Supplementary Material for: Toward a new history and geography of human genes informed by ancient DNA

Joseph K. Pickrell[1,2], David Reich[3,4,5]

[1] New York Genome Center, New York, NY

[2] Department of Biological Sciences, Columbia University, New York, NY

[3] Department of Genetics, Harvard Medical School, Boston, MA

[4] Howard Hughes Medical Institute, Harvard Medical School, Boston, MA

[5] Broad Institute of MIT and Harvard, Cambridge, MA

June 9, 2014

1 Simulations of smooth linear declines in heterozygosity under admixture models.

In this section, we detail the models presented in Figure 1B and 1C.

1.1 Approximation of heterozygosity under admixture models

Consider an admixed population A that has admixture proportions α and $1-\alpha$ from two ancestral populations: Y and Z, respectively. The heterozygosity in A is the expected number of differences between two random haplotypes sampled from the population. To approximate this, we let the time of admixture be the present (in the models later on, we have admixture at 60 generations in the past and 800 generations in the past, so in some cases this will be a poor approximation). With this simplification:

$$H_A \approx \alpha^2 H_Y + 2\alpha(1-\alpha)H_x^{YZ} + (1-\alpha)^2 H_Z, \tag{1}$$

where H_Y is the heterozygosity in population Y, H_Z is the corresponding quantity for population Z, and H_x^{YZ} is the heterozygosity when sampling a single haplotype from population Y and a single haplotype from population Z. Rearranging this gives:

$$\alpha^2(H_Y - 2H_x^{YZ} + H_Z) + \alpha(2H_x^{YZ} - 2H_Z) + H_Z - H_A = 0. \tag{2}$$

Solving this for α gives the admixture proportion necessary to produce a given H_A.

1.2 Model with severe bottlenecks and recent admixture

In the demographic model in Figure 1B in the main text (reproduced in the upper panel of Supplementary Figure 1), all 42 simulated populations are a mixture between three ancestral populations. We set the ancestral effective population size to 10,000 individuals and the mutation rate to $\mu = 2.5 \times 10^{-8}$. To apply Equation 2, we need the heterozygosity in populations 1-3 (H_1, H_2, and H_3, labeled from left to right in Supplementary Figure 1), and the relevant cross-population heterozygosities (H_x^{12}, H_x^{23}) (in Supplementary Figure 1 there are no populations with ancestry from both populations 1 and 3). H_1 is 0.001. The extreme bottlenecks consist of a reduction in population size to 25 individuals for 10 generations, so H_2 is approximately $(1-0.02)^{10}H_1 = 0.00081$, and H_3 is approximately $(1-0.02)^{20}H_1 = 0.00067$. The cross-population heterozygosities are $H_x^{12} = 0.001 + 4400\mu$ and $H_x^{23} = 0.00081 + 2200\mu$. We applied Equation 2 to get a set of 42 populations with an approximately linear decline in heterozygosity; these admixture proportions are shown in the lower panel of Supplementary Figure 1.

1.3 Model with no bottlenecks and ancient admixture

In the demographic model in Figure 1C in the main text (reproduced in the upper panel of Supplementary Figure 2), all simulated populations are again a mixture between three ancestral pop-

ulations. In this case, however, the first two populations experienced admixture with an archaic population 800 generations in the past, with admixture proportions β_1 and β_2, respectively (these are 20% and 9% in the simulations). We again set the effective population size to 10,000 individuals and the mutation rate to $\mu = 2.5 \times 10^{-8}$. Again, we need the expected heterozygosities in populations 1-3 (H_1, H_2, and H_3, labeled from left to right in Supplementary Figure 2), and the relevant expected cross-population heterozygosities (H_x^{12}, H_x^{23}). With the approximation that the archaic admixture occurred in the present:

$$H_1 = \beta_1^2 0.001 + 2\beta_1(1-\beta_1)[0.001 + 32000\mu] + (1-\beta_1)^2 0.001 \quad (3)$$

$$H_2 = \beta_2^2 0.001 + 2\beta_2(1-\beta_2)[0.001 + 32000\mu] + (1-\beta_2)^2 0.001 \quad (4)$$

$$H_3 = 0.001 \quad (5)$$

$$H_x^{12} = \beta_1\beta_2 0.001 + \beta_1(1-\beta_2)[0.001 + 32000\mu] + (1-\beta_1)\beta_2[0.001 + 4400\mu] + (1-\beta_1)(1-\beta_2)0.001 \quad (6)$$

$$H_x^{23} = \beta_2[0.001 + 32000\mu] + (1-\beta_2)[0.001 + 2200\mu] \quad (7)$$

We used these approximations and Equation 2 to get a set of 42 populations with an approximately linear decline in heterozygosity; these admixture proportions are shown in the lower panel of Supplementary Figure 2.

