# Supplementary Material for: Towards 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

*To whom correspondence should be addressed: jkpickrell@nygenome.org (JKP), reich@genetics.med.harvard.edu (DR)

**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 -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 0.0132146341463 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
0.0239463414634 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 0.0346780487805 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
```

In the demographic model in Figure 1B, there are three ancestral populations, two of which experienced strong bottlenecks. All observed populations are mixtures (in different proportions) of two of these populations. To simulate this scenario, 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 0, 22, and 44 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 -em 0.001475 2 1 38181.8181818 -ej 0.0015 2
```

```
23 -em 0.001475 3 1 36363.6363636 -ej 0.0015 3 23 -em 0.001475 4 1
34545.4545455 -ej 0.0015 4 23 -em 0.001475 5 1 32727.2727273 -ej 0.0015
5 23 -em 0.001475 6 1 30909.0909091 -ej 0.0015 6 23 -em 0.001475 7 1
29090.9090909 -ej 0.0015 7 23 -em 0.001475 8 1 27272.7272727 -ej 0.0015
8 23 -em 0.001475 9 1 25454.5454545 -ej 0.0015 9 23 -em 0.001475 10 1
23636.3636364 -ej 0.0015 10 23 -em 0.001475 11 1 21818.1818182 -ej
0.0015 11 23 -em 0.001475 12 1 20000.0 -ej 0.0015 12 23 -em 0.001475 13
1 18181.8181818 -ej 0.0015 13 23 -em 0.001475 14 1 16363.6363636 -ej
0.0015 14 23 -em 0.001475 15 1 14545.4545455 -ej 0.0015 15 23 -em
0.001475 16 1 12727.2727273 -ej 0.0015 16 23 -em 0.001475 17 1
10909.0909091 -ej 0.0015 17 23 -em 0.001475 18 1 9090.90909091 -ej
0.0015 18 23 -em 0.001475 19 1 7272.72727273 -ej 0.0015 19 23 -em
0.001475 20 1 5454.54545455 -ej 0.0015 20 23 -em 0.001475 21 1
3636.36363636 -ej 0.0015 21 23 -em 0.001475 22 1 1818.18181818 -ej
0.0015 22 23 -em 0.001475 24 23 38181.8181818 -ej 0.0015 24 45 -em
0.001475 25 23 36363.6363636 -ej 0.0015 25 45 -em 0.001475 26 23
34545.4545455 -ej 0.0015 26 45 -em 0.001475 27 23 32727.2727273 -ej
0.0015 27 45 -em 0.001475 28 23 30909.0909091 -ej 0.0015 28 45 -em
0.001475 29 23 29090.9090909 -ej 0.0015 29 45 -em 0.001475 30 23
27272.7272727 -ej 0.0015 30 45 -em 0.001475 31 23 25454.5454545 -ej
0.0015 31 45 -em 0.001475 32 23 23636.3636364 -ej 0.0015 32 45 -em
0.001475 33 23 21818.1818182 -ej 0.0015 33 45 -em 0.001475 34 23
20000.0 -ej 0.0015 34 45 -em 0.001475 35 23 18181.8181818 -ej 0.0015 35
45 -em 0.001475 36 23 16363.6363636 -ej 0.0015 36 45 -em 0.001475 37 23
14545.4545455 -ej 0.0015 37 45 -em 0.001475 38 23 12727.2727273 -ej
0.0015 38 45 -em 0.001475 39 23 10909.0909091 -ej 0.0015 39 45 -em
0.001475 40 23 9090.90909091 -ej 0.0015 40 45 -em 0.001475 41 23
7272.72727273 -ej 0.0015 41 45 -em 0.001475 42 23 5454.54545455 -ej
0.0015 42 45 -em 0.001475 43 23 3636.36363636 -ej 0.0015 43 45 -em
0.001475 44 23 1818.18181818 -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
```

In the demographic model in Figure 1C, there are three ancestral populations, two of which have admixed with an archaic population in different proportions. All observed populations are mixtures (in different proportions) of two of these populations. To simulate this scenario, 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 0, 22, 44 and 45 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 -em 0.001475 2 1 38181.8181818 -ej 0.0015
