Supporting Information

Hammer et al. 10.1073/pnas.1109300108

SI Materials and Methods

Two-Population Model. *Estimating demographic parameters.* For each pair of sub-Saharan African populations we consider the following demographic model: an ancestral panmictic population having effective population size, $N_e = 10^4$ splits at time T_1 into two descendant panmictic populations each also having $N_e = 10^4$. Given a per-generation migration rate m, these descendant populations exchange migrants at the scaled migration rate $M = 4\,N_e\,m$ until the present day. The two populations have 100-fold population growth starting at times g_1 and g_2, respectively (Fig. 1*A*, main text).

We use previously published composite-likelihood methodology (1, 2) to estimate parameters $\psi = (g_1, g_2, T_1, M)$. This method uses information from levels of diversity and the joint frequency spectrum—but not linkage disequilibrium (LD)—for estimating (composite) likelihoods. Likelihoods are calculated over a grid of parameter values, with increments of 2,000 y for g_1 and g_2, 5,000 y for T_1, and 1 for M. Scaled recombination rates were assumed to be fixed within loci but to vary across loci, within-locus recombination rates were chosen from a Γ distribution with mean equal to half of the average mutation rate as estimated by θ_W (3). We ran 5×10^5 simulations for each parameter combination.

First, we simulated 15 replicates under each of the following three scenarios: $\psi_1 = (0, 4, 450, 10)$, $\psi_2 = (0, 4, 35, 5)$, and $\psi_3 = (0, 4, 25, 0)$. The first scenario corresponds to the Mandenka-Biaka maximum-likelihood estimates from the data, whereas ψ_2 and ψ_3 are comparable parameter values (with the same values of g_1 and g_2) that produce roughly the same average value of F_{ST}. A summary of the simulation results is shown in Table S1. We note that the approximate 95% confidence intervals (CIs) (based on asymptotic likelihood assumptions) cover the true parameter value roughly 93% of the time (167 of 180), which suggests that CIs based on standard assumptions are reasonably accurate. For analyzing the actual data, we empirically determined a new likelihood-ratio cutoff value for estimating 95% CIs. This cutoff (which takes log-likelihood values within 2.8 of the maximum-likelihood estimate) has the correct coverage level for the simulated data.

Detecting archaic admixture. For each pair of African populations, we used the parameters estimated above as a null model and tested for the presence of additional ancient population structure (2). If archaic admixture occurs at a locus, then "archaic" SNPs on introgressed sequences would be in strong LD with each other. Simulations suggest that both the number of such SNPs and the total distance spanned by such SNPs are elevated when archaic admixture occurs (4). To exploit these two observations, and to account for the effects of intragenic recombination (1), we calculated, for each locus, a statistic, S^*, shown to be sensitive to archaic admixture (1). S^* looks for population-specific SNPs (excluding singletons) that are in strong LD with each other (e.g., the square correlation $r^2 \approx 1$). We determine the significance of S^* values from the actual data by running simulations using the previously estimated demographic parameters to obtain a distribution of S^* values under the null hypothesis of no (archaic) admixture. Significantly high S^* values are interpreted as departures from the null model in the direction of some unknown ancient population structure. We estimate P values for each locus by running 10^4 simulations under the null model. The P values across loci were combined (assuming independence) using the method of Fisher (5).

Three-Population Model. To more closely model our population sampling strategy, we introduce a second, more comprehensive three-population model (Fig. 1*B*, main text). Our goal is to estimate simultaneously the time of admixture (T_a), the ancestral split time (T_0), and the admixture proportion (a). Our approach has several modeling assumptions, including that the San are ancestral to the Mandenka and Biaka (6), that the migration rate between all three populations is symmetric and constant, that recent population growth leads to a 100-fold increase in effective population size, and that generation time is 25 y The model is specified by the parameters $\psi = (N_A, T_1, T_2, g_1, g_2, M, T_a, T_0, a)$, where

- N_A is the ancestral effective population size,
- T_1 is the time when the San split from the Biaka-Mandenka,
- T_2 is the time when the Biaka and Mandenka split,
- g_1 is the time since the start of population growth in the San,
- g_2 is the time since the start of population growth in the Biaka and Mandenka, and
- M is the scaled migration rate.

