,1, or 2 pulses of Neanderthal admixture

Model	P1	P2	Num P1	Num P2	D	SE
No admixture	ASN	EUR	84987	85221	-0.0014	0.0025
	ASN	AFR	90844	90332	-0.0028	0.0023
	EUR	AFR	90063	89785	-0.0015	0.0024
One pulse	ASN	EUR	102233	102151	0.0004	0.0022
	ASN	AFR	91001	109182	0.0908	0.0023
	EUR	AFR	90798	108897	0.0906	0.0022
Two pulse	ASN	EUR	106000	101487	0.0218	0.0021
	ASN	AFR	90837	113021	0.1088	0.0022
	EUR	AFR	90881	108552	0.0886	0.0022

D statistics were computed from data simulated using ms[26] under the demographic model estimated for human populations in Gravel et al.[23] with our own modifications and those suggested by Vernot and Akey.[6,25] Specifically, recent population growth, as used in Vernot and Akey was included in the model. We simulated the three human populations and a Neanderthal population that split from the human population 400,000 years ago. The Neanderthal population had a constant size of 1500 individuals. The one pulse model includes a 500-year period of migration between the ancestral non-African population and Neanderthals. The two-pulse model includes the same migration as in the one pulse model, except it includes an additional 500 years of migration between the Neanderthal and East Asian populations. Note, we decreased the human-Neanderthal migration rates by 2 relative to the values given in Vernot and Akey[6] to give *D* statistics more comparable to those observed in actual data. The precise ms commands for these models are given in Table S5.

The *D* test was computed as: *D*=(Num_P1-Num_P2)/(Num_P1+Num_P2). Our simulations assume that the derived allele can be accurately inferred. As such, we simulated the three human populations (EUR, AFR, ASN) and a Neanderthal population.

Standard errors were computed using a nonparametric bootstrap of the values shown in the table. This is appropriate as each site was simulated independently of the others.

The *D* statistics for all the simulations without any Neanderthal admixture are within 2 standard errors of 0. Further, the *D* statistic computed using ASN and EUR under the one pulse model also is within 2 standard errors of 0. This suggests that a model with one pulse of Neanderthal admixture cannot explain the higher Neanderthal ancestry in East Asia, even with a higher migration rate between African and Europe than between Africa and East Asia. The two-pulse model, however, predicts *D* statistics significantly >0.

Table S5: ms commands for neutral coalescent simulations in Table S4

Model	Command
No admixture	ms 4 1 -s 1 -I 4 1 1 1 1 0 -n 4 2.051984e-01 -n 1 58.002735978 -n 2 70.041039672 -n 3 187.55 -eg 0 1 482.46 -eg 0 2 570.18 -eg 0 3 720.23 -em 0 2 4 0 -em 0 2 4 0 -em 0 3 4 0 -em 0 3 4 0 -em 0 1 2 0.7310 -em 0 2 1 0.7310 -em 0 1 3 0.228072 -em 0 3 1 0.228072 -em 0 2 3 0.909364 -em 0 3 2 0.909364 -eg 0.006997264 1 0 -eg 0.006997264 2 2.089166e+01 -eg 0.006997264 3 3.006376e+01 -en 0.006997264 1 1.98002736 -en 0.031463748 2 7.774282e-01 -en 0.031463748 3 5.820793e-01 -ej 5.453352e-02 3 2 -en 5.453352e-02 2 7.774282e-01 -em 5.453352e-02 1 2 4.386 -em 5.453352e-02 2 1 4.386 -ej 8.207934e-02 2 1 -en 8.207934e-02 1 1.98002736 -en 0.20246238 1 1 -ej 9.575923e-01 4 1
One pulse	ms 4 1 -s 1 -I 4 1 1 1 1 0 -n 4 2.051984e-01 -n 1 58.002735978 -n 2 70.041039672 -n 3 187.55 -eg 0 1 482.46 -eg 0 2 570.18 -eg 0 3 720.23 -em 6.635294e-02 2 4 0 -em 0 3 4 0 -em 0 3 4 0 -em 0 1 2 0.7310 -em 0 2 1 0.7310 -em 0 1 3 0.228072 -em 0 3 1 0.228072 -em 0 2 3 0.909364 -em 0 3 2 0.909364 -eg 0.006997264 1 0 -eg 0.006997264 2 2.089166e+01 -eg 0.006997264 3 3.006376e+01 -en 0.006997264 1 1.98002736 -en 0.031463748 2 7.774282e-01 -en 0.031463748 3 5.820793e-01 -ej 5.453352e-02 3 2 -en 5.453352e-02 2 7.774282e-01 -em 5.453352e-02 1 2 4.386 -em 5.453352e-02 2 1 4.386 -ej 8.207934e-02 2 1 -en 8.207934e-02 1 1.98002736 -en 0.20246238 1 1 -em 6.566895e-02 2 4 4.386000e+01 -ej 9.575923e-01 4 1
Two pulse	ms 4 1 -s 1 -I 4 1 1 1 1 0 -n 4 2.051984e-01 -n 1 58.002735978 -n 2 70.041039672 -n 3 187.55 -eg 0 1 482.46 -eg 0 2 570.18 -eg 0 3 720.23 -em 0 1 2 0.7310 -em 0 2 1 0.7310 -em 0 1 3 0.228072 -em 0 3 1 0.228072 -em 0 2 3 0.909364 -em 0 3 2 0.909364 -eg 0.006997264 1 0 -eg 0.006997264 2 2.089166e+01 -eg 0.006997264 3 3.006376e+01 -en 0.006997264 1 1.98002736 -en 0.031463748 2 7.774282e-01 -en 0.031463748 3 5.820793e-01 -ej 5.453352e-02 3 2 -en 5.453352e-02 2 7.774282e-01 -em 5.453352e-02 1 2 4.386 -em 5.453352e-02 2 1 4.386 -ej 8.207934e-02 2 1 -en 8.207934e-02 1 1.98002736 -en 0.20246238 1 1 -em 6.566895e-02 2 4 4.386000e+01 -em 6.635294e-02 2 4 0 -em 5.316553e-02 3 4 8.832178e+00 -em 5.384952e-02 3 4 0 -ej 9.575923e-01 4 1

See Table S4 for a description of the demographic model.