The complete genome sequence of a Neanderthal from the Altai Mountains

Kay Prüfer[1], Fernando Racimo[2], Nick Patterson[3], Flora Jay[2], Sriram Sankararaman[3,4], Susanna Sawyer[1], Anja Heinze[1], Gabriel Renaud[1], Peter H. Sudmant[5], Cesare de Filippo[1], Heng Li[3], Swapan Mallick[3,4], Michael Dannemann[1], Qiaomei Fu[1,6], Martin Kircher[1,5], Martin Kuhlwilm[1], Michael Lachmann[1], Matthias Meyer[1], Matthias Ongyerth[1], Michael Siebauer[1], Christoph Theunert[1], Arti Tandon[3,4], Priya Moorjani[4], Joseph Pickrell[4], James C. Mullikin[7], Samuel H. Vohr[8], Richard E. Green[8], Ines Hellmann[9]†, Philip L. F. Johnson[10], Hélène Blanche[11], Howard Cann[11], Jacob O. Kitzman[5], Jay Shendure[5], Evan E. Eichler[5,12], Ed S. Lein[13], Trygve E. Bakken[13], Liubov V. Golovanova[14], Vladimir B. Doronichev[14], Michael V. Shunkov[15], Anatoli P. Derevianko[15], Bence Viola[16], Montgomery Slatkin[2], David Reich[3,4,17], Janet Kelso[1] & Svante Pääbo[1]

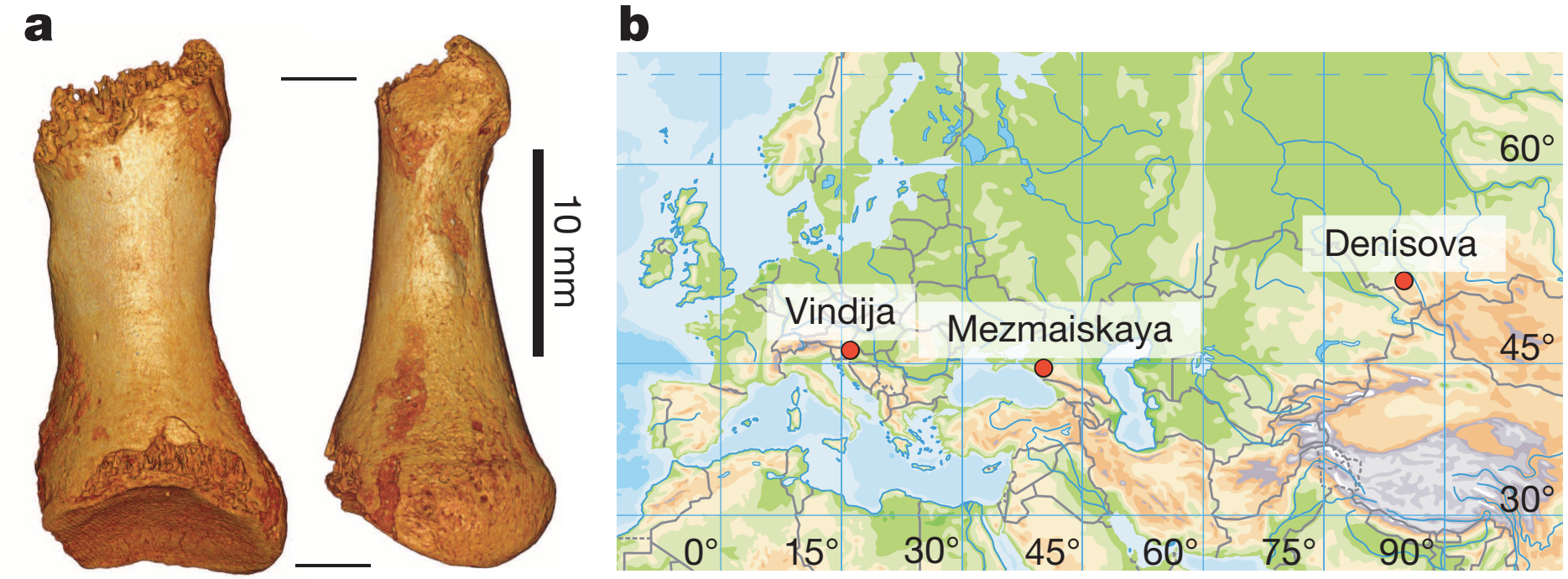
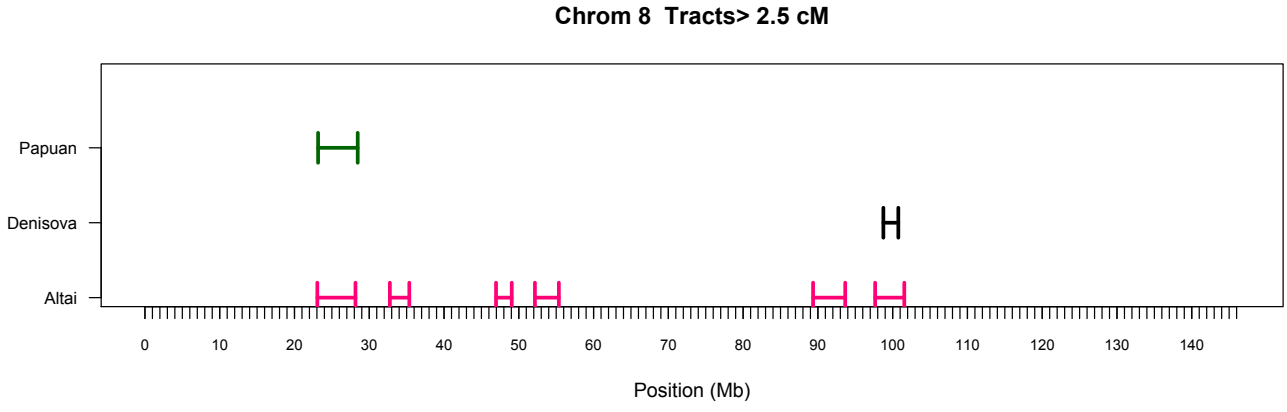


Figure 1 | Toe phalanx and location of Neanderthal samples for which genome-wide data are available. **a**, The toe phalanx found in the east gallery of Denisova Cave in 2010. Dorsal view (left image), left view (right image). Total length of the bone is 26 mm. **b**, Map of Eurasia showing the location of Vindija Cave, Mezmaiskaya Cave and Denisova Cave, where Neanderthal samples used here were found.



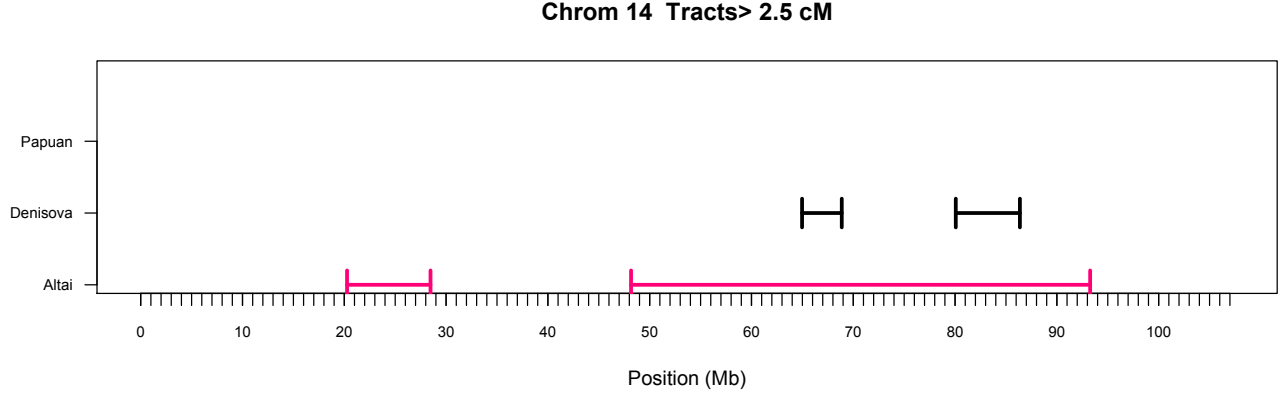


Figure S10.1 HBD tracts identified in chromosomes 8 and 14 for Papuan (top line, green), Denisova (middle line, black), and Altai (bottom line, pink).

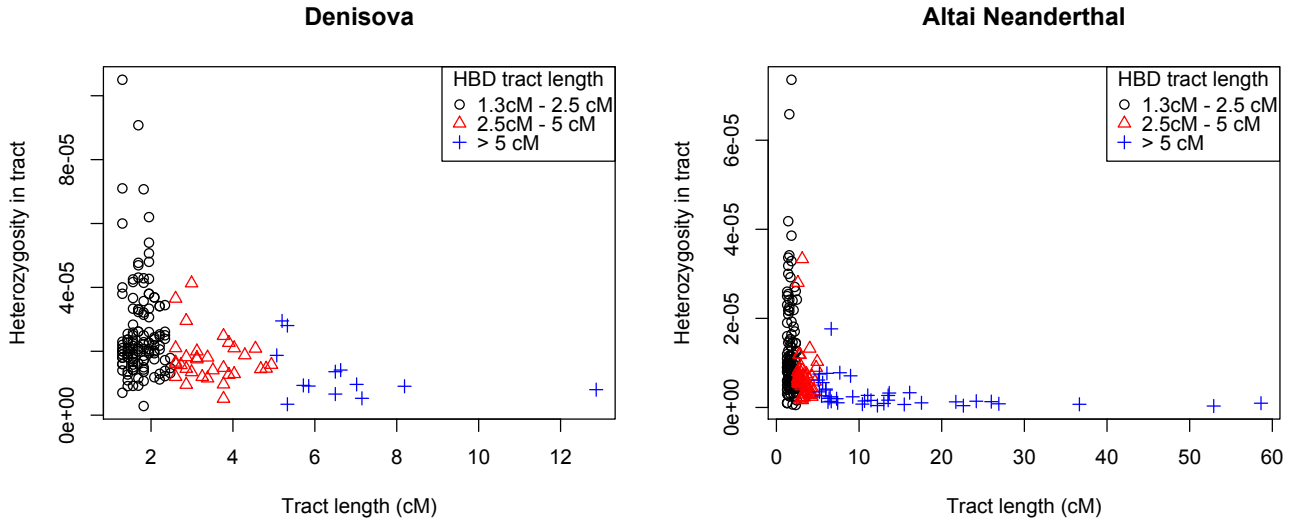
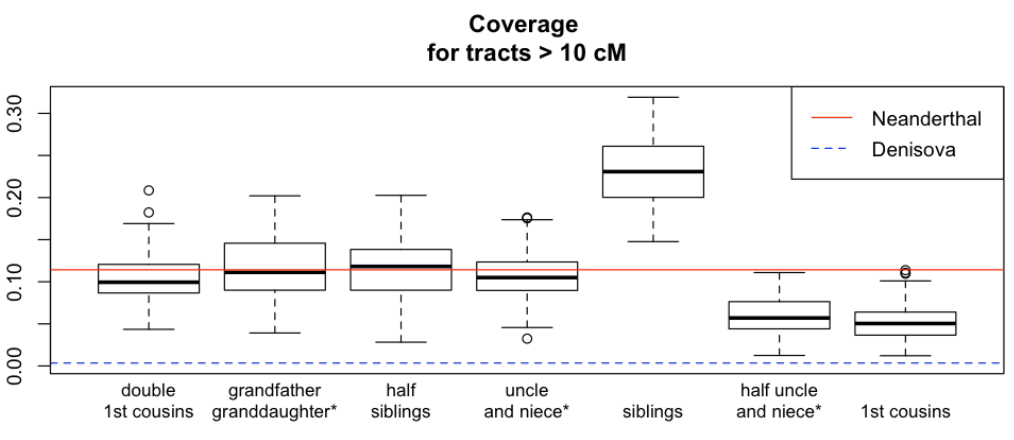


Figure S10.3. Heterozygosity in HBD tracts detected by the scan for Denisova and Altai Neanderthal as a function of the tract length.