1.4 Simulation parameters

To generate Figure 1, we performed simulations of different demographic parameters using ms (Hudson, 2002). For the serial bottleneck model in Figure 1A, following a demographic model similar to DeGiorgio et al., 2009, we used the following command to generate 20 haplotypes from each of 42 populations:

```
ms 840 1 -t 100 -r 10 100000 -I 42 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20
20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 -en
0.00114146341463 42 0.025 -ej 0.00134146341463 42 41 -en 0.00248292682927 41 0.025
-ej 0.00268292682927 41 40 -en 0.0038243902439 40 0.025 -ej 0.0040243902439 40 39 -en
0.00516585365854 39 0.025 -ej 0.00536585365854 39 38 -en 0.00650731707317 38 0.025
-ej 0.00670731707317 38 37 -en 0.0078487804878 37 0.025 -ej 0.0080487804878 37 36 -en
0.00919024390244 36 0.025 -ej 0.00939024390244 36 35 -en 0.0105317073171 35 0.025 -ej
0.0107317073171 35 34 -en 0.0118731707317 34 0.025 -ej 0.0120731707317 34 33 -en 0.013214634146
33 0.025 -ej 0.0134146341463 33 32 -en 0.014556097561 32 0.025 -ej 0.014756097561 32
31 -en 0.0158975609756 31 0.025 -ej 0.0160975609756 31 30 -en 0.0172390243902 30 0.025
-ej 0.0174390243902 30 29 -en 0.0185804878049 29 0.025 -ej 0.0187804878049 29 28 -en
0.0199219512195 28 0.025 -ej 0.0201219512195 28 27 -en 0.0212634146341 27 0.025 -ej
0.0214634146341 27 26 -en 0.0226048780488 26 0.025 -ej 0.0228048780488 26 25 -en 0.023946341463
25 0.025 -ej 0.0241463414634 25 24 -en 0.025287804878 24 0.025 -ej 0.025487804878 24
23 -en 0.0266292682927 23 0.025 -ej 0.0268292682927 23 22 -en 0.0279707317073 22 0.025
-ej 0.0281707317073 22 21 -en 0.029312195122 21 0.025 -ej 0.029512195122 21 20 -en
0.0306536585366 20 0.025 -ej 0.0308536585366 20 19 -en 0.0319951219512 19 0.025 -ej
```

```
0.0321951219512 19 18 -en 0.0333365853659 18 0.025 -ej 0.0335365853659 18 17 -en 0.034678048780
17 0.025 -ej 0.0348780487805 17 16 -en 0.0360195121951 16 0.025 -ej 0.0362195121951
16 15 -en 0.0373609756098 15 0.025 -ej 0.0375609756098 15 14 -en 0.0387024390244 14
0.025 -ej 0.0389024390244 14 13 -en 0.040043902439 13 0.025 -ej 0.040243902439 13 12
-en 0.0413853658537 12 0.025 -ej 0.0415853658537 12 11 -en 0.0427268292683 11 0.025
-ej 0.0429268292683 11 10 -en 0.0440682926829 10 0.025 -ej 0.0442682926829 10 9 -en
0.0454097560976 9 0.025 -ej 0.0456097560976 9 8 -en 0.0467512195122 8 0.025 -ej 0.0469512195122
8 7 -en 0.0480926829268 7 0.025 -ej 0.0482926829268 7 6 -en 0.0494341463415 6 0.025
-ej 0.0496341463415 6 5 -en 0.0507756097561 5 0.025 -ej 0.0509756097561 5 4 -en 0.0521170731707
4 0.025 -ej 0.0523170731707 4 3 -en 0.0534585365854 3 0.025 -ej 0.0536585365854 3 2
-en 0.0548 2 0.025 -ej 0.055 2 1
```