2 23 -em 0.001475 3 1 36363.6363636 -ej 0.0015 3 23 -em 0.001475 4 1
34545.4545455 -ej 0.0015 4 23 -em 0.001475 5 1 32727.2727273 -ej 0.0015
5 23 -em 0.001475 6 1 30909.0909091 -ej 0.0015 6 23 -em 0.001475 7 1
29090.9090909 -ej 0.0015 7 23 -em 0.001475 8 1 27272.7272727 -ej 0.0015
8 23 -em 0.001475 9 1 25454.5454545 -ej 0.0015 9 23 -em 0.001475 10 1
23636.3636364 -ej 0.0015 10 23 -em 0.001475 11 1 21818.1818182 -ej
0.0015 11 23 -em 0.001475 12 1 20000.0 -ej 0.0015 12 23 -em 0.001475 13
1 18181.8181818 -ej 0.0015 13 23 -em 0.001475 14 1 16363.6363636 -ej
0.0015 14 23 -em 0.001475 15 1 14545.4545455 -ej 0.0015 15 23 -em
0.001475 16 1 12727.2727273 -ej 0.0015 16 23 -em 0.001475 17 1
10909.0909091 -ej 0.0015 17 23 -em 0.001475 18 1 9090.90909091 -ej
```

```
0.0015 18 23 -em 0.001475 19 1 7272.72727273 -ej 0.0015 19 23 -em
0.001475 20 1 5454.54545455 -ej 0.0015 20 23 -em 0.001475 21 1
3636.36363636 -ej 0.0015 21 23 -em 0.001475 22 1 1818.18181818 -ej
0.0015 22 23 -em 0.001475 24 23 38181.8181818 -ej 0.0015 24 45 -em
0.001475 25 23 36363.6363636 -ej 0.0015 25 45 -em 0.001475 26 23
34545.4545455 -ej 0.0015 26 45 -em 0.001475 27 23 32727.2727273 -ej
0.0015 27 45 -em 0.001475 28 23 30909.0909091 -ej 0.0015 28 45 -em
0.001475 29 23 29090.9090909 -ej 0.0015 29 45 -em 0.001475 30 23
27272.7272727 -ej 0.0015 30 45 -em 0.001475 31 23 25454.5454545 -ej
0.0015 31 45 -em 0.001475 32 23 23636.3636364 -ej 0.0015 32 45 -em
0.001475 33 23 21818.1818182 -ej 0.0015 33 45 -em 0.001475 34 23
20000.0 -ej 0.0015 34 45 -em 0.001475 35 23 18181.8181818 -ej 0.0015 35
45 -em 0.001475 36 23 16363.6363636 -ej 0.0015 36 45 -em 0.001475 37 23
14545.4545455 -ej 0.0015 37 45 -em 0.001475 38 23 12727.2727273 -ej
0.0015 38 45 -em 0.001475 39 23 10909.0909091 -ej 0.0015 39 45 -em
0.001475 40 23 9090.90909091 -ej 0.0015 40 45 -em 0.001475 41 23
7272.72727273 -ej 0.0015 41 45 -em 0.001475 42 23 5454.54545455 -ej
0.0015 42 45 -em 0.001475 43 23 3636.36363636 -ej 0.0015 43 45 -em
0.001475 44 23 1818.18181818 -ej 0.0015 44 45 -em 0.02 23 46 4000 -em
0.020025 23 46 0  -em 0.02 1 46 20000 -em 0.020025 1 46 0 -ej 0.0275 45
23 -ej 0.055 23 1 -ej 1 46 1
```

**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 used in tests for admixture**

| Panel | Population | Country | Region | Lat | Long | P1 | P2 |
|---|---|---|---|---|---|---|---|
| Pagani | AFAR | Ethiopia | EASTAF | 12 | 41 | Sardinian | SUDANESE |
| Pagani | AMHARA | Ethiopia | EASTAF | 10 | 39 | Sardinian | SUDANESE |
| Pagani | ANUAK | Ethiopia | EASTAF | 8 | 34 | ARIBLACKSMITH | SUDANESE |
| Pagani | ARIBLACKSMITH | Ethiopia | EASTAF | 6 | 37 | NA | NA |
| Pagani | ARICULTIVATOR | Ethiopia | EASTAF | 6 | 39 | Sardinian | Juhoansi |
| Pagani | ESOMALI | Ethiopia | EASTAF | 9 | 42 | Sardinian | SUDANESE |
| Pagani | GUMUZ | Ethiopia | EASTAF | 10.78 | 35.57 | NA | NA |
| Pagani | OROMO | Ethiopia | EASTAF | 8 | 37 | Sardinian | SUDANESE |
| Pagani | SOMALI | Ethiopia | EASTAF | 5.15 | 46.2 | Sardinian | SUDANESE |
| Pagani | SUDANESE | Sudan | EASTAF | 12.86 | 30.21 | NA | NA |