Coverage
for tracts > 10 cM
Neanderthal
Denisova
0.00
0.10
0.20
0.30
double
1st cousins
grandfather
granddaughter*
half
siblings
uncle
and niece*
siblings
half uncle
and niece*
1st cousins
Offspring of

Conclusion. The inbreeding coefficient of the Altai individual is likely to be 1/8, which implies that her parents were double first cousins, grandfather and granddaughter, grandmother and grandson, half siblings, uncle and niece, or aunt and nephew, but that we cannot distinguish among these possibilities using runs of homozygosity on the autosomes. This number provides an upper bound for the inbreeding coefficient as a smaller false positive rate or unobserved heterozygous sites (due to missing data) might decrease the total length of homozygous tracts.

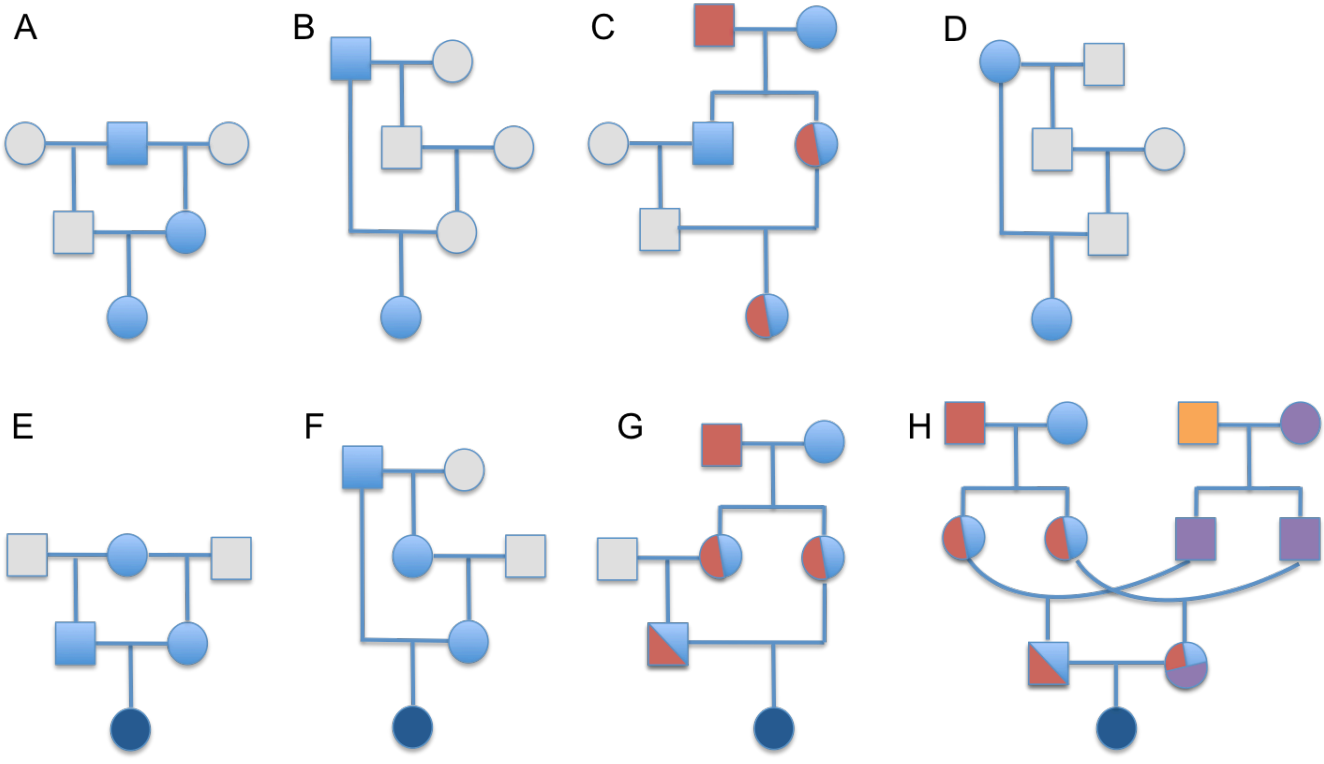


Figure S10.9. Non-exhaustive illustration of pedigrees that can be excluded (top, A-D) or not excluded (bottom, E-H), using X chromosome information. Gray denotes the absence of X sequence coming from the recent common ancestor(s). Other colors denote the potential presence of X sequence coming from the common ancestor(s). Dark blue indicates that both parents might carry X chunks inherited from the same recent common ancestor, thus the individual might be inbred for X. The pedigrees depict cases of the following scenarios: offspring of half-siblings (A,E), grandfather-granddaughter (B, F), aunt-nephew (C,G), grandmother-grandson (D), double-first-cousins (H)

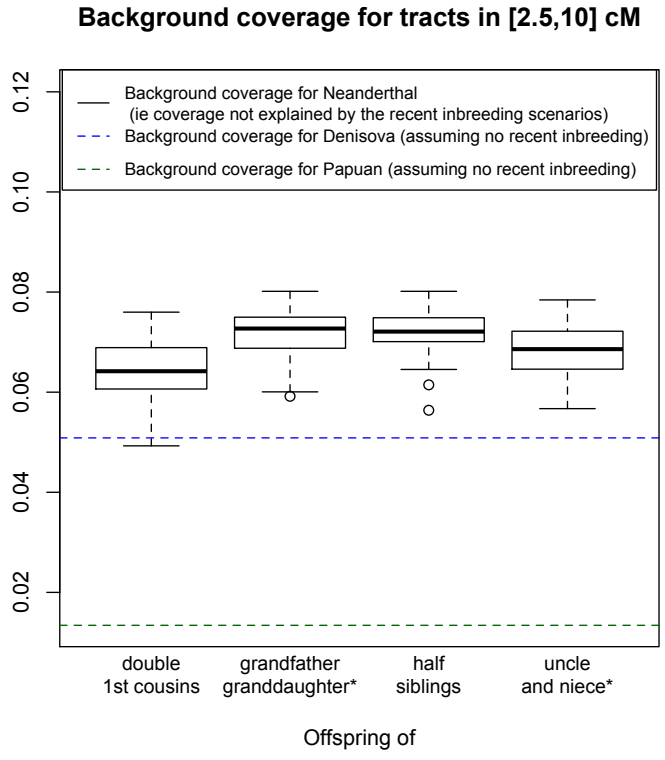


Figure S10.10 Background coverage for tracts between 2.5 and 10 cM.

Conclusion: The observed background coverage of HBD tracts could be explained by the presence of background inbreeding in the population. Alternatively, a demographic scenario of random mating with successive bottlenecks starting after the split from modern humans that induce a very small population size at time of sampling (~600 individuals) also provides a reasonable fit to the data. Note that when a population is very small for a long time the chance of mating between distant cousins is not negligible even in case of random mating.

Patterns of ancient selection in modern humans around candidate sites

Fernando Racimo,*[,1,2] Martin Kuhlwilm,[2] Montgomery Slatkin,[1]

[1]Department of Integrative Biology, University of California, Berkeley, CA, USA