To simulate the demographic scenario in Supplementary Figure 1 (Figure 1B in the main text), we used the following command. The output of this command is 20 haplotypes from 45 populations (the three ancestral populations and the 42 observed populations). For plotting purposes, we ignore the three ancestral populations (populations 1, 23, and 45 in the command).

```
ms 900 1 -t 100 -r 10 100000 -I 45 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20
20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20
20 20 -em 0.001475 2 1 16088.5663 -ej 0.0015 2 23 -em 0.001475 3 1 14874.5283 -ej 0.0015
3 23 -em 0.001475 4 1 13756.3749 -ej 0.0015 4 23 -em 0.001475 5 1 12714.4406 -ej 0.0015
5 23 -em 0.001475 6 1 11734.9843 -ej 0.0015 6 23 -em 0.001475 7 1 10807.9558 -ej 0.0015
7 23 -em 0.001475 8 1 9925.7415 -ej 0.0015 8 23 -em 0.001475 9 1 9082.4094 -ej 0.0015
9 23 -em 0.001475 10 1 8273.2310 -ej 0.0015 10 23 -em 0.001475 11 1 7494.3650 -ej 0.0015
11 23 -em 0.001475 12 1 6742.6404 -ej 0.0015 12 23 -em 0.001475 13 1 6015.4032 -ej
0.0015 13 23 -em 0.001475 14 1 5310.4056 -ej 0.0015 14 23 -em 0.001475 15 1 4625.7243
-ej 0.0015 15 23 -em 0.001475 16 1 3959.6978 -ej 0.0015 16 23 -em 0.001475 17 1 3310.8797
-ej 0.0015 17 23 -em 0.001475 18 1 2678.0013 -ej 0.0015 18 23 -em 0.001475 19 1 2059.9424
-ej 0.0015 19 23 -em 0.001475 20 1 1455.7085 -ej 0.0015 20 23 -em 0.001475 21 1 864.4112
-ej 0.0015 21 23 -em 0.001475 22 1 285.2536 -ej 0.0015 22 23 -em 0.001475 24 23 21885.6254
-ej 0.0015 24 45 -em 0.001475 25 23 19913.4599 -ej 0.0015 25 45 -em 0.001475 26 23
18230.7956 -ej 0.0015 26 45 -em 0.001475 27 23 16738.4399 -ej 0.0015 27 45 -em 0.001475
28 23 15383.5083 -ej 0.0015 28 45 -em 0.001475 29 23 14133.8441 -ej 0.0015 29 45 -em
0.001475 30 23 12968.1634 -ej 0.0015 30 45 -em 0.001475 31 23 11871.5139 -ej 0.0015
31 45 -em 0.001475 32 23 10832.9150 -ej 0.0015 32 45 -em 0.001475 33 23 9844.0210 -ej
0.0015 33 45 -em 0.001475 34 23 8898.3127 -ej 0.0015 34 45 -em 0.001475 35 23 7990.5827
-ej 0.0015 35 45 -em 0.001475 36 23 7116.5926 -ej 0.0015 36 45 -em 0.001475 37 23 6272.8385
-ej 0.0015 37 45 -em 0.001475 38 23 5456.3837 -ej 0.0015 38 45 -em 0.001475 39 23 4664.7385
-ej 0.0015 39 45 -em 0.001475 40 23 3895.7698 -ej 0.0015 40 45 -em 0.001475 41 23 3147.6338
-ej 0.0015 41 45 -em 0.001475 42 23 2418.7240 -ej 0.0015 42 45 -em 0.001475 43 23 1707.6300
-ej 0.0015 43 45 -em 0.001475 44 23 1013.1060 -ej 0.0015 44 45 -en 0.02725 45 0.0025
-ej 0.0275 45 23 -en 0.05475 23 0.0025 -ej 0.055 23 1
```

To simulate the scenario in Supplementary Figure 2 (Figure 1C in the main text), we used the following command. The output of this command is 20 haplotypes from 46 populations (the archaic population, the three ancestral populations and the 42 observed populations). For plotting purposes, we ignore the archaic population and the three ancestral populations (populations 1, 23, 45 and 46 in the command):

```
ms 920 1 -t 100 -r 10 100000 -I 46 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20
20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20
20 20 20 -em 0.001475 2 1 23612.4348 -ej 0.0015 2 23 -em 0.001475 3 1 21403.0630 -ej
0.0015 3 23 -em 0.001475 4 1 19523.4041 -ej 0.0015 4 23 -em 0.001475 5 1 17859.0223
-ej 0.0015 5 23 -em 0.001475 6 1 16349.4741 -ej 0.0015 6 23 -em 0.001475 7 1 14958.2141
-ej 0.0015 7 23 -em 0.001475 8 1 13661.1475 -ej 0.0015 8 23 -em 0.001475 9 1 12441.3941
-ej 0.0015 9 23 -em 0.001475 10 1 11286.5839 -ej 0.0015 10 23 -em 0.001475 11 1 10187.3311
-ej 0.0015 11 23 -em 0.001475 12 1 9136.3139 -ej 0.0015 12 23 -em 0.001475 13 1 8127.6902
-ej 0.0015 13 23 -em 0.001475 14 1 7156.7095 -ej 0.0015 14 23 -em 0.001475 15 1 6219.4477
-ej 0.0015 15 23 -em 0.001475 16 1 5312.6183 -ej 0.0015 16 23 -em 0.001475 17 1 4433.4364
-ej 0.0015 17 23 -em 0.001475 18 1 3579.5175 -ej 0.0015 18 23 -em 0.001475 19 1 2748.8016
-ej 0.0015 19 23 -em 0.001475 20 1 1939.4941 -ej 0.0015 20 23 -em 0.001475 21 1 1150.0205
-ej 0.0015 21 23 -em 0.001475 22 1 378.9902 -ej 0.0015 22 23 -em 0.001475 24 23 35742.7437
-ej 0.0015 24 45 -em 0.001475 25 23 31262.2470 -ej 0.0015 25 45 -em 0.001475 26 23
28095.0878 -ej 0.0015 26 45 -em 0.001475 27 23 25500.8143 -ej 0.0015 27 45 -em 0.001475
28 23 23249.1179 -ej 0.0015 28 45 -em 0.001475 29 23 21232.0517 -ej 0.0015 29 45 -em
0.001475 30 23 19388.6924 -ej 0.0015 30 45 -em 0.001475 31 23 17680.6523 -ej 0.0015
31 45 -em 0.001475 32 23 16081.8924 -ej 0.0015 32 45 -em 0.001475 33 23 14573.7869
-ej 0.0015 33 45 -em 0.001475 34 23 13142.4676 -ej 0.0015 34 45 -em 0.001475 35 23
11777.2781 -ej 0.0015 35 45 -em 0.001475 36 23 10469.8202 -ej 0.0015 36 45 -em 0.001475
37 23 9213.3360 -ej 0.0015 37 45 -em 0.001475 38 23 8002.2910 -ej 0.0015 38 45 -em
0.001475 39 23 6832.0854 -ej 0.0015 39 45 -em 0.001475 40 23 5698.8467 -ej 0.0015 40
45 -em 0.001475 41 23 4599.2783 -ej 0.0015 41 45 -em 0.001475 42 23 3530.5463 -ej 0.0015
42 45 -em 0.001475 43 23 2490.1931 -ej 0.0015 43 45 -em 0.001475 44 23 1476.0715 -ej
0.0015 44 45 -em 0.02 23 46 3507.578 -em 0.020025 23 46 0 -em 0.02 1 46 8000 -em 0.020025
1 46 0 -ej 0.0275 45 23 -ej 0.055 23 1 -ej 0.4 46 1
```