| Pagani | TYGRAY | Ethiopia | EASTAF | 9 | 38 | Sardinian | SUDANESE |
| Pagani | WOLAYTA | Ethiopia | EASTAF | 6 | 37 | MbutiPygmy | Sardinian |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Henn | Khomani | SouthAfrica | SAF | -26.97 | 20.79 | Juhoansi | Belorussians |
| KSP | Khwe | Angola | SAF | -17.4 | 22.95 | Mozabite | Juhoansi |
| KSP | Xun | Angola | SAF | -14.6 | 17.66 | Yoruba | Juhoansi |
| KSP | GuiGana | Botswana | SAF | -23.6 | 24.66 | Yoruba | Juhoansi |
| KSP | Juhoansi | Namibia | SAF | -19.59 | 20.49 | NA | NA |
| KSP | Nama | Namibia | SAF | -22.55 | 17.07 | Basque | Juhoansi |
| KSP | Karretjie | SouthAfrica | SAF | -30.71 | 25.10 | Russian | Juhoansi |
| Hgdp | BiakaPygmy | CAR | CAF | 4 | 17 | NA | NA |
| Hgdp | MbutiPygmy | Congo | CAF | 1 | 29 | NA | NA |
| Henn | HADZA | Tanzania | EASTAF | -3.38 | 36.68 | NA | NA |
| Henn | SANDAWE | Tanzania | EASTAF | -6.18 | 35.74 | Sardinian | Juhoansi |
| HapMap | LWK | Kenya | EASTAF | 0.61 | 34.76 | BiakaPygmy | Sardinian |
| Hgdp | BantuKenya | Kenya | EASTAF | -3 | 37 | BiakaPygmy | Samaritians |
| Hgdp | BantuSouthAfrica | SouthAfrica | SAF | -21 | 18.7 | YRI | Juhoansi |
| HapMap | YRI | Nigeria | WAF | 8 | 5 | NA | NA |
| Hgdp | Yoruba | Nigeria | WAF | 8 | 5 | NA | NA |
| Hgdp | Mandenka | Senegal | WAF | 12 | -12 | NA | NA |
| HapMap | MKK | Kenya | EASTAF | -1.32 | 36.82 | MbutiPygmy | Sardinian |
| KSP | ColouredWellington | SouthAfrica | SAF | -33.6 | 19.01 | Russian | Juhoansi |
| Hgdp | Mozabite | Algeria-Mzab | MIDDLEEAST | 32 | 3 | YRI | Sardinian |
| Hgdp | Druze | Israel-Carmel | MIDDLEEAST | 32 | 35 | NA | NA |
| Hgdp | Palestinian | Israel-Central | MIDDLEEAST | 32 | 35 | Yoruba | Sardinian |
| Hgdp | Bedouin | Israel-Negev | MIDDLEEAST | 31 | 35 | Sardinian | SUDANESE |
| HapMap | GIH | India | CSASIA | 23 | 72 | Paniya | Georgians |
| Hgdp | Uygur | China | CSASIA | 44 | 81 | Italian | Japanese |
| Hgdp | Xibo | China | CSASIA | 43.5 | 81.5 | Lithuanians | Japanese |
| Hgdp | Balochi | Pakistan | CSASIA | 30.5 | 66.5 | Cypriots | Paniya |
| Hgdp | Brahui | Pakistan | CSASIA | 30.5 | 66.5 | Cypriots | Paniya |
| Hgdp | Burusho | Pakistan | CSASIA | 36.5 | 74 | Georgians | Naxi |
| Hgdp | Hazara | Pakistan | CSASIA | 33.5 | 70 | Italian | Japanese |
| Hgdp | Kalash | Pakistan | CSASIA | 36 | 71.5 | NA | NA |
| Hgdp | Makrani | Pakistan | CSASIA | 26 | 64 | Legzins | BantuSouthAfrica |
| Hgdp | Pathan | Pakistan | CSASIA | 33.5 | 70.5 | Samaritians | Karitiana |
| Hgdp | Sindhi | Pakistan | CSASIA | 25.5 | 69 | Paniya | Georgians |
| HapMap | CEU | CEPH | NEUROPE | 55 | -3 | Karitiana | Sardinian |
| HapMap | TSI | Italy | SEUROPE | 43 | 11 | Karitiana | Sardinian |
| Hgdp | French | France | NEUROPE | 46 | 2 | Karitiana | Sardinian |
| Hgdp | Basque | France | SEUROPE | 43 | 0 | NA | NA |
| Hgdp | Italian | Italy | SEUROPE | 46 | 10 | Karitiana | Sardinian |
| Hgdp | Sardinian | Italy | SEUROPE | 40 | 9 | NA | NA |
| Hgdp | Orcadian | Orkney | NEUROPE | 59 | -3 | Karitiana | Sardinian |
| Hgdp | Russian | Russia | NEUROPE | 61 | 40 | Karitiana | Sardinian |
| Hgdp | Adygei | Russia-Caucasus | NEUROPE | 44 | 39 | Karitiana | Sardinian |
| HapMap | CHB | China | EASTASIA | 32.5 | 114 | Dai | Daur |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HapMap | JPT | Japan | NEASTASIA | 38 | 138 | NA | NA |