[2]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

***Corresponding author:** E-mail: fernandoracimo@gmail.com.

Associate Editor: X

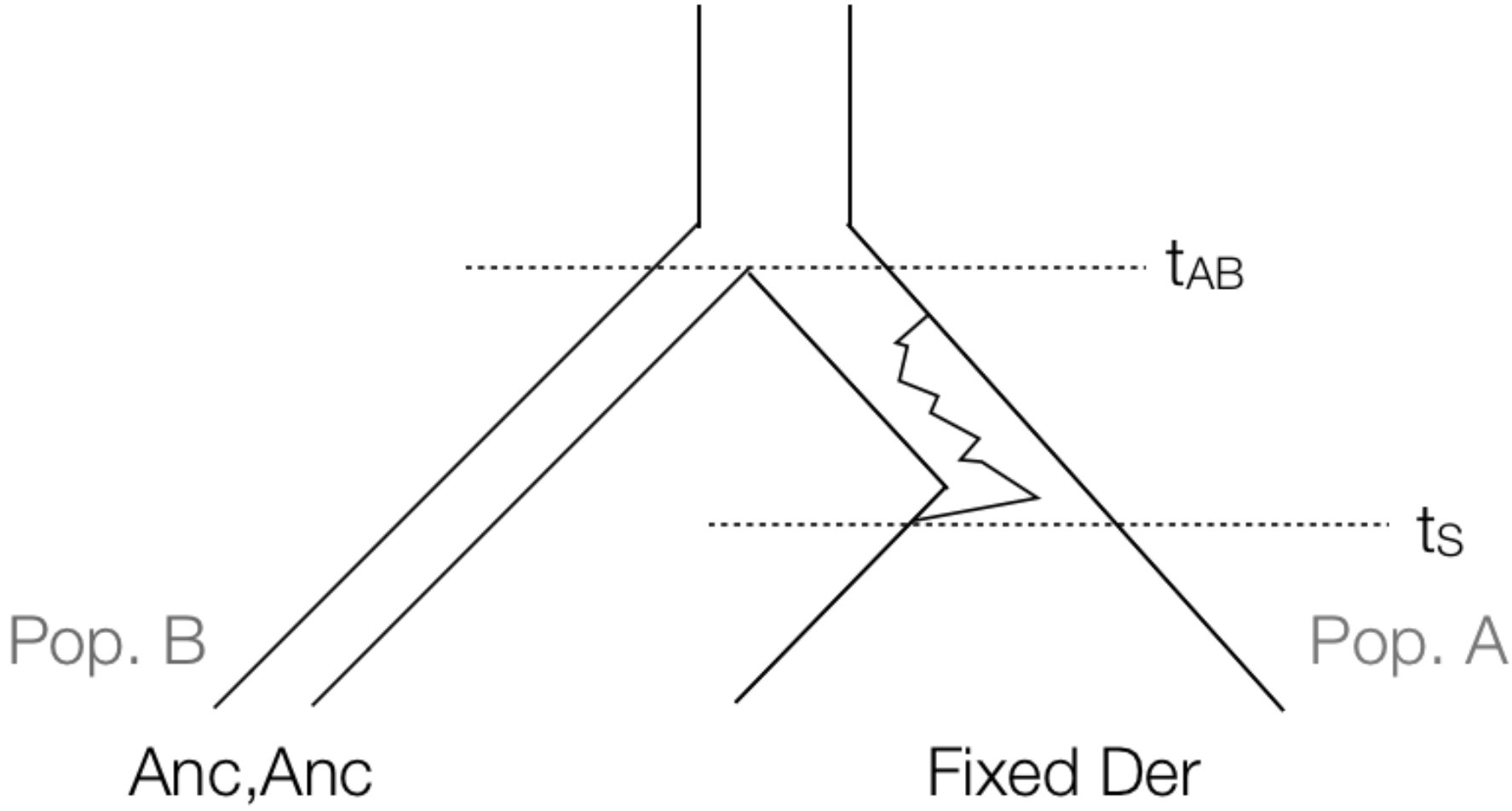


FIG. 5. Tree representing msms runs to simulate a change in a site that is homozygous ancestral in an archaic human (Pop. B) and rises to fixation in modern humans (Pop. A). t_{AB}=modern-archaic split time. t_S=derived allele fixation time.

Test statistics for sets of 4 adjacent segregating sites

H_E, $= \sum_{i=1}^{4} 2p_i(1-p_i)$ (i. e. not π).

H_M, the frequency of the most common haplotype.

H_S, evenness of haplotype frequency distribution

H_I, the inconsistency of the majority haplotype with the outgroup genotype.

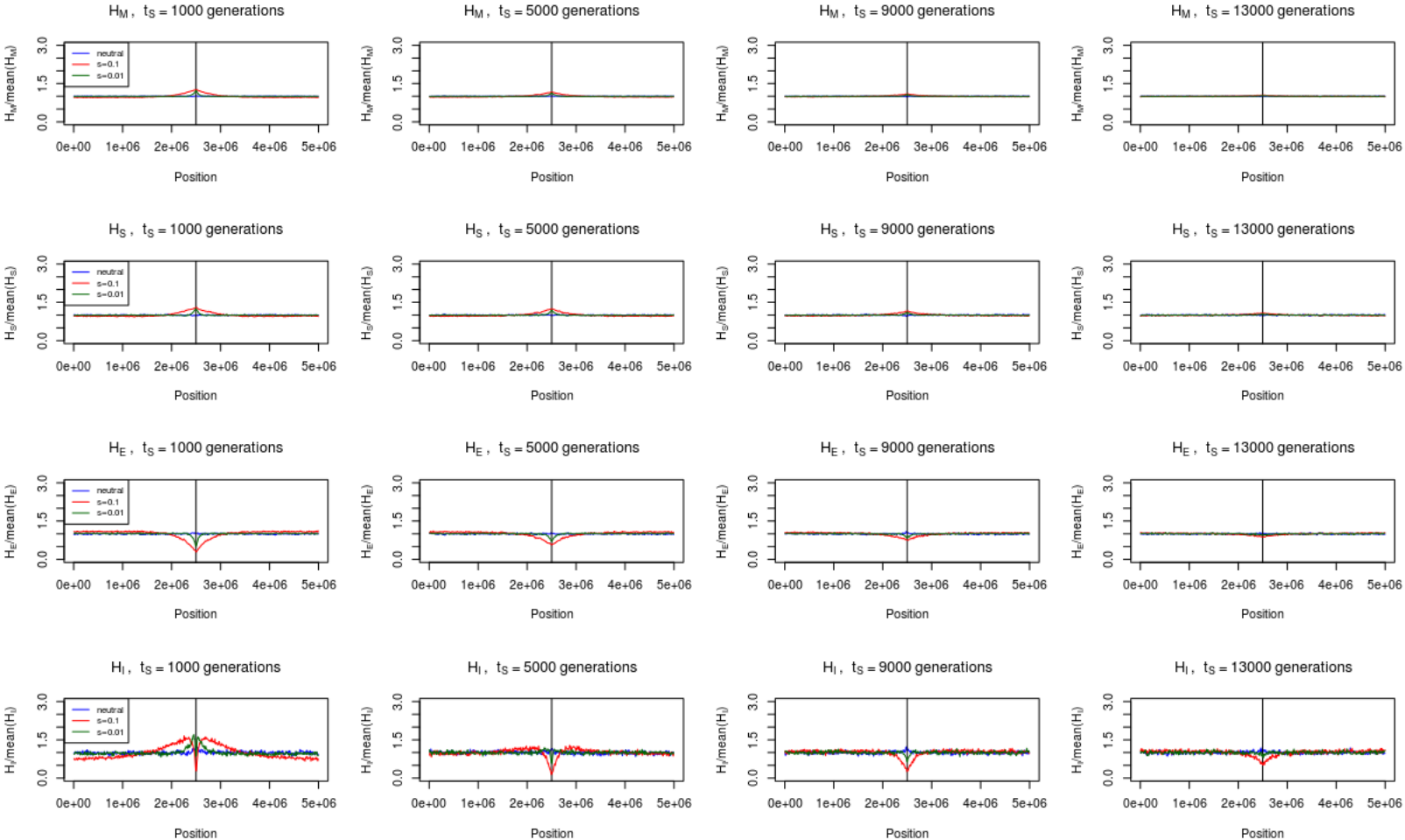


Figure S3. The mean values of the H_E, H_M, H_S and H_I statistics from 200 simulations run under the same parameters were calculated along windows of 100 kb (=0.1 cM) in a 5 Mb region and divided by their mean value along the entire region.

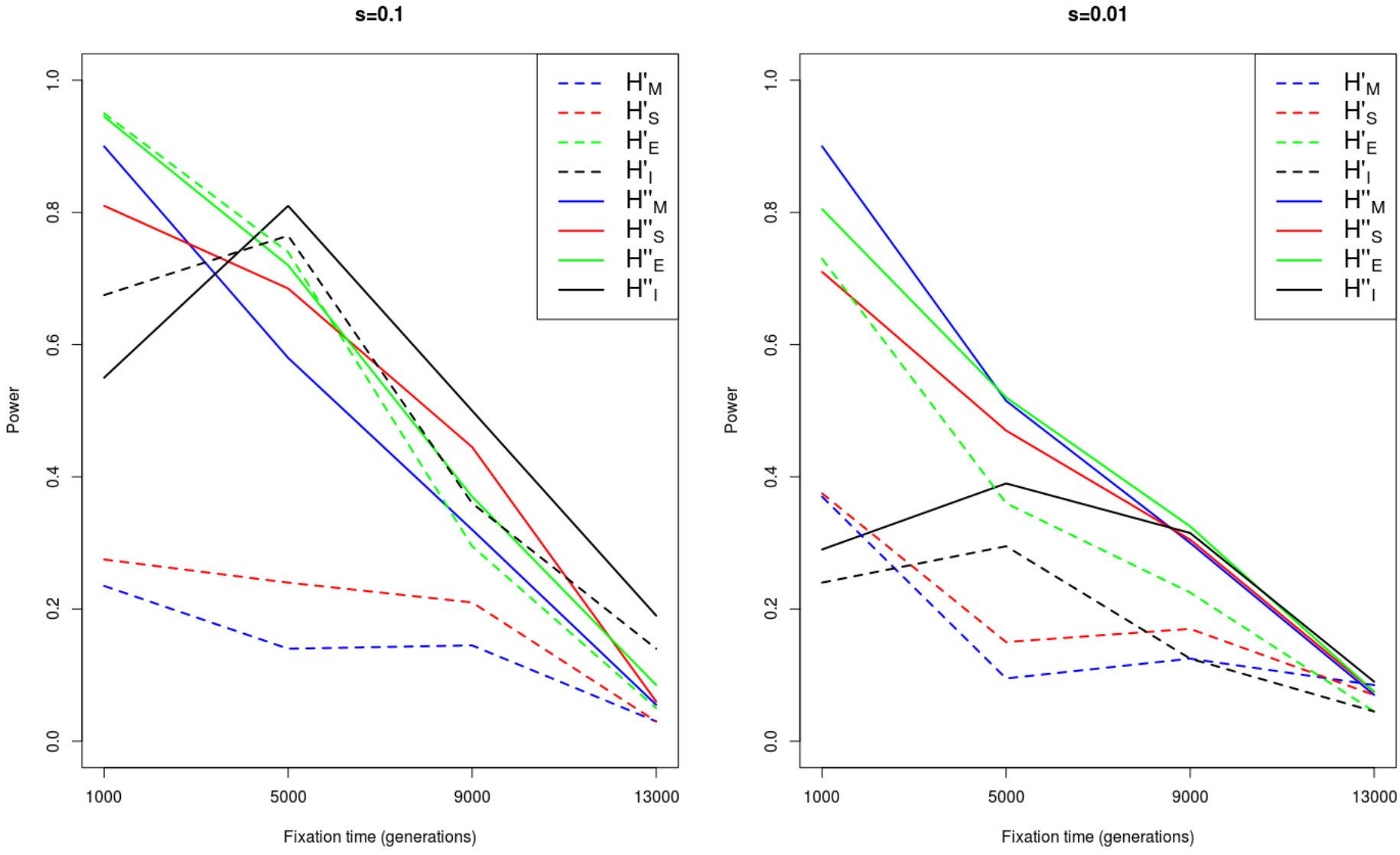


FIG. 6. Power to reject neutrality for different statistics under two different selection coefficients and a range of times since fixation, estimated by calculating the proportion of simulations (out of 200) that have a value more extreme than 95% of 200 neutral simulations.