1.5 Calculation of summary statistics

For each population in each simulation, we calculated heterozygosity by pairing the 20 haplotypes from the population to form 10 diploid individuals, then counting the average number of heterozygous sites per individual. We then averaged this across simulations. These averages are shown in Figure 1 in the main text.

We also calculated the measure of linkage disequilibrium r^2 for all pairs of SNPs in each simulation. To summarize the qualitative LD patterns in different populations, following DeGiorgio

et al. [2], for each population in each simulation, we identified all SNPs separated by a distance of 10-11kb and calculated the average r^2 between these SNPs. We then averaged this across all simulations separately for each population. These averages are shown In Supplementary Figure 3. In all three scenarios, we recapitulate the qualitative pattern of an increase in LD with distance from a reference population. Note that this implies that archaic admixture *decreases* levels of LD (Supplementary Figure 3F), which seems counterintuitive. However, this is simply a consequence of a shift in the allele frequency spectrum to lower frequency SNPs in populations with archaic admixture (DeGiorgio et al. [2] see the same effect of a decrease in LD with archaic admixture in their simulations, which they attribute to an interaction between population bottlenecks and archaic admixture. However, the demographic model in Supplementary Figure 3C has no bottlenecks, so this cannot be the case in our simulations).

2 Admixture tests

To generate Figure 2, we combined SNP data generated on Illumina chips from a number of sources (Li et al., 2008; Altshuler et al., 2010; Behar et al., 2010; Henn et al., 2011; Schlebusch et al., 2012). We excluded the Jewish populations from Behar et al. 2010. In total, the data set consisted of 103 populations and 256,540 SNPs (Supplementary Table 1).

We used admixtools (Patterson et al., 2012) to compute all possible f_3 statistics of the form $f_3(A; B, C)$ on these populations. We considered an f_3 statistic to be significant evidence for admixture in population A if it was at least three standard errors less than zero (corresponding to a P-value of about 0.001). In Supplementary Table 1, we list all populations, their approximate latitudes and longitudes, and the representatives of the admixing populations (if any). These representatives were chosen as the population pair B and C that give the minimum f_3 statistic.

Table 1**: Populations tested for admixture.** For each population, we show the name of the publication from which we received the data, the name of the population according to the publication from which the data was obtained, the country of the population, the region we into which we classified the population, the approximate latitude and longitude of the population, and the names of the two source populations that give the most negative f_3 statistics. This latter information is only shown for populations with at least one significantly negative f_3 statistic.