Summary Statistics. To identify candidate introgressed sequences, we adopt the following approach. For each locus, we cluster all sequences into two (putatively basal) groups, G_1 and G_2, as follows:

1. Identify the two most diverged sequences.
2. Assign the remaining sequences to one of two groups according to genetic similarity to the two individuals identified in step 1.
3. For a tie in step 2, calculate the average genetic distance between the target individual and all individuals in each group. Assign membership to the closer group. In case of a tie, assign group membership randomly.

Then, define the statistics:

- K_{max}, the number of differences between the sequences chosen in step 1,
- S_s, the number of polymorphisms shared between the two groups,
- S, the total number of polymorphisms, and
- d, the number of fixed differences between human and chimpanzee sequence.

We now define our summary statistics for inference $D_1 = S_s/S$, $D_2 = K_{\max}/d$, and $D_3 = \min\{|G_1|, |G_2|\}$.

Because our null model of no admixture, H_0, is a subspace of our alternative model of admixture, H_1, we can make inference using likelihood ratio tests. Further, we can use χ^2_3, the χ^2 statistic with three degrees of freedom, as a test statistic for the difference in log-likelihood values under H_0 and H_1. This is a conservative approximation, however, because the null space represents a corner of our alternative space. Unless otherwise stated, P values are those that come from this approximation.

We approximate the likelihood of the summary statistics $\mathbf{D} = (D_1, D_2, D_3)$ using tolerance levels $\delta = (\delta_1, \delta_2, \delta_3)$. Thus, for each set of model parameters ψ we estimate

$$\Pr_\psi\{|d_1 - D_1| < \delta_1, |d_2 - D_2| < \delta_2, |d_3 - D_3| < \delta_3\},$$

where d_1, d_2, and d_3 are calculated from data simulated under the parameter values ψ. The initial tolerances were selected to maximize power for 1% admixture. Loci are assumed to be in-

dependent, so likelihoods for the full data are computed as the product of the 61 locus-specific likelihoods.

To simulate the data to compute approximate likelihoods, we first need to determine fine-scale estimates of recombination. To this end, we used Phase 2.1 (7, 8) using two qualitatively different strategies. We examined genotypes for our loci in 30 HAPMAP (9) Yoruba parent–offspring trios. For each trio we replaced genotype calls at SNPs showing Mendelian inconsistencies with "missing data." Using parental genotypes, we constructed Phase files for each locus, adding an additional 10 kb of flanking data to mitigate any possible edge effects in our estimate of ρ. Using all individuals to estimate ρ, we used a 10,000-step burn-in and sampled 100,000 points in the posterior. We then estimated ρ for each locus according to the median per-SNP ρ estimates from Phase. To validate this approach we performed 100 coalescent simulations using *ms* (10) of a single Wright-Fisher population using a recombination rate of 1 cM/Mb over 40 kb of sequence, and then we ran Phase on these simulations. We contrasted computing the mean vs. the median recombination rate estimates, to the per-SNP and to the per-locus estimates, and found that the per-SNP median estimate better recovered, although slightly underestimated, the simulated value.

In the second strategy we used Phase to estimate ρ using the major clade of each locus in our own resequencing data using the same Phase parameters as above. This approach is, in general, less powerful because of the locus trio design. With only ≈6 kb of data collected over a ≈20-kb genomic window, the ability to infer the recombination rates will be hampered by the small number of segregating sites. In addition, we are restrained to the assumption that ρ at the locus trio is constant across a 20-kb region. We used the same validation approach as in the first strategy, modified to run Phase on the major clade of each simulation using an archetype locus-trio design, and again found that the median per-SNP estimate better recovered the simulated value. Despite numerous attempts, Phase failed to complete on locus *1pMB4*, and thus this locus was dropped from all subsequent analyses.

Rejecting the Null Hypothesis. We simulated ancestral recombination graphs (ARGs) over a grid of parameter values to estimate each locus's approximate likelihood using the recombination rate estimates described above and tolerances $\delta_1 = 0.06$, $\delta_2 = 0.05$, and $\delta_3 = 2$. Parameter values ranged from 6,000 to 16,000 for N_A, 60 to 120 kya for T_1, 30 to 60 kya for T_2, 20 to 40 kya for g_1 and g_2, 0 to 10 for M, 10 to 100 kya for T_a, 0.125 to 1.5 Mya for T_0, and 0 to 8% for a.