| Hgdp | Cambodian | Cambodia | EASTASIA | 12 | 105 | Dai | Samaritians |
| Hgdp | Dai | China | EASTASIA | 21 | 100 | NA | NA |
| Hgdp | Daur | China | NEASTASIA | 48.5 | 124 | Han | Yakut |
| Hgdp | Han | China | EASTASIA | 32.5 | 114 | Dai | Daur |
| Hgdp | Hezhen | China | NEASTASIA | 47.5 | 133.5 | She | Yakut |
| Hgdp | Lahu | China | EASTASIA | 22 | 100 | NA | NA |
| Hgdp | Miaozu | China | EASTASIA | 28 | 109 | NA | NA |
| Hgdp | Mongola | China | NEASTASIA | 48.5 | 119 | Lithuanians | Japanese |
| Hgdp | Naxi | China | EASTASIA | 26 | 100 | NA | NA |
| Hgdp | Oroqen | China | NEASTASIA | 50.5 | 126.5 | Han | Yakut |
| Hgdp | She | China | EASTASIA | 27 | 119 | NA | NA |
| Hgdp | Tu | China | EASTASIA | 36 | 101 | CEU | Tujia |
| Hgdp | Tujia | China | EASTASIA | 29 | 109 | Dai | Hezhen |
| Hgdp | Yizu | China | EASTASIA | 28 | 103 | NA | NA |
| Hgdp | Japanese | Japan | EASTASIA | 38 | 138 | NA | NA |
| Hgdp | Yakut | Siberia | NEASTASIA | 70 | 129.5 | NA | NA |
| Hgdp | Melanesian | Bougainville | OCEANIA | -6 | 155 | NA | NA |
| Hgdp | Papuan | NewGuinea | OCEANIA | -4 | 143 | NA | NA |
| Hgdp | Karitiana | Brazil | AMERICA | -10 | -63 | NA | NA |
| Hgdp | Surui | Brazil | AMERICA | -11 | -62 | NA | NA |
| Hgdp | Colombian | Colombia | AMERICA | 3 | -68 | NA | NA |
| Hgdp | Maya | Mexico | AMERICA | 19 | -91 | Moroccans | Surui |
| Hgdp | Pima | Mexico | AMERICA | 29 | -108 | NA | NA |
| Behar | Armenians | Armenia | CAUCASUS | 40.07 | 45.04 | GIH | Sardinian |
| Behar | Belorussians | Belarus | NEUROPE | 53.71 | 27.95 | Karitiana | Sardinian |
| Behar | Chuvaths | Russia | NEUROPE | 55.56 | 46.93 | Oroqen | Lithuanians |
| Behar | Cypriots | Cyprus | SEUROPE | 35.13 | 33.43 | Sardinian | ANUAK |
| Behar | Egyptans | Egypt | NAFRICA | 26.82 | 30.8 | Yoruba | Sardinian |
| Behar | Georgians | Georgia | CAUCASUS | 42.32 | 43.36 | NA | NA |
| Behar | Hungarians | Hungary | NEUROPE | 47.16 | 19.5 | Karitiana | Sardinian |
| Behar | Iranians | Iran | MIDDLEEAST | 32.43 | 53.69 | Samaritians | Colombian |
| Behar | Jordanians | Jordan | MIDDLEEAST | 30.59 | 36.24 | Sardinian | SUDANESE |
| Behar | Legzins | Russia | CAUCASUS | 42.14 | 47.09 | Samaritians | Colombian |
| Behar | Lithuanians | Lithuania | NEUROPE | 55.17 | 23.88 | NA | NA |
| Behar | Moroccans | Morocco | NAFRICA | 31.79 | -7.09 | Yoruba | Sardinian |
| Behar | North_Kannadi | India | CSASIA | 10.85 | 76.27 | NA | NA |
| Behar | Paniya | India | CSASIA | 11.71 | 76.1 | NA | NA |
| Behar | Romanians | Romania | NEUROPE | 45.94 | 24.97 | Karitiana | Sardinian |
| Behar | Sakilli | India | CSASIA | 11.13 | 78.66 | NA | NA |
| Behar | Samaritians | Israel | MIDDLEEAST | 32.07 | 34.78 | NA | NA |
| Behar | Saudis | SaudiArabia | MIDDLEEAST | 23.89 | 45.08 | GuiGana | Sardinian |
| Behar | Spaniards | Spain | SEUROPE | 40.46 | -3.75 | Karitiana | Sardinian |
| Behar | Syrians | Syria | MIDDLEEAST | 34.8 | 39 | MbutiPygmy | Sardinian |
| Behar | Turks | Turkey | MIDDLEEAST | 38.96 | 35.24 | Karitiana | Sardinian |
| Behar | Uzbeks | Uzbekistan | CAUCASUS | 41.38 | 64.59 | Oroqen | Sardinian |
| Behar | Yemenese | Yemen | MIDDLEEAST | 15.55 | 48.52 | MbutiPygmy | Sardinian |