Pub.	Population	Country	Region	Lat	Long	Source 1	Source 2
[6]	AFAR	Ethiopia	EASTAF	12	41	Sardinian	SUDANESE
[6]	AMHARA	Ethiopia	EASTAF	10	39	Sardinian	SUDANESE
[6]	ANUAK	Ethiopia	EASTAF	8	34	ARIBLACKSMITH	SUDANESE
[6]	ARIBLACKSMITH	Ethiopia	EASTAF	5	36		
[6]	ARICULTIVATOR	Ethiopia	EASTAF	3	40	Juhoansi	Sardinian
[6]	ESOMALI	Ethiopia	EASTAF	9	42	Sardinian	SUDANESE
[6]	GUMUZ	Ethiopia	EASTAF	10.8	35.6		
[6]	OROMO	Ethiopia	EASTAF	8	37	Sardinian	SUDANESE
[6]	SOMALI	Ethiopia	EASTAF	5.2	46.2	Sardinian	SUDANESE
[6]	SUDANESE	Sudan	EASTAF	12.9	30.2		
[6]	TYGRAY	Ethiopia	EASTAF	13	38	Sardinian	SUDANESE
[6]	WOLAYTA	Ethiopia	EASTAF	6	39	MbutiPygmy	Sardinian
[7]	Khwe	Angola	SAF	-17.4	23.0	Mozabite	Juhoansi
[7]	Xun	Angola	SAF	-14.7	17.7	Yoruba	Juhoansi
[7]	GuiGana	Botswana	SAF	-23.7	24.7	Yoruba	Juhoansi
[7]	Juhoansi	Namibia	SAF	-19.6	20.5		
[7]	Nama	Namibia	SAF	-22.6	17.1	Basque	Juhoansi
[7]	Karretjie	SouthAfrica	SAF	-30.8	25.1	Russian	Juhoansi
[7]	ColouredWellington	SouthAfrica	SAF	-33.7	19.0	Russian	Juhoansi
[3]	HADZA	Tanzania	EASTAF	-3.4	33.7		
[3]	SANDAWE	Tanzania	EASTAF	-6.2	35.7	Juhoansi	Sardinian
[3]	Khomani	SouthAfrica	SAF	-27.0	20.8	Belorussians	Juhoansi
[4]	LWK	Kenya	EASTAF	0.618	34.8	BiakaPygmy	Sardinian
[4]	YRI	Nigeria	WAF	8	5		
[4]	MKK	Kenya	EASTAF	-0.321	37.8	MbutiPygmy	Sardinian
[4]	GIH	India	CSASIA	27	72	Paniya	Georgians
[4]	CEU	CEPH	NEUROPE	55	-3	Karitiana	Sardinian
[4]	TSI	Italy	SEUROPE	43	11	Karitiana	Sardinian
[4]	CHB	China	EASTASIA	32.500	114	Dai	Daur
[4]	JPT	Japan	NEASTASIA	38	138		
[5]	Mozabite	Algeria-Mzab	MIDDLEEAST	32	3	YRI	Sardinian
[5]	Druze	Israel-Carmel	MIDDLEEAST	32	35		
[5]	Palestinian	Israel-Central	MIDDLEEAST	32	37	Yoruba	Sardinian
[5]	Bedouin	Israel-Negev	MIDDLEEAST	29	35	Sardinian	SUDANESE
[5]	Sindhi	Pakistan	CSASIA	25.5	69	Paniya	Georgians
[5]	Uygur	China	CSASIA	44	81	Italian	Japanese
[5]	Yoruba	Nigeria	WAF	8	5		
[5]	Mandenka	Senegal	WAF	12	-12		
[5]	BantuKenya	Kenya	EASTAF	-3	37	BiakaPygmy	Samaritians
[5]	BantuSouthAfrica	SouthAfrica	SAF	-23	20.7	YRI	Juhoansi