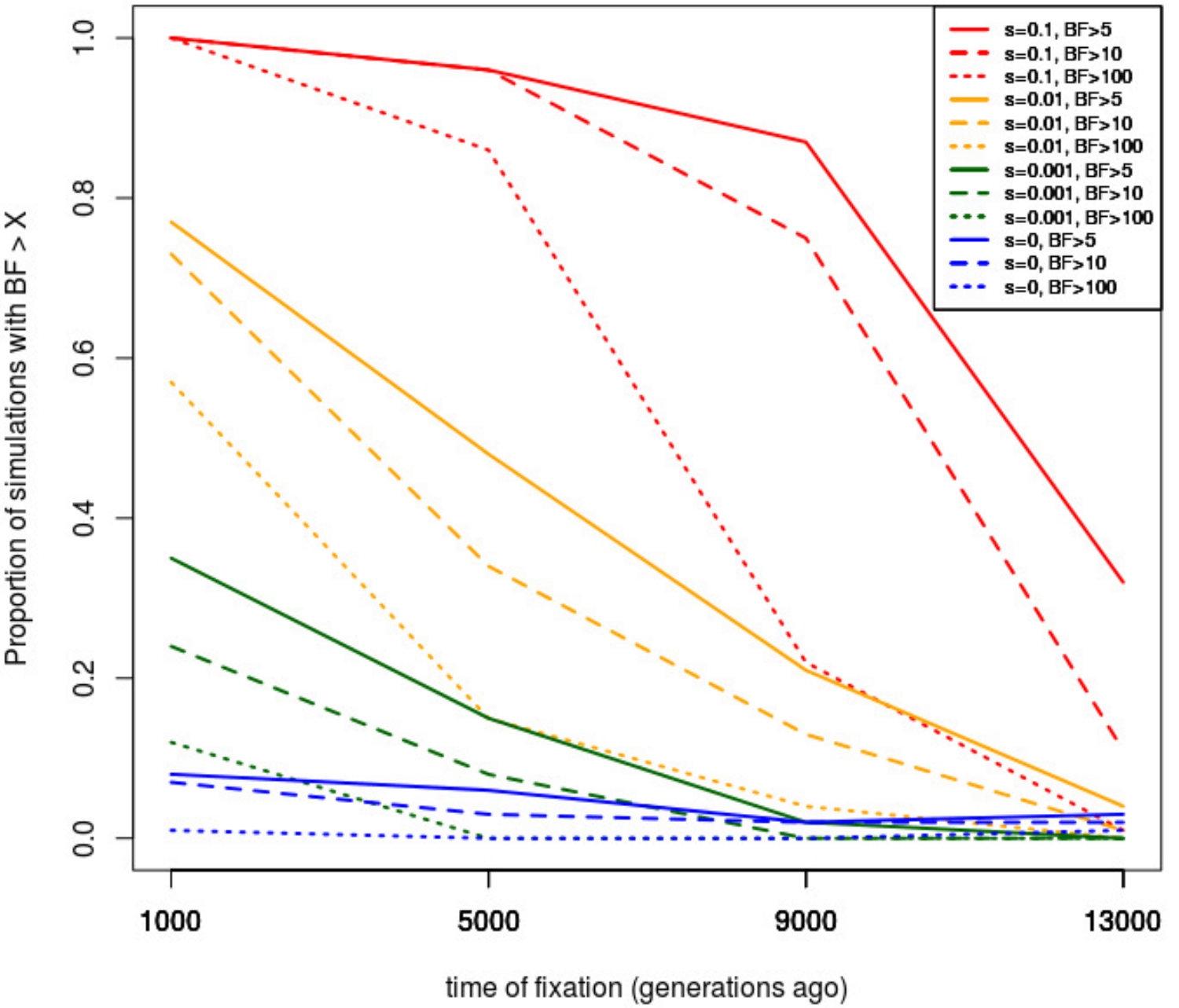


FIG. 7. Sets of 100 simulations were run through the ABC pipeline to obtain Bayes factors in favor of selection (versus neutrality) under different known parameters. The lines show the proportion of the simulations that have a Bayes factor larger than the specified cutoffs. BF = Bayes factor, s=selection coefficient, t=time since derived allele fixation, in generations.

Numbers of sites fixed derived in humans (p>0.99) and fixed ancestral in the high-coverage Neanderthal and Denisovan genomes.

Nonsynonymous: 109
Synonymous: 120
Splice: 45
3' UTR: 364
5' UTR: 93
Regulatory motif: 26

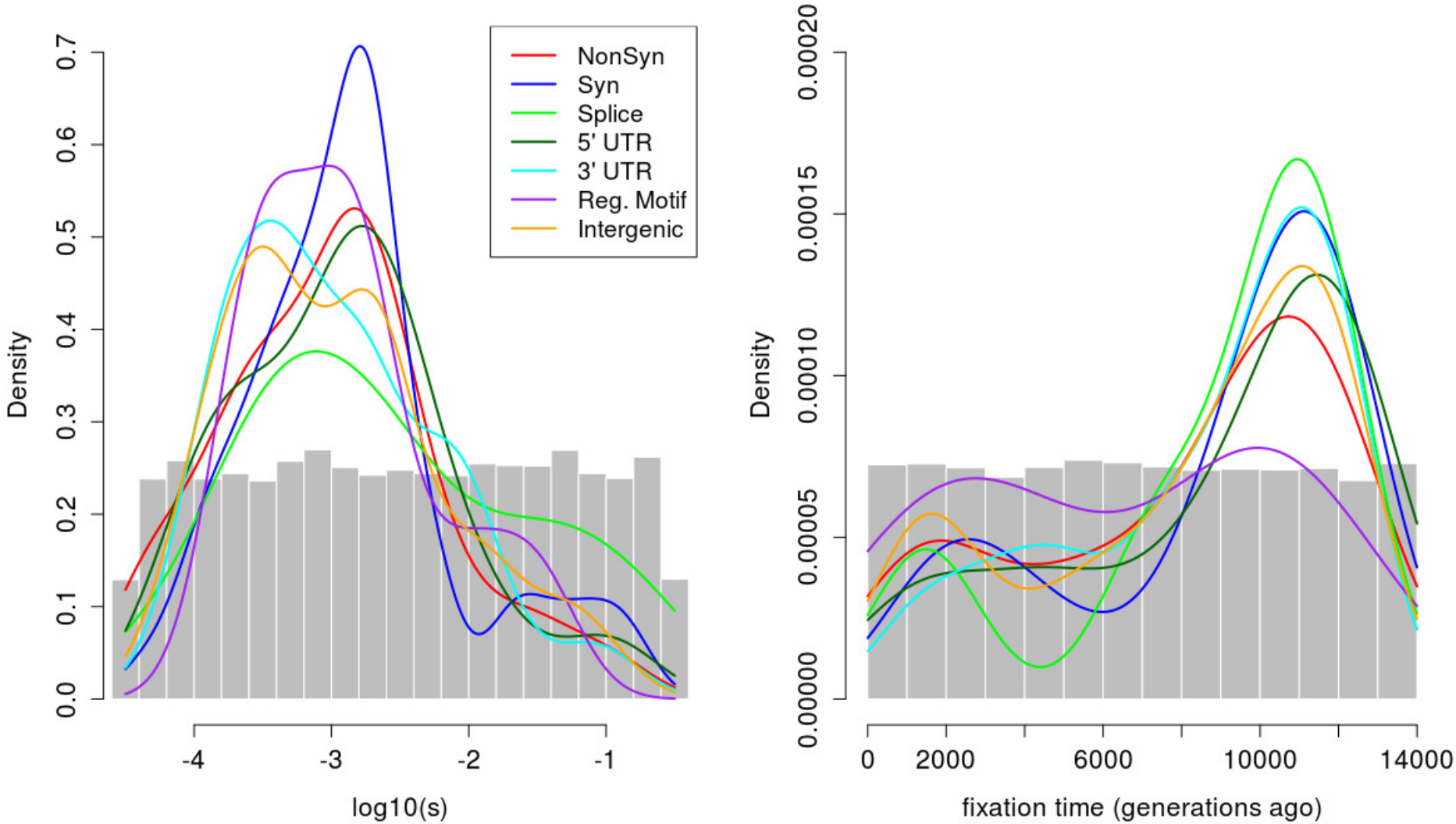


FIG. 2. Density of estimated posterior modes from ABC analyses under the positive selection model, across different genomic classes. The grey bars represent the prior used for each parameter.

Table 1. Modern-human specific changes that lead to an amino acid replacement, affect a splice site or are located in a UTR, and that: 1) have Bayes factors >5 in favor of selection and 2) are a good fit ($P>0.05$) to the selection model.

Position	Bayes factor	log(s)	t_S (generations)	P neutral	P selection	Class	Gene
chr1:27425606	18.48	-1.79286	3111	0.08	0.4	3' UTR	SLC9A1
chr1:27426756	11.65	-1.83326	4101	0.1	0.42	3' UTR	SLC9A1
chr1:27430334	8.46	-1.87366	3960	0.11	0.38	5' UTR	SLC9A1
chr1:78183739	6.98	-0.702133	10181	0.92	0.97	Splice	USP33
chr3:28476768	9.17	-0.702133	10181	0.88	0.98	Splice	ZCWPW2
chr3:28503157	8.6	-0.702133	9757	0.84	0.99	3' UTR	ZCWPW2
chr3:52009091	5.58	-0.823325	11878	1	1	5' UTR	ABHD14B
chr7:73113999	13.74	-1.1465	8202	0.69	1	3' UTR	STX1A
chr8:133771663	5.21	-3.16638	12444	0.04	0.18	5' UTR	TMEM71
chr10:50820543	21.58	-0.984915	9474	0.99	1	3' UTR	SLC18A3
chr11:117778820	11.6	-2.8028	6222	0.05	0.32	3' UTR	TMPRSS13
chr11:129769974	10.11	-2.8432	4667	0.07	0.1	3' UTR	TMPRSS13
chr11:129771185	13.44	-2.8028	4667	0.16	0.39	3' UTR	PRDM10
chr11:129771376	14.2	-2.8432	4667	0.17	0.42	3' UTR	PRDM10
chr11:129771773	16.55	-0.90412	10181	0.11	0.22	3' UTR	PRDM10
chr11:129772293	16.99	-0.90412	9757	0.2	0.53	NonSyn	PRDM10
chr11:64813918	6.53	-1.2273	11878	0.99	1	NonSyn	NAALADL1
chr11:64889467	5.38	-2.7624	11030	0.76	0.71	5' UTR	FAU
chr11:64889626	5.19	-2.7624	11030	0.76	0.69	5' UTR	FAU
chr11:64889767	6.72	-2.8028	11171	0.7	0.71	5' UTR	MRPL49
chr11:64893151	12.41	-2.92399	11312	0.11	0.3	NonSyn	MRPL49
chr11:64900743	10.19	-0.984915	11454	0.74	0.8	5' UTR	SYVN1
chr12:73058827	19.19	-3.44916	3253	0	0.2	3' UTR	TRHDE
chr12:73058885	7.05	-3.44916	3253	0.01	0.2	3' UTR	TRHDE
chr14:76249759	61.22	-1.2273	6081	0.05	0.15	NonSyn	TTLL5
chr17:47867139	5.17	-1.59088	11030	1	1	5' UTR	KAT7
chr19:42731306	20.65	-1.79286	3535	0.28	0.87	3' UTR	ZNF526
chr19:42732059	15.95	-1.79286	3535	0.39	0.91	3' UTR	ZNF526

NOTE.—Parameters listed are the posterior modes inferred using ABC. We also list the P-values for the fit to the neutral and selection models.