To provide a coarse-grained likelihood surface, we generated 5,000 ARGs over a reduced grid of the parameter space. We use a goodness of fit (GOF) test to identify loci for finer-scale estimation. This yielded three loci (*4qMB105*, *16pMB17*, and *13qMB64*) with poor fit GOF ($P < 0.05$) across our entire parameter space, with P values of 0.022, 0.015, and 0.026, respectively. These three loci were then rerun using the major clade fine-scale estimate of recombination, and all three exhibited improved GOF, with P values of 0.400, 0.145, and 0.526, respectively. Further, in the initial run, two additional loci, *13qMB107* and *18qMB73*, had fine-scale estimates of recombination that were exceedingly high (estimates for ρ per locus are 147.97 and 118.73, respectively), leading to coalescent runtimes that were prohibitively long. For all future simulations the major clade estimate was used for these loci. (Estimates for ρ per locus are 95.46 and 107.28, respectively.)

To obtain a more refined point estimate, we reduced the parameter space to the null space and to those values within the 99% CI of the coarse-grain estimates. We then ran simulations using 100,000 ARGs for each parameter value. In addition, we store, for each parameter value, an approximation of the summary statistic distribution in a 3D histogram (for our three summaries) using a reduced tolerances $\delta_1 = 0.01$, $\delta_2 = 0.01$, and $\delta_3 = 0$.

The result is a maximum-likelihood estimate of $T_a = 40$ kya, $T_0 = 750$ kya, and $a = 1\%$ with a log-likelihood ratio of −2.01. To estimate the significance of this value we drew 10,000 points from the maximum-likelihood location under H_0 using our 3D histogram and tabulated the probability of observing a log-likelihood ratio as small (or smaller) than −2.01 with an archaic split time no more recent than 750 kya. The bootstrapped P value for this is 0.0493, allowing us to reject the null hypothesis. Although this P value is only marginally significant, as seen in the sections that follow, more refined analyses yield even smaller P values under the conservative χ^2 approximation of the likelihood ratio test.

Describing H_1. We chose two different approaches to describing our alternative model. The first, and simplest, uses the minimum tolerances for each of the summary statistics, D_1, D_2, and D_3, keeping the other two at their original tolerances. This gave us three sets of three likelihood profiles (for each of the three admixture parameters). Minimizing the tolerance for δ_1 best restricted the parameter space. Under this method, the point estimates are: $T_0 = 375$ kya, $T_a = 20$ kya, and $a = 2\%$ with a log-likelihood ratio of −4.14 ($P = 0.04$). Moreover, this method allowed us to estimate the following 95% CIs for T_0, T_a, a: 125 kya $< T_0 <$ 1.5 Mya, $0 < T_a < 70$ kya, and $0 < a < 1$.

There was one exception to this analysis. The log-likelihood difference between the parameter value for $T_a = 100$ kya ($T_0 = 1$ Mya, $a = 0.5\%$) and the maximum is −1.915, marginally inside of our CIs based on the χ^2 approximation. To assess the accuracy of this approximation, we drew 10,000 samples from this point in the alternative space and estimated the probability of observing a maximum log-likelihood ratio at or more extreme than −1.915 at an introgression time at most 20 kya. The bootstrapped probability of this occurring by chance is 0.021, allowing us to place this single point in the alternative model outside of our 95% CI. As seen in Fig. 2 (main text) and Fig. S5, the alternative space can best be described as multimodal.

Custom tolerances. From our bootstrap analysis, we found that locus-specific critical values are largely determined by the basal recombination rate. More precisely, loci with higher recombination rates required much smaller likelihood ratios to reject H_0. To determine optimal tolerance values to discriminate between values in the parameter space, given a fixed number of ARGs, we chose at random 100 parameter values for each locus. For each pair of values, we evaluated tolerance levels from the minimal tolerance up to our original level of acceptable tolerance. We then asked the question: what is the level of tolerance that maximizes our discriminatory power given that 1% of the points will yield an observed likelihood of 0?