[5]	Xibo	China	CSASIA	43.5	83.5	Lithuanians	Japanese
[5]	Balochi	Pakistan	CSASIA	30.5	69.5	Cypriots	Paniya
[5]	BiakaPygmy	CAR	CAF	4	17		
[5]	MbutiPygmy	Congo	CAF	1	29		
[5]	Brahui	Pakistan	CSASIA	30.5	66.5	Cypriots	Paniya
[5]	Burusho	Pakistan	CSASIA	36.5	74	Georgians	Naxi
[5]	Hazara	Pakistan	CSASIA	33.5	73	Italian	Japanese
[5]	Kalash	Pakistan	CSASIA	36	71.5		
[5]	Makrani	Pakistan	CSASIA	26	64	Legzins	BantuSouthAfrica
[5]	Pathan	Pakistan	CSASIA	33.5	70.5	Samaritians	Karitiana
[5]	French	France	NEUROPE	46	2	Karitiana	Sardinian
[5]	Basque	France	SEUROPE	43	0		
[5]	Italian	Italy	SEUROPE	46	10	Karitiana	Sardinian
[5]	Sardinian	Italy	SEUROPE	40	9		
[5]	Orcadian	Orkney	NEUROPE	59	-3	Karitiana	Sardinian
[5]	Russian	Russia	NEUROPE	61	40	Karitiana	Sardinian
[5]	Adygei	Russia-Caucasus	NEUROPE	44	39	Karitiana	Sardinian
[5]	Cambodian	Cambodia	EASTASIA	12	105	Dai	Samaritians
[5]	Dai	China	EASTASIA	21	100		
[5]	Daur	China	NEASTASIA	46.500	124	Han	Yakut
[5]	Han	China	EASTASIA	32.500	114	Dai	Daur
[5]	Hezhen	China	NEASTASIA	47.500	136.5	She	Yakut
[5]	Lahu	China	EASTASIA	22	102		
[5]	Miaozu	China	EASTASIA	27	109		
[5]	Mongola	China	NEASTASIA	48.5	119	Lithuanians	Japanese
[5]	Naxi	China	EASTASIA	26	100		
[5]	Oroqen	China	NEASTASIA	50.5	126.500	Han	Yakut
[5]	She	China	EASTASIA	27	119		
[5]	Tu	China	EASTASIA	36	101	CEU	Tujia
[5]	Tujia	China	EASTASIA	29	109	Dai	Hezhen
[5]	Yizu	China	EASTASIA	28	103		
[5]	Japanese	Japan	EASTASIA	38	138		
[5]	Yakut	Siberia	NEASTASIA	70	129.500		
[5]	Melanesian	Bougainville	OCEANIA	-6	155		
[5]	Papuan	NewGuinea	OCEANIA	-4	143		
[5]	Karitiana	Brazil	AMERICA	-10	-63		
[5]	Surui	Brazil	AMERICA	-11	-62		
[5]	Colombian	Colombia	AMERICA	3	-68		
[5]	Maya	Mexico	AMERICA	19	-91	Moroccans	Surui
[5]	Pima	Mexico	AMERICA	29	-108		
[1]	Armenians	Armenia	CAUCASUS	40.1	45.0	GIH	Sardinian
[1]	Belorussians	Belarus	NEUROPE	53.7	28.0	Karitiana	Sardinian
[1]	Chuvaths	Russia	NEUROPE	55.6	46.930	Oroqen	Lithuanians
[1]	Cypriots	Cyprus	SEUROPE	35.1	33.4	Sardinian	ANUAK
[1]	Egyptans	Egypt	NAFRICA	26.8	30.8	Yoruba	Sardinian
[1]	Georgians	Georgia	CAUCASUS	42.3	43.3		
[1]	Hungarians	Hungary	NEUROPE	47.2	19.5	Karitiana	Sardinian
[1]	Iranians	Iran	MIDDLEEAST	32.4	53.7	Samaritians	Colombian
[1]	Jordanians	Jordan	MIDDLEEAST	29.6	38.2	Sardinian	SUDANESE

[1]	Legzins	Russia	CAUCASUS	42.1	47.1	Samaritians	Colombian
[1]	Lithuanians	Lithuania	NEUROPE	55.2	23.9		
[1]	Moroccans	Morocco	NAFRICA	31.8	-7.1	Yoruba	Sardinian
[1]	North_Kannadi	India	CSASIA	10.9	76.3		
[1]	Paniya	India	CSASIA	13.710	76.1		
[1]	Romanians	Romania	NEUROPE	46.0	25.0	Karitiana	Sardinian
[1]	Sakilli	India	CSASIA	11.130	78.7		
[1]	Samaritians	Israel	MIDDLEEAST	35.1	35.8		
[1]	Saudis	SaudiArabia	MIDDLEEAST	23.9	45.1	GuiGana	Sardinian
[1]	Spaniards	Spain	SEUROPE	40.5	-3.8	Karitiana	Sardinian
[1]	Syrians	Syria	MIDDLEEAST	34.8	39	MbutiPygmy	Sardinian
[1]	Turks	Turkey	MIDDLEEAST	39.0	35.2	Karitiana	Sardinian
[1]	Uzbeks	Uzbekistan	CAUCASUS	41.4	64.6	Oroqen	Sardinian
[1]	Yemenese	Yemen	MIDDLEEAST	15.6	48.5	MbutiPygmy	Sardinian

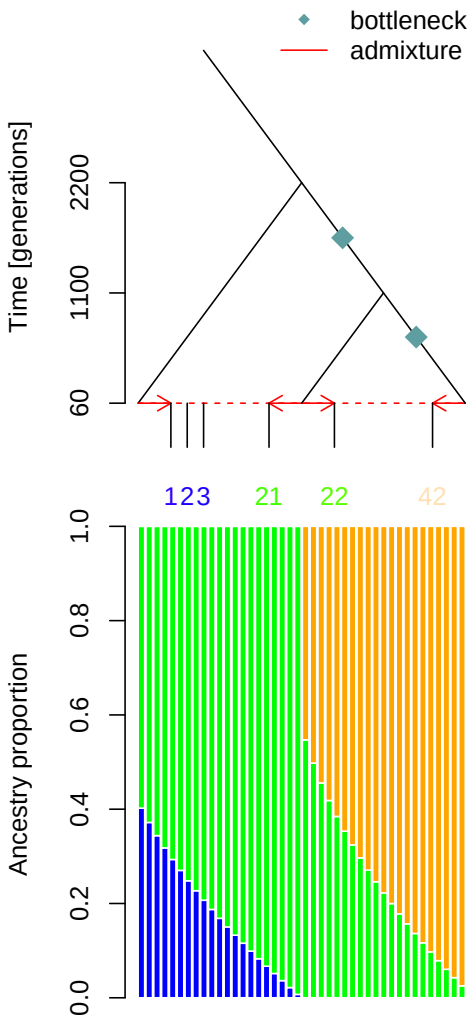


Figure 1. **A more detailed schematic of the model in Figure 1B.** In the top panel we show a schematic representation of the model. All ancestral population sizes are 10,000 individuals, and each bottleneck represents a reduction to a population size of 25 individuals for 10 generations. In the simulations we placed the bottlenecks immediately after the population splits rather than in the middle of the branches. In the lower panel we show the admixture proportions of the 42 sampled populations. Blue, green and orange represent the admixture proportions from the three ancestral populations pictured in the top panel from left to right, respectively.