Projection analysis (Melinda Yang)

x is the derived allele frequency in the **reference** population.

At each segregating site in the reference population, assign a weight to that site in the **test** genome

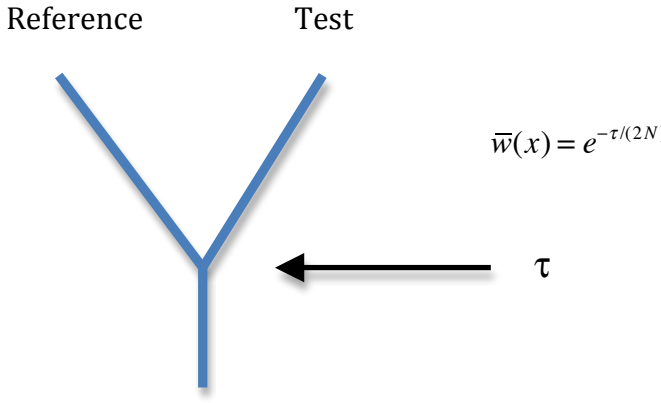
$w = 0$ if the test genome is homozygous ancestral

$w = \frac{1}{2x}$ if the test genome is heterozygous

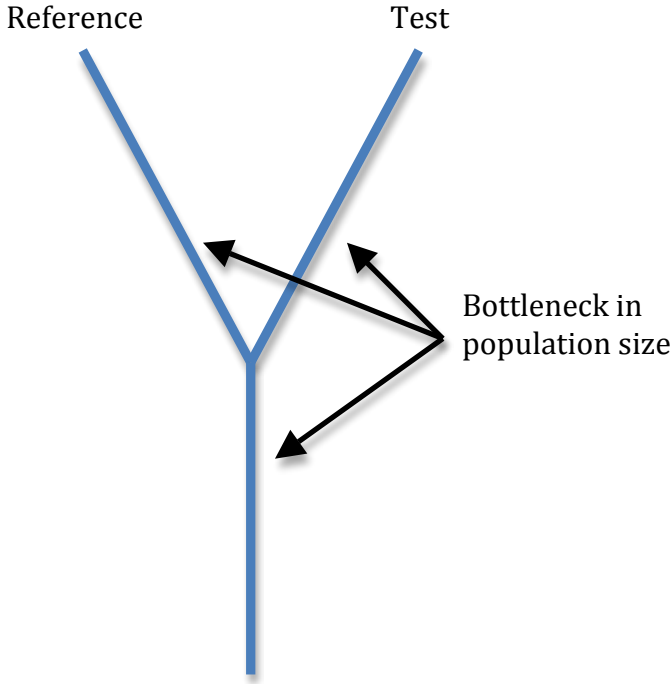
$w = \frac{1}{x}$ if the test genome is homozygous derived

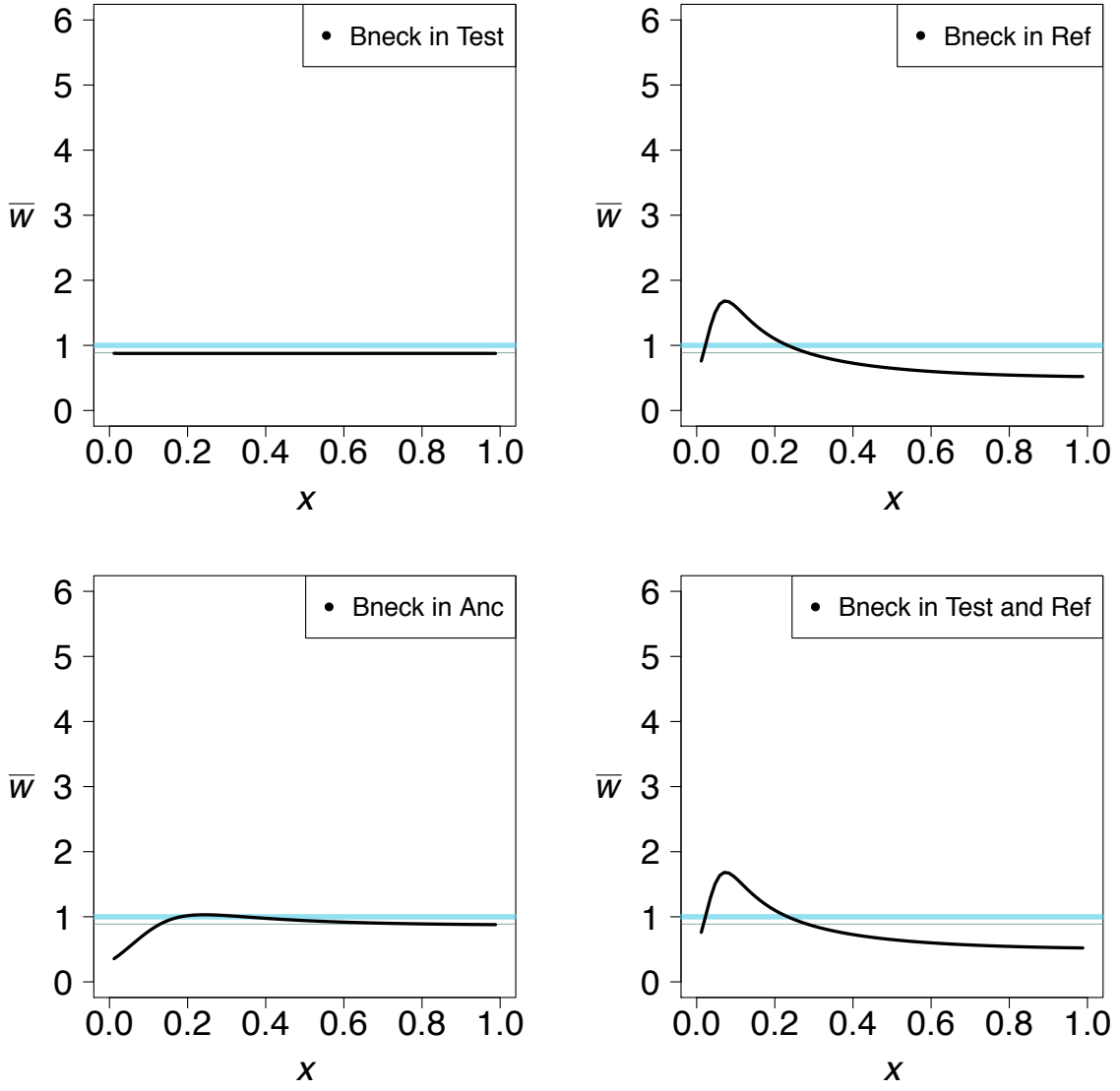
$\bar{w}(x)$ is the projection of the test genome on the reference population.

If the test genome is a random sample from the reference population $\bar{w}(x) = 1$