Applying these custom tolerances to our loci yielded pronounced evidence of two distinct maxima: $T_0 = 375$ kya, $T_a = 10$ kya, and $a = 0.5\%$ and $T_0 = 750$ kya, $T_a = 40$ kya, and $a = 2\%$, with essentially equal log-likelihood values of −468.48 and −468.67, respectively, and the former yielding a log-likelihood ratio of −5.02 ($P < 0.02$). Four loci (*1pMB101*, *12qMB46*, *5pMB35*, and *5qMB123*) had fewer than 10 ARGs that matched their empirical values in either of the maxima, and for these loci an additional 100,000 ARGs were generated, elevating the minimum number of matching simulations to 10 for all loci. This slightly adjusted the likelihood surface, favoring instead the older archaic split time as the maximum likelihood estimate (likelihoods of −468.78 and −468.51), giving a likelihood ratio of −5.00, $P < 0.02$ (Fig. 2, main text). The same strategy was used to elevate the minimum number of matching ARGs to 20, this time favoring the local maxima $T_0 = 500$ kya, $T_a = 20$ kya, and $a = 2\%$ and $T_0 = 750$ kya, $T_a = 40$ kya, and $a = 2\%$, with the first of the two points moving perhaps more than expected.

At least 20 matches. To test whether this movement was due to the sampling variance associated with estimating exceedingly small likelihoods, we designed an iterative variant to the above procedure designed ensure that all loci have at least 20 matching ARGs for each point within our 95% CI using the χ^2 approximation. To accommodate this, we used our initial set of custom tolerances, and rather than keeping the tolerances static and adding more simulations as needed, we instead relaxed tolerances for all loci having fewer than 20 matches and looked at the minimum number of matching ARGs over the 95% confidence region. The 2D likelihood surface shows two distinct maxima: $T_0 = 625$ kya, $T_a = 30$ kya, and $a = 3\%$ and $T_0 = 250$ kya, $T_a = 10$ kya, and $a = 5\%$, the latter of which has estimated times similar to those discovered in our two population approach.

Goodness of Fit. We used a parametric bootstrap to address GOF. In particular, we drew 1,000 samples from our 3D histogram for each locus for both maxima in H_1. We then estimated the likelihood of each of our 1,000 samples and calculated the probability of our empirical likelihood value in this distribution. This generated a probability value for each locus, and these probability values were combined using the method of Fisher (5) to give a single GOF P value for the data set. This procedure was run on the *at least 20 matches* maxima, yielding GOF P values of 0.059 and 0.071 for the earlier and later archaic split maxima, respectively. These P values are conservative, because any maxima we find will only be a maximum with respect to our parameter space discretization; finer discretization will likely result in higher maxima and, thus, in an improved fit of the model. Uncertainties in our recombination rate estimates also influence the fit. Notably, results that are based on the deCODE estimates of recombination, which are estimated over much larger physical distance, produced a substantially smaller GOF ($P < 10^{-4}$).

Likelihood Ratios of Individual Loci. Although this approach examines the likelihood of the set of 61 loci together, it also can be used to evaluate whether a particular locus better fits the alternative model. To identify individual loci that are likely to harbor archaic lineages, we allow all nine parameters to vary freely among loci. In addition, rather than selecting points from the 3D histogram, which are only defined for our initial estimate of the 99% CI for all loci together, we instead selected our maxima and calculated our bootstrapped P values from our original coarse scan of the parameter space. Table 1 (main text) describes the three loci exhibiting the lowest P value.

Describing Two Maxima. Throughout our attempts to describe the alternative space we have seen pronounced evidence for two peaks in our likelihood surface: one with more recent time characteristics (ψ_{recent}), with $T_0 \approx 375$ kya and $T_a \approx 15$ kya and the other at an older time (ψ_{old}), $T_0 \approx 700$ kya and $T_a \approx 35$ kya. This leads to the question: do some loci favor one maximum over the other, and if so, which ones? To address this we compute the likelihood ratio:

$$L(\psi_{old}|data)/L(\psi_{recent}|data)$$

for each locus (Fig. S2) using the approach guaranteeing at least 10 matching simulations for each locus. Notably, the three loci that individually favor H_1 (Table 1, main text) are among four most extreme likelihood ratios.