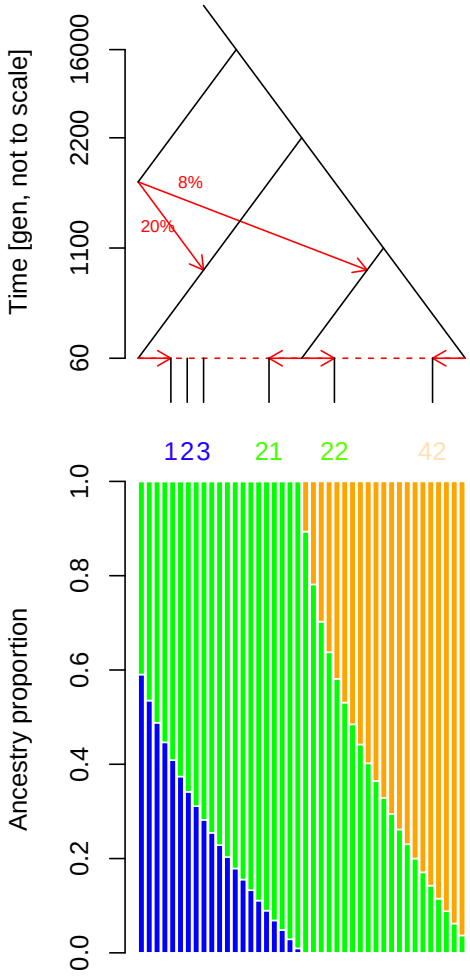


Figure 2. **A more detailed schematic of the model in Figure 1C.** In the top panel we show a schematic representation of the model. All ancestral population sizes are 10,000 individuals. Note that the split time of the archaic population is not to scale. The time of the archaic admixture is 800 generations in the past. In the lower panel we show the admixture proportions of the 42 sampled populations. Blue, green and orange represent the admixture proportions from the three ancestral populations pictured in the top panel from left to right, respectively.

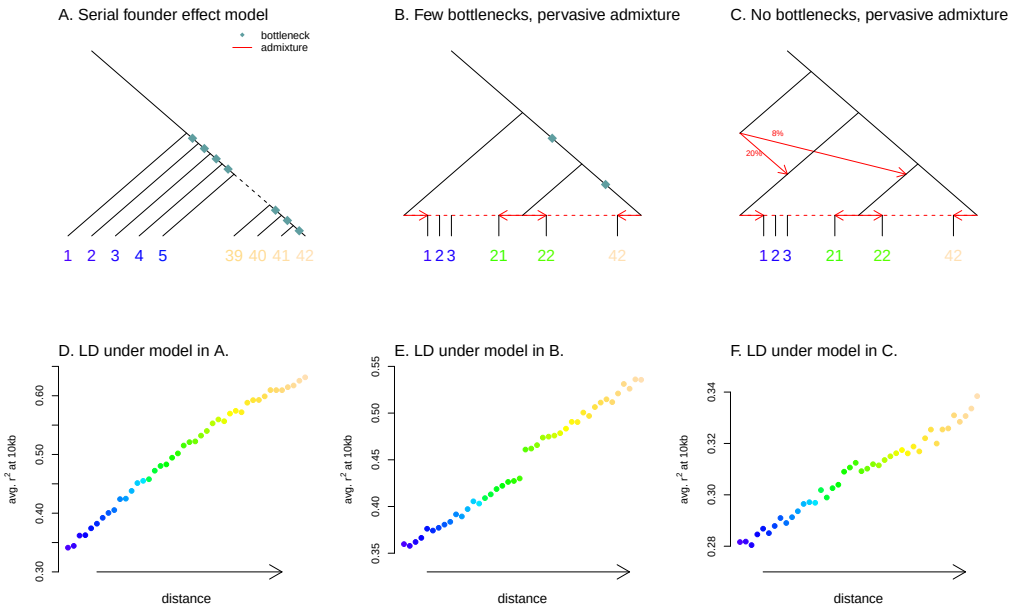


Figure 3. **A positive correlation between LD and geographic distance from a source population can be generated by qualitatively different, historically plausible demographic models**. We simulated genetic data under different demographic models and calculated the average level of LD (measured by r^2 between SNPs separated by 10-11kb) in each simulated population. **A.** Schematic of a serial founder effect model. **B.** Schematic of a demographic model with two bottlenecks and extensive admixture. **C.** Schematic of a demographic model with no bottlenecks and extensive admixture. **D,E,F.** Average LD in each population simulated under the demographic models in A,B, and C respectively. Each point represents a population, ordered along the x-axis according as in A.