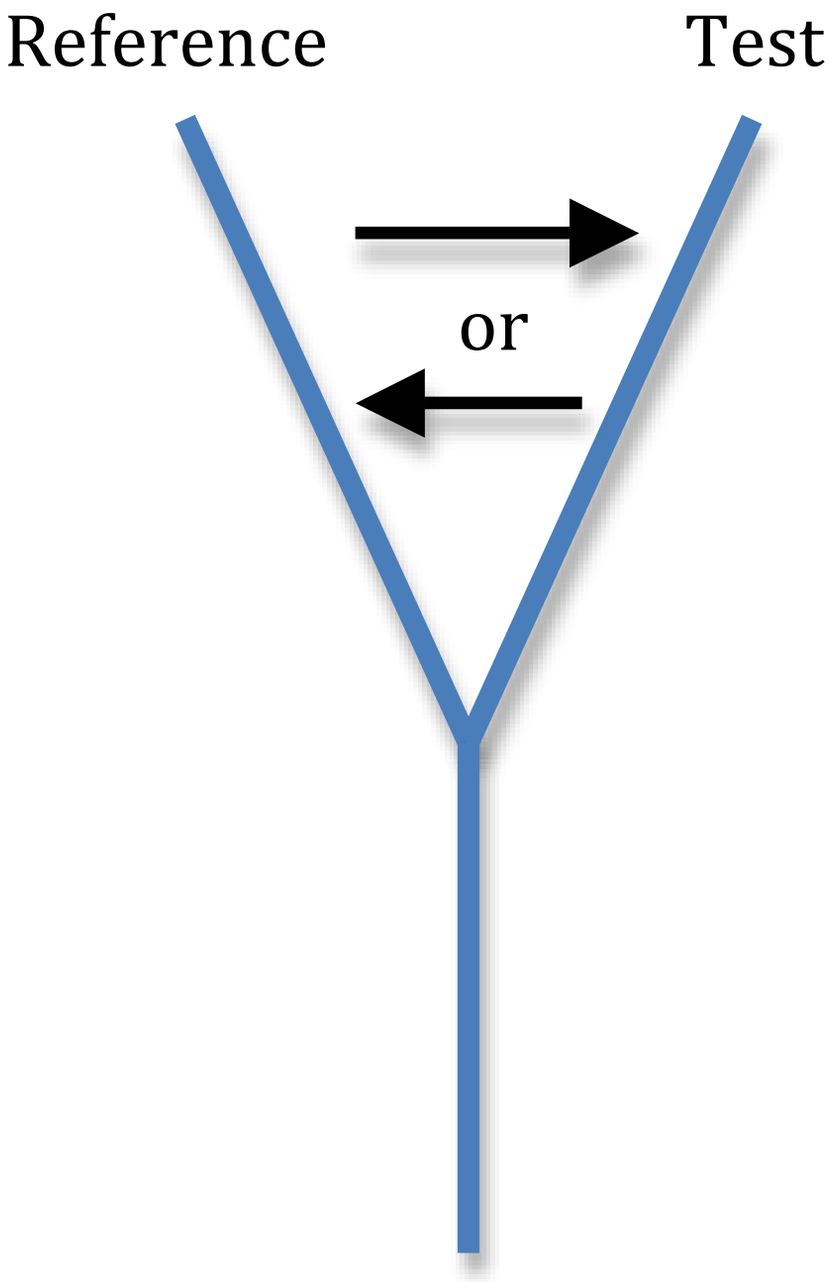


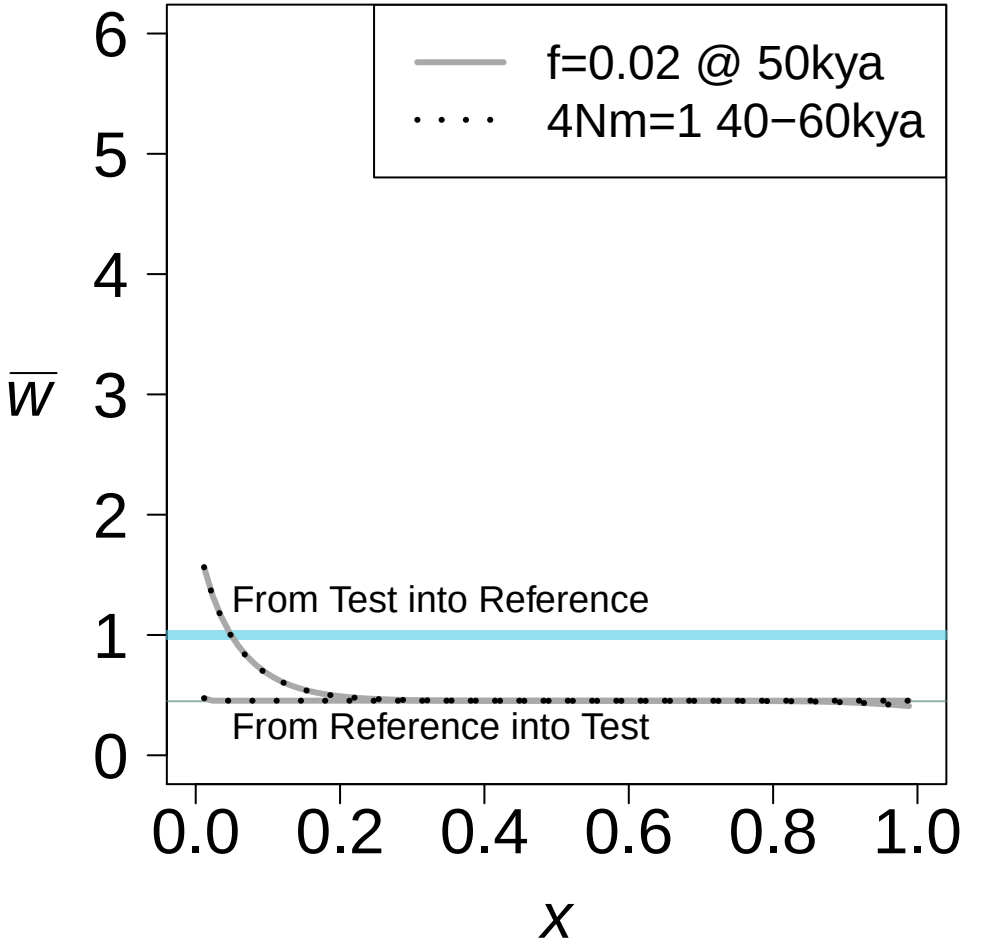
Chen et al. (2007, *Genetics*)

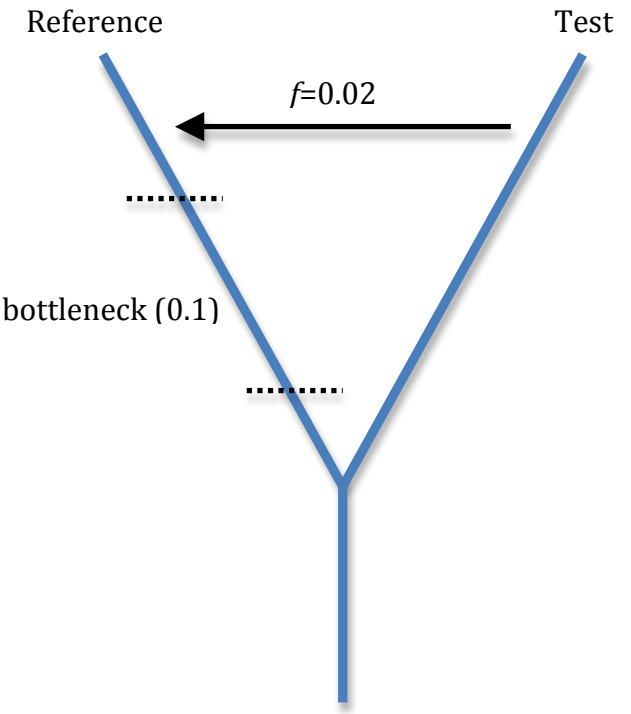

Reference
Test
Bottleneck in
population size

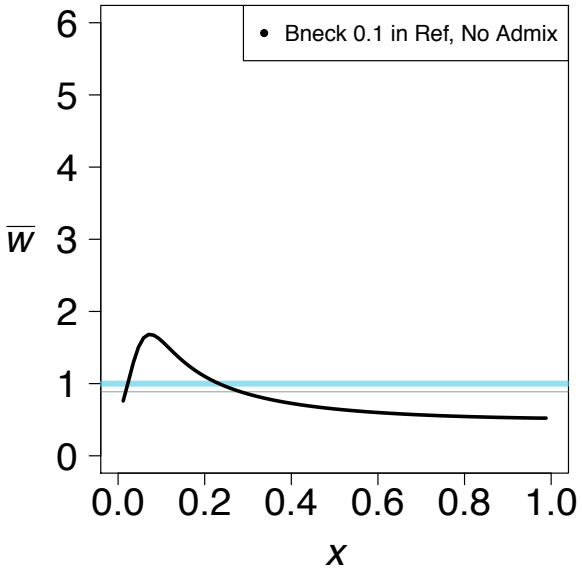


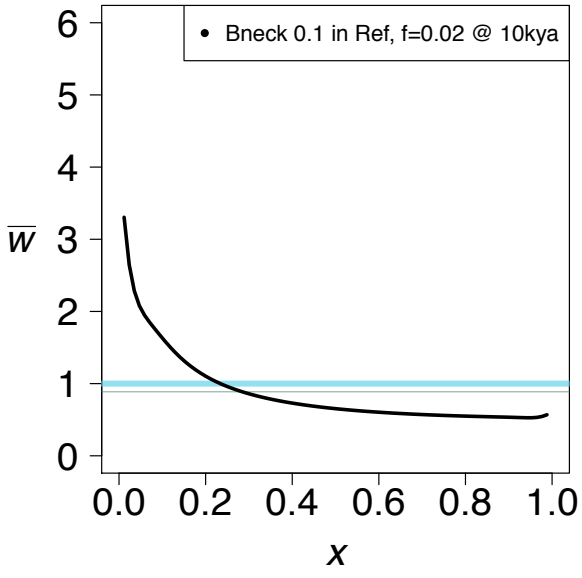
Population divergence time, τ=60,000 years, N=10,000, 25 years per generation
Bottleneck (0.1) in test and reference 20,000-50,000 years ago
Bottleneck (0.1) in ancestral population 70,000-100,000 years ago.

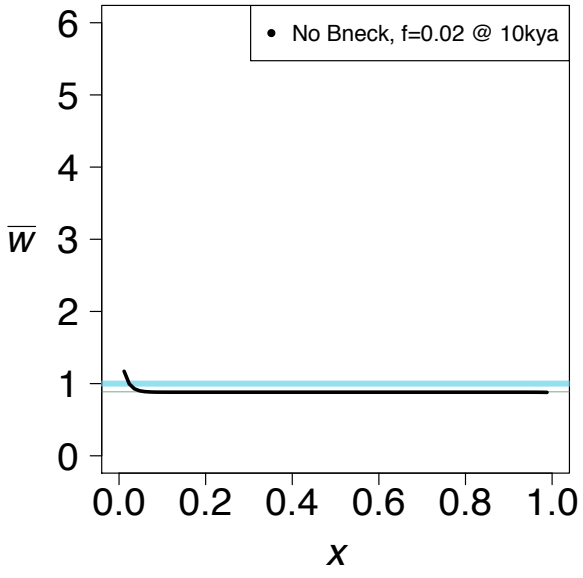

Reference
Test
or


f=0.02 @ 50kya
4Nm=1 40−60kya
From Test into Reference
From Reference into Test
$\overline{w}$
x
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0

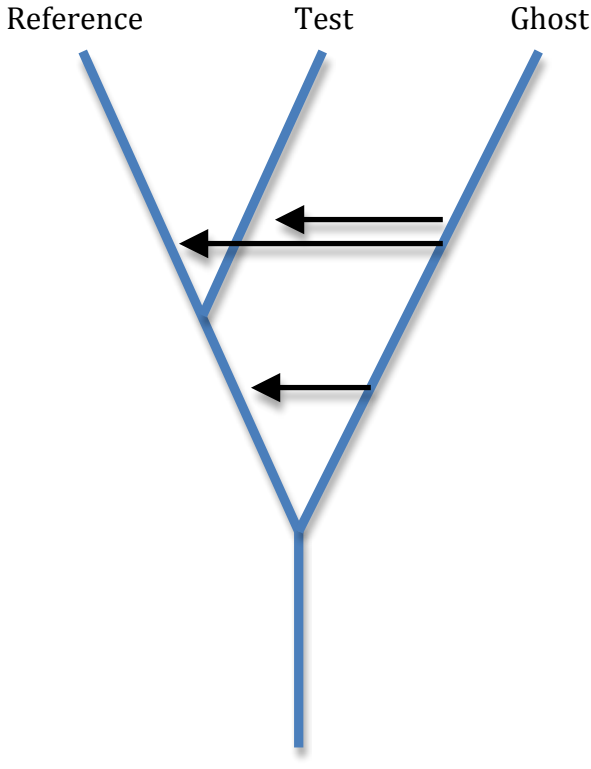

Reference
Test
f=0.02
bottleneck (0.1)