Genotyping Candidate Alleles. A sample of ≈500 individuals from 14 sub-Saharan African populations was genotyped at a single insertion and two SNPs that marked divergent alleles at the three loci exhibiting the lowest P value in the likelihood test described above. A 4-nt insertion (GCCA) at position 179598847 (hg18) within *4qMB179* was genotyped by using an allele-specific PCR. We obtained the DNA sequence of all samples containing the insertion to confirm heterozygosity. A G→A nucleotide polymorphism site at 107495053 (hg18) within *13qMB107* was genotyped via a PCR and subsequent restriction enzyme digestion (ApoI, NEB catalog no. R0566). An A→G nucleotide polymorphism site at site 60718922 (hg18) within *18qMB60* was genotyped via a PCR and subsequent restriction enzyme digestion (DdeI, NEB catalog no. R0175).

1. Plagnol V, Wall JD (2006) Possible ancestral structure in human populations. *PLoS Genet* 2:e105.
2. Wall JD, Lohmueller KE, Plagnol V (2009) Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Mol Biol Evol* 26: 1823–1827.
3. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276.
4. Wall JD (2000) Detecting ancient admixture in humans using sequence polymorphism data. *Genetics* 154:1271–1279.
5. Mosteller F, Fisher RA (1948) Questions and answers #14. *The American Statistician* 2:30–31.
6. Wall JD, et al. (2008) A novel DNA sequence database for analyzing human demographic history. *Genome Res* 18:1354–1361.
7. Crawford DC, et al. (2004) Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat Genet* 36:700–706.
8. Li N, Stephens M (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165: 2213–2233.
9. Frazer KA, et al.; International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861.
10. Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.

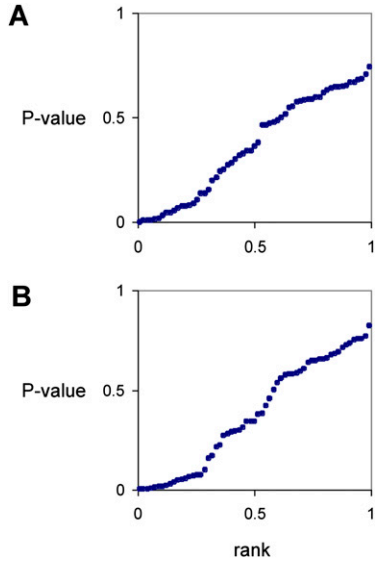

Fig. S1. Quantile–quantile plots for the *P* values for S*, calculated for each locus. Results are shown for the (*A*) Biaka and (*B*) San, in pairwise analyses with the Mandenka.