References

[1] Behar, D. M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., *et al.*, 2010. The genome-wide structure of the Jewish people. *Nature*, **466**(7303):238–42.

[2] DeGiorgio, M., Jakobsson, M., and Rosenberg, N. A., 2009. Out of Africa: modern human origins special feature: explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. *Proc Natl Acad Sci U S A*, **106**(38):16057–62.

[3] Henn, B. M., Gignoux, C. R., Jobin, M., Granka, J. M., Macpherson, J. M., Kidd, J. M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A., *et al.*, 2011. Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc Natl Acad Sci U S A*, **108**(13):5154–62.

[4] International HapMap 3 Consortium, Altshuler, D. M., Gibbs, R. A., Peltonen, L., Altshuler, D. M., Gibbs, R. A., Peltonen, L., Dermitzakis, E., Schaffner, S. F., Yu, F., *et al.*, 2010. Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**(7311):52–8.

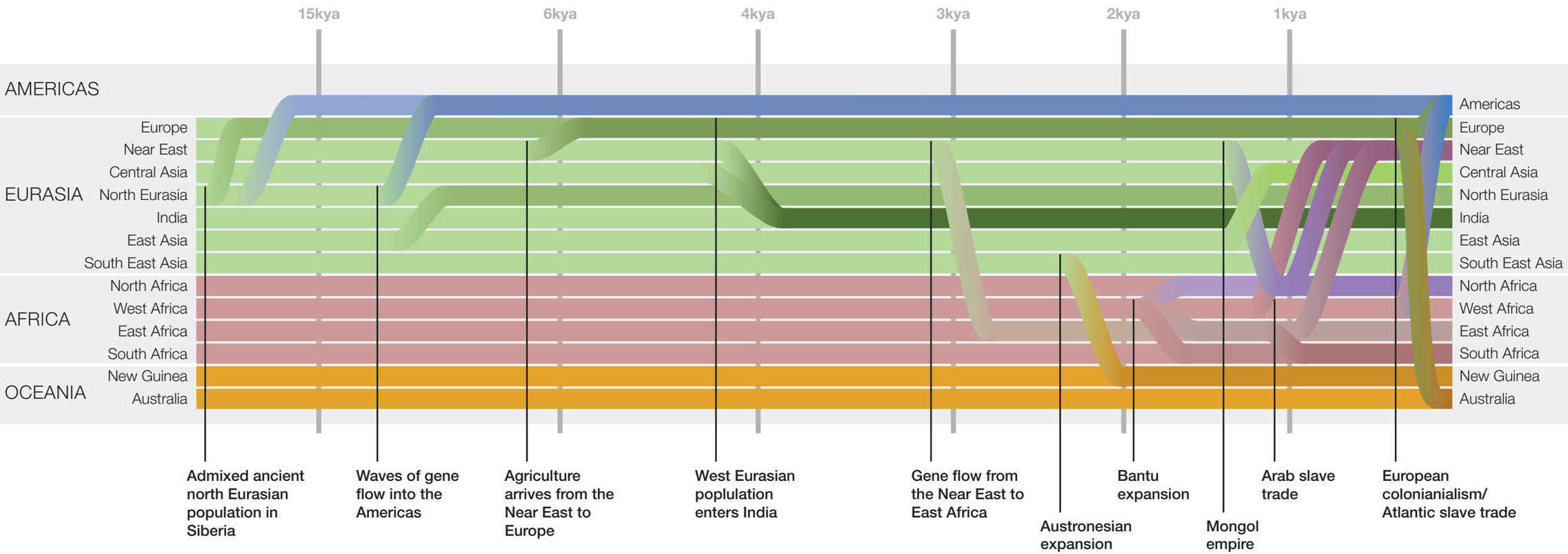
[5] Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., Cann, H. M., Barsh, G. S., Feldman, M., Cavalli-Sforza, L. L., *et al.*, 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, **319**(5866):1100–1104.

[6] Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S. Q., Thomas, M. G., Luiselli, D., *et al.*, 2012. Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am J Hum Genet*, **91**(1):83–96.

[7] Schlebusch, C. M., Skoglund, P., Sjödin, P., Gattepaille, L. M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A., Blum, M. G. B., *et al.*, 2012. Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science*, **338**(6105):374–9.

Trends in Genetics

Pickrell and Reich review recent progress in understanding human migration and admixture, arguing that present-day inhabitants of many geographic locations no longer resemble the original populations that settled these areas, which has implications for assessing human ancestry. Shown here are some of the human migration patterns from the past 20,000 years. Future studies relying on ancient DNA will continue to refine our understanding of human history.



Pickrell & Reich, Trends in Genetics, 2014.