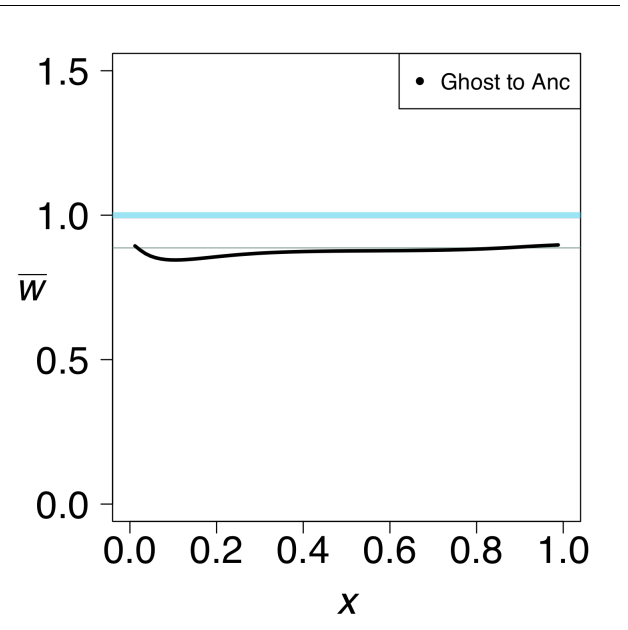


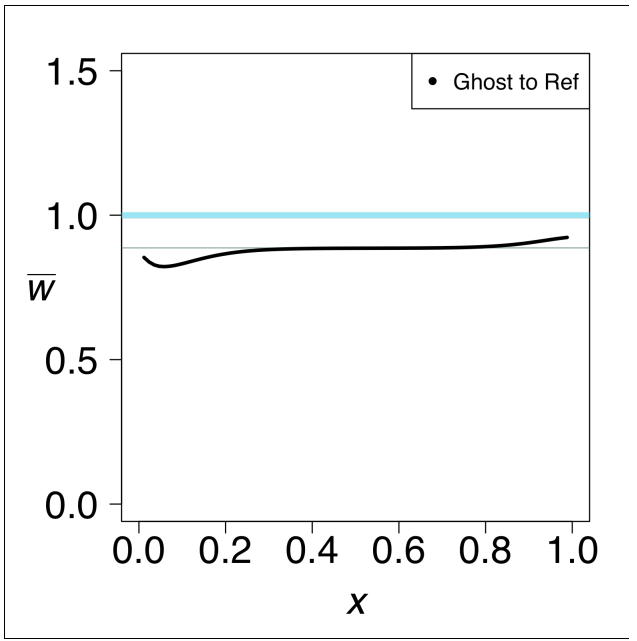


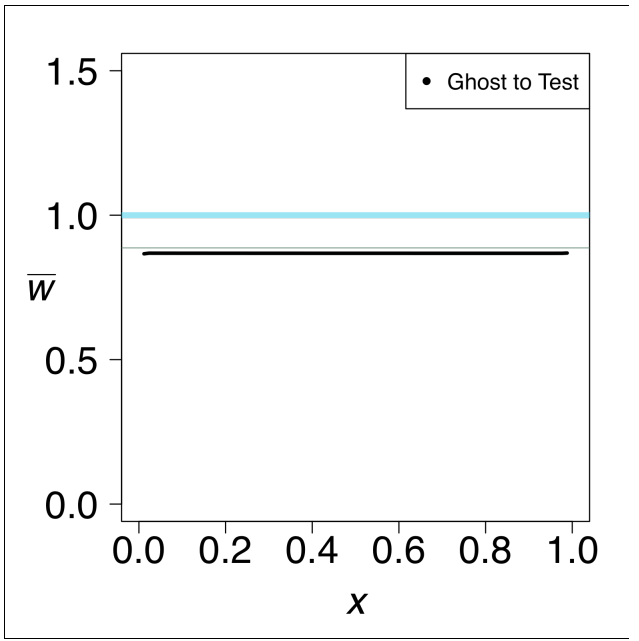


Population separation τ=60,000 years
Bottleneck (0.1) 20,000-50,000 years ago
Admixture from test to reference (f=0.02) 10,000 years ago