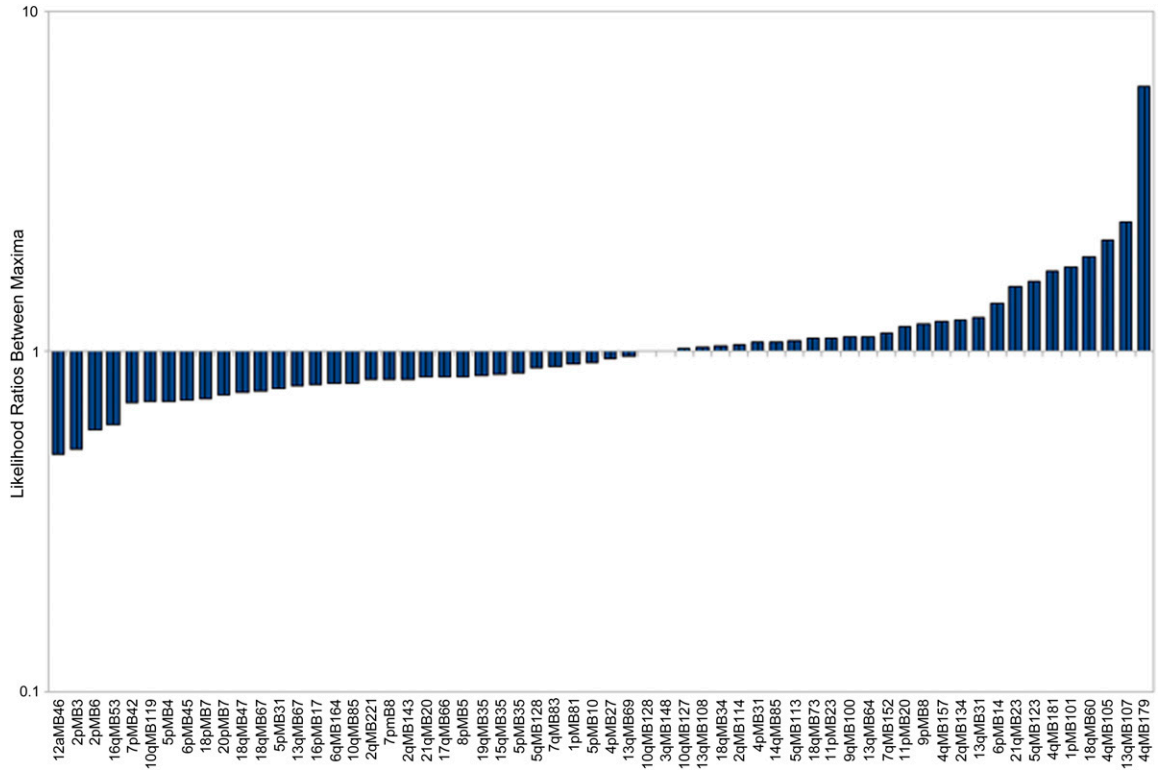

Fig. S2. The likelihood ratios, $L(\psi_{old}|data)/L(\psi_{recent}|data)$, of each locus in the distinct two maxima in H_1. The loci *18qMB60*, *13qMB107*, and *4qMB179* all favor H_{old}, and all three are loci, individually, favor the H_1.

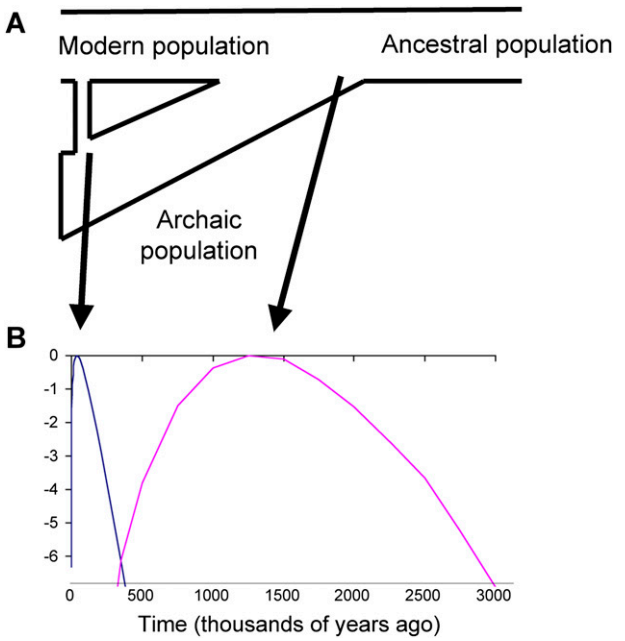

Fig. S3. Schematic of a simple isolation model (*A*) and *4qMB179* profile likelihood curves (*B*) for estimates of the split time T_0 and the admixture time T_a, in thousands of years ago.

A

Position (hg18)	Position in Garrigan et al. (12).	Coverage in Neandertal	State in divergent lineage	State in Neandertal
143,220,506	1,343	Yes	A (derived)	A (derived)
143,220,802	1,639	No	A (Indeterminant)	?
143,221,114	1,951	Yes	G (derived)	G (derived)
143,221,183	2,020	Yes	A (derived)	A (derived)
143,221,426	2,263	Yes	C/A	C (ancestral)
143,221,467	2,304	Yes	T (ancestral)	T (ancestral)