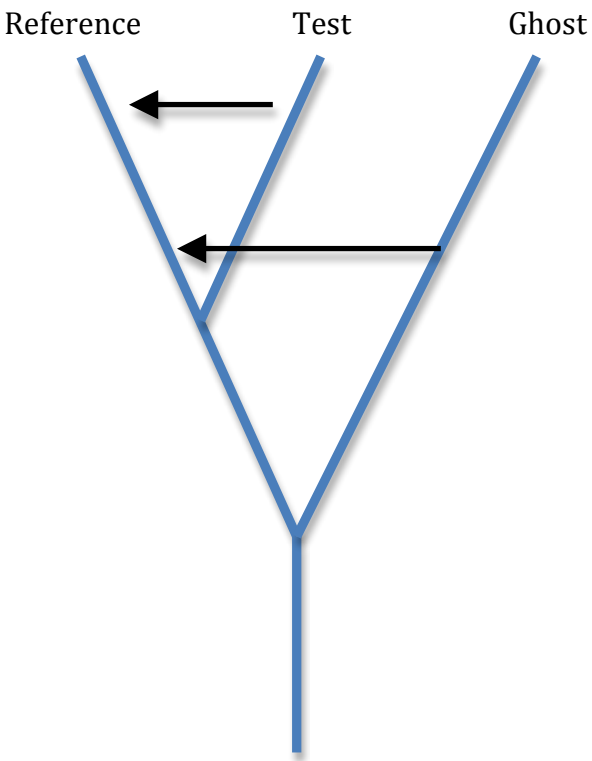

Reference
Test
Ghost

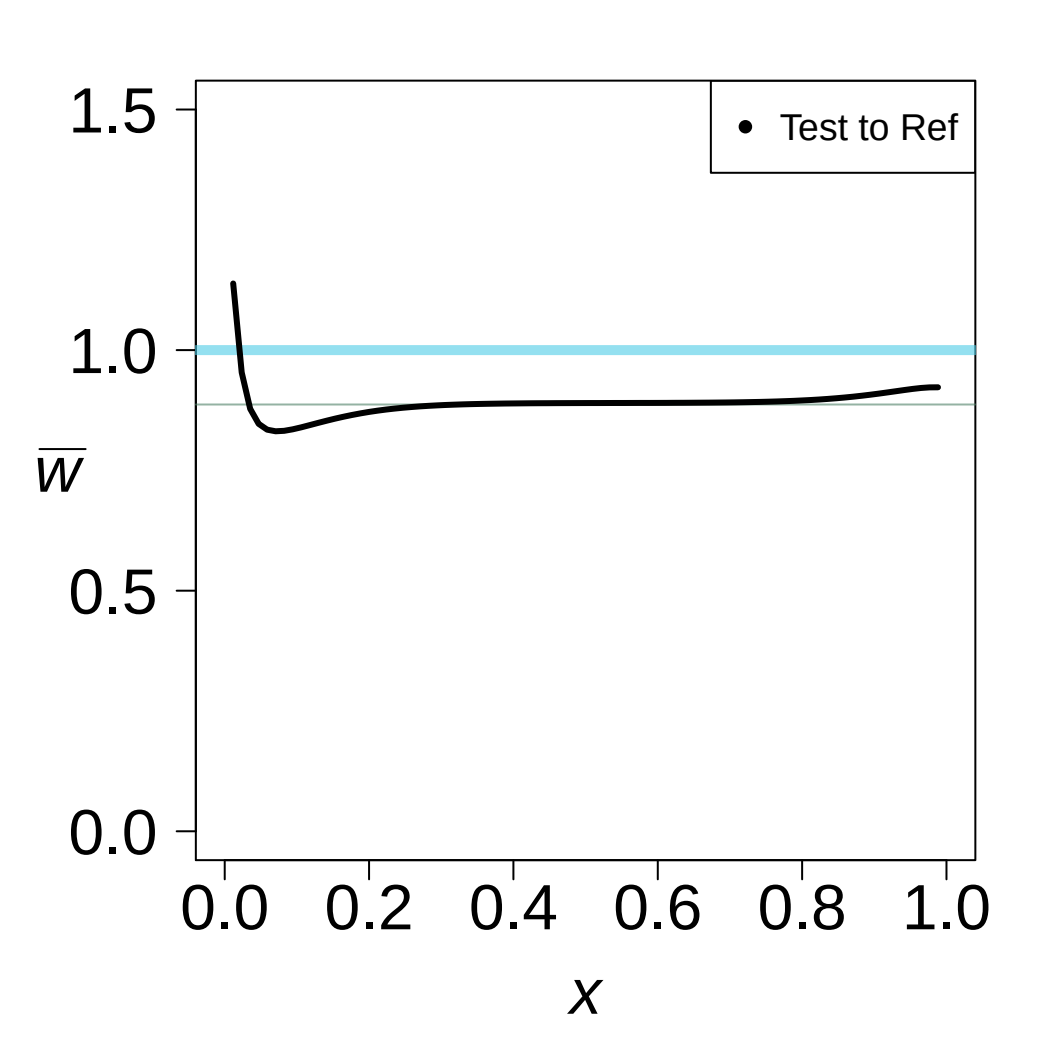


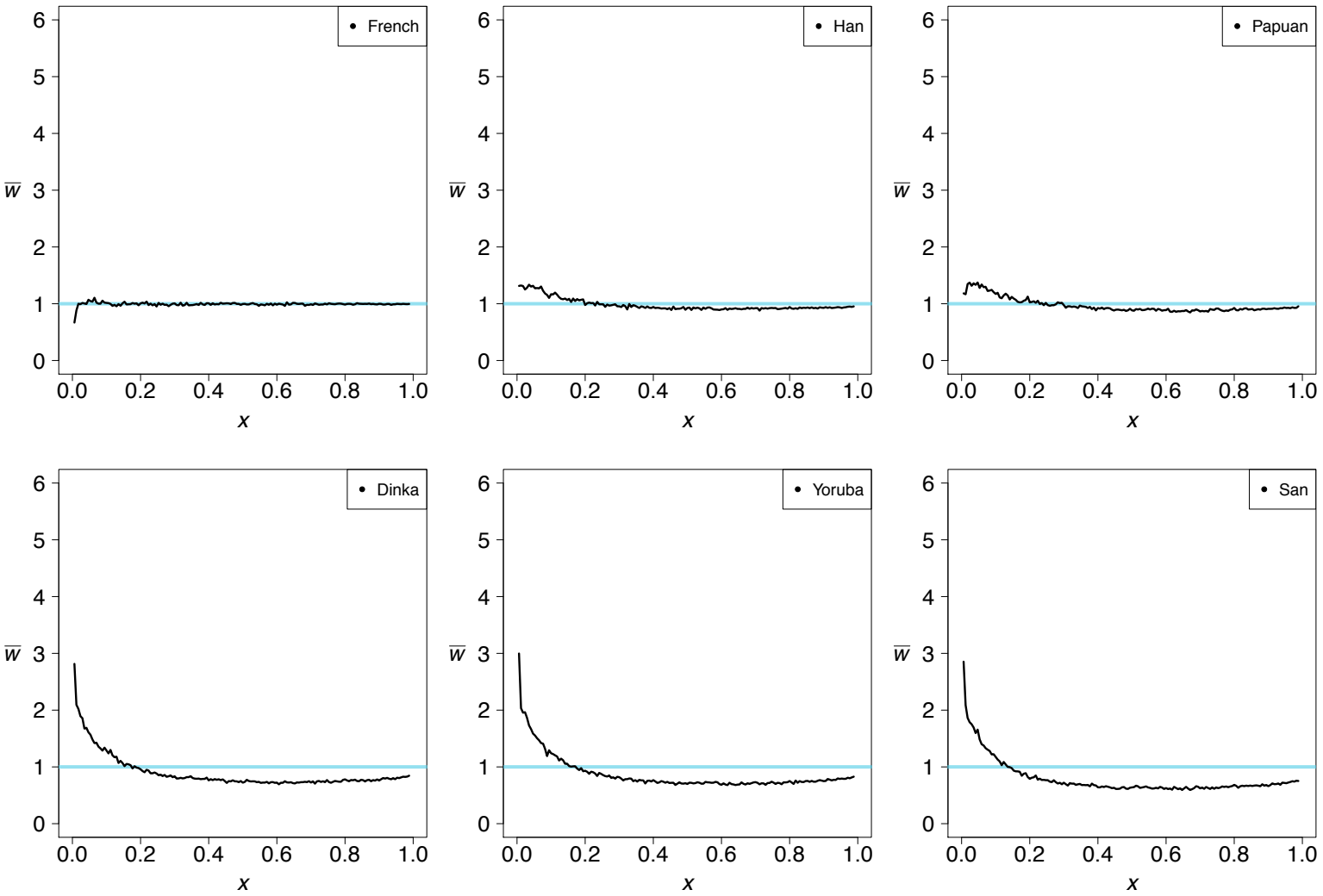




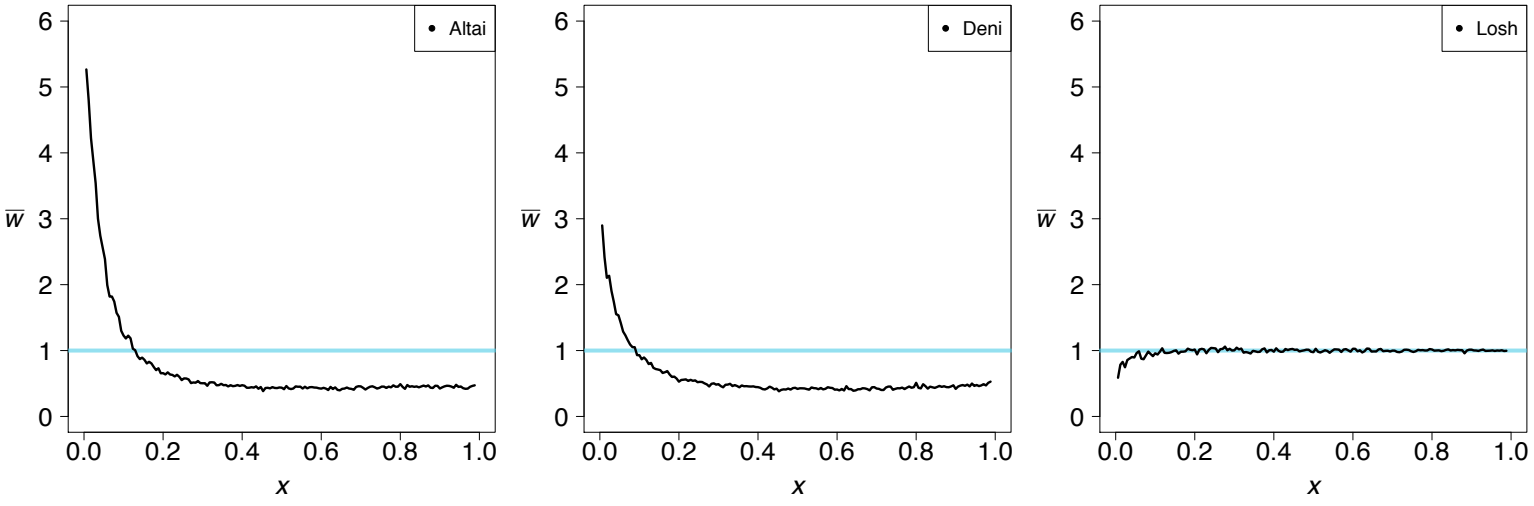
Test and reference diverged 60,000 years ago
Ghost diverged 400,000 years ago
Admixture rate f=0.02, 50,000 years ago into test or reference, 70,000 years ago into ancestor


Reference
Test
Ghost

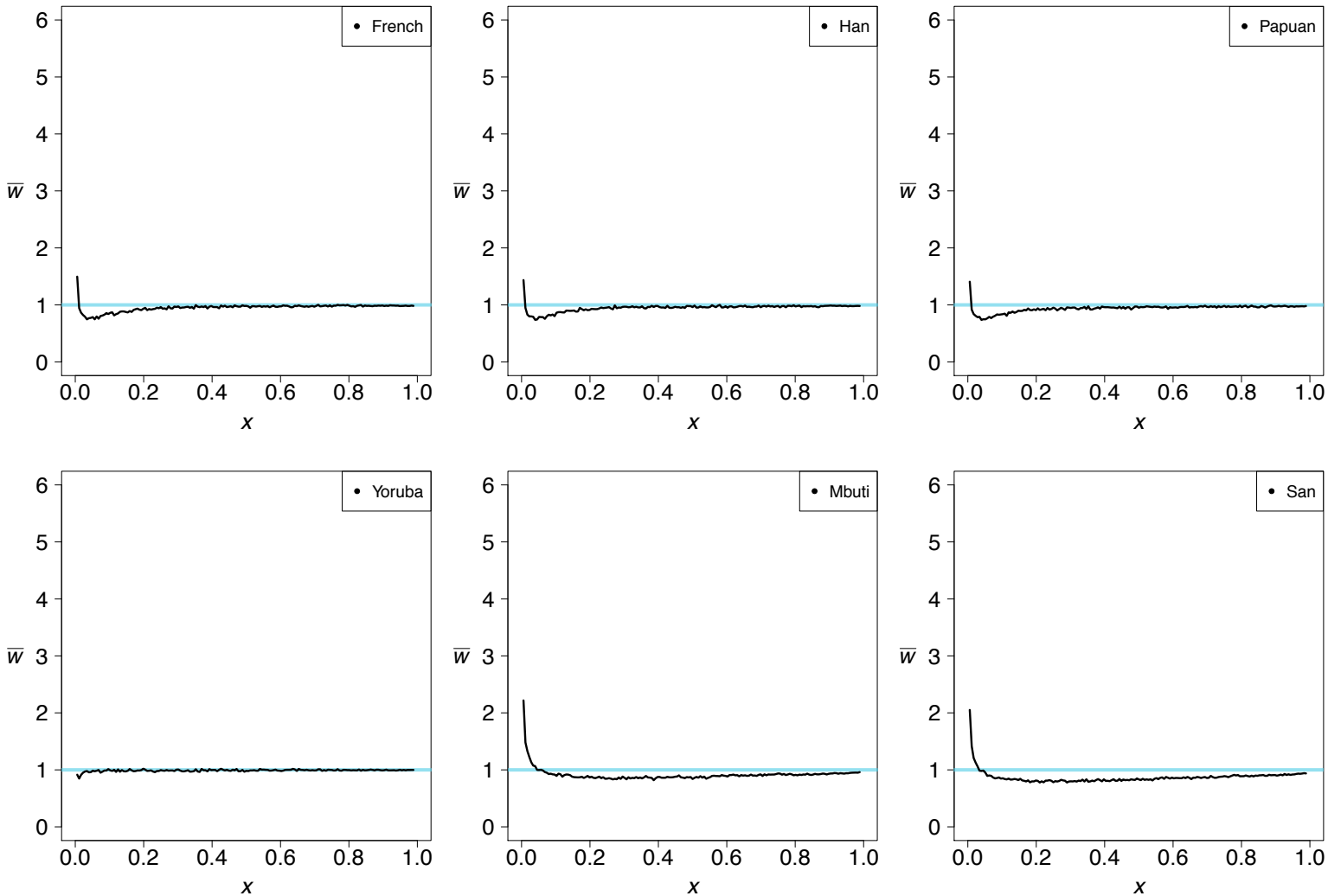

Test to Ref
1.5
1.0
0.5
0.0
$\overline{w}$
0.0
0.2
0.4
0.6
0.8
1.0
x



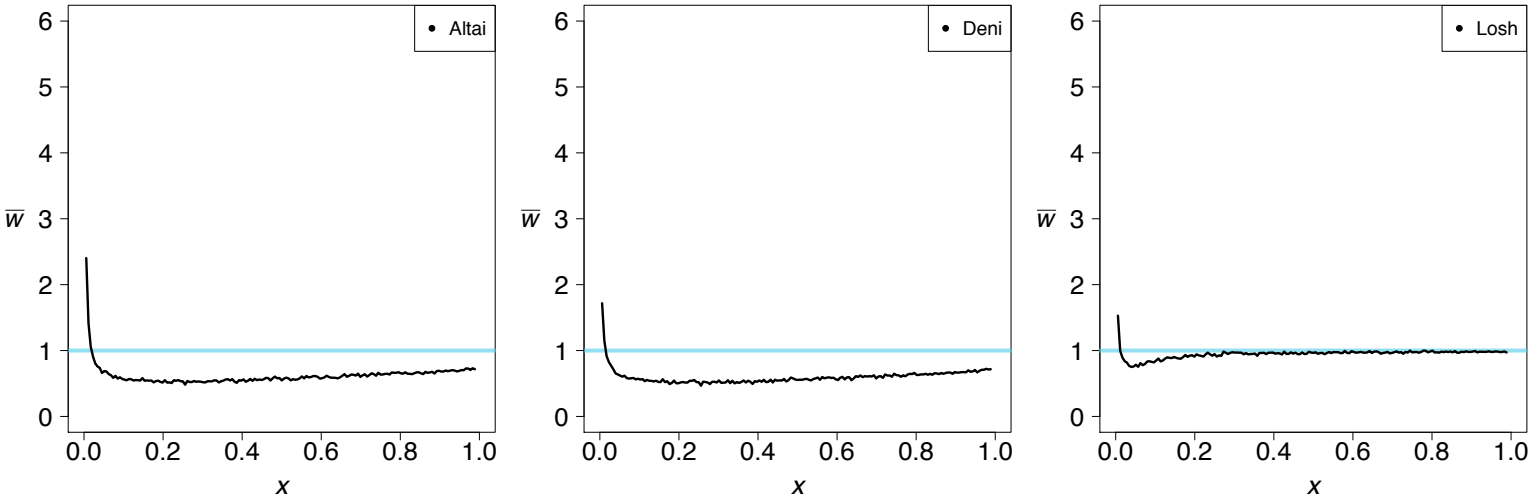
CEU reference population



CEU reference population



YRI reference population



YRI reference population