B

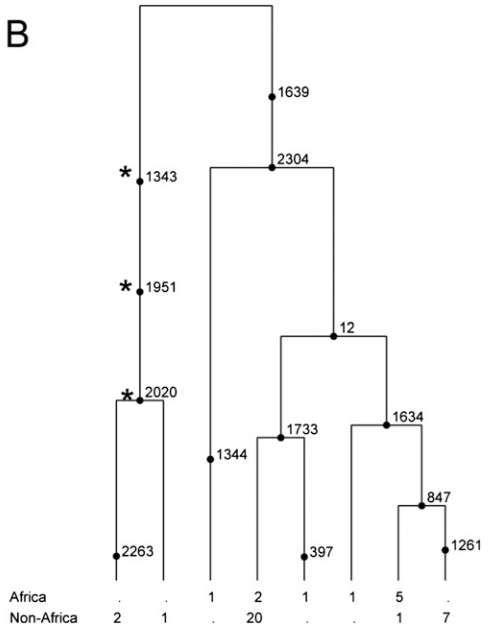

Fig. S4. (*A*) Sharing of SNPs between Neandertal and human divergent lineage at *RRM2P4*. (*B*) Asterisk indicates sharing of derived state in Neandertal and human divergent lineage on the phylogenetic tree shown in Garrigan et al. (1).

1. Garrigan D, Mobasher Z, Severson T, Wilder JA, Hammer MF (2005) Evidence for archaic Asian ancestry on the human X chromosome. *Mol Biol Evol* 22:189–192.

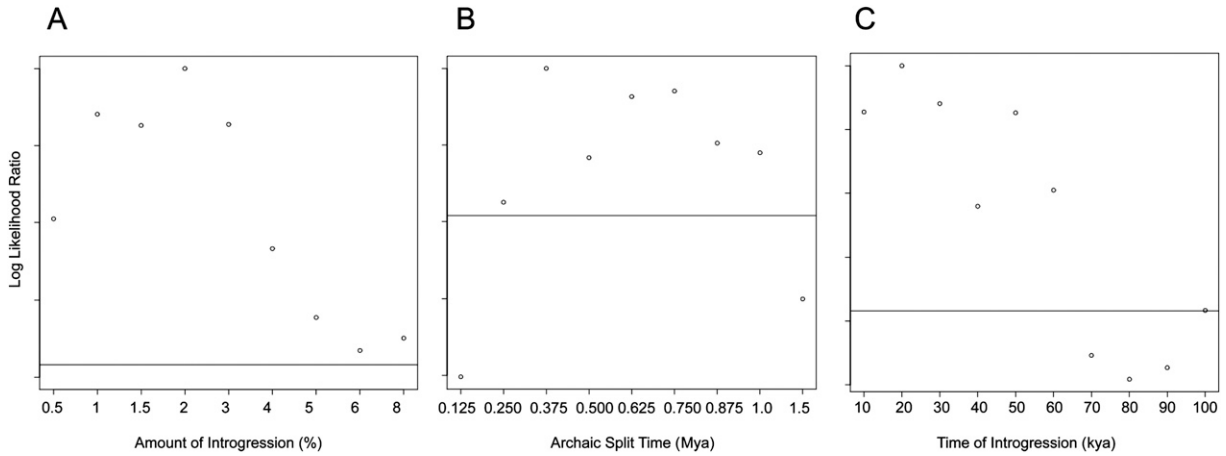

Fig. S5. Likelihood profiles for the archaic admixture parameters for (*A*) amount of admixture, (*B*) archaic split time, and (*C*) time of introgression. Horizontal line represents the 95% CI cutoff using the χ^2 approximation. The T_a = 100 kya point was shown to be outside of our confidence region using a parametric bootstrap.

Table S1. Point estimates (simulation-based 95% CI) for the actual data

Parameter	Man–Bia	Man–San	Bia–San
g_1 (kya)	0 (0–5.2)	0 (0–5.5)	10 (0–22)
g_2 (kya)	4 (0–11)	2 (0–11)	4 (0–20)
T (kya)	450 (280–690)	100 (64–500)	55 (40–230)
M	10 (8.2–12)	3 (1.6–4.2)	1.5 (0–5.3)

Table S2. Mean values of parameter estimates on simulated data (g_1 = 0, g_2 = 4 kya)

Model	T_1 = 25, M = 0	T_1 = 35, M = 5	T_1 = 450, M = 10
g_1	0.9	2.3	2.4
g_2	4.1	5.9	7.7
T_1	25	44	580
M	1.1	4.3	9.6
Coverage* *(%)*	97	88	93

*Coverage denotes the fraction of times that the estimated 95% CIs contained the true parameter value.

Table S3. Biaka 4qMB179 haplotypes

		179717925	179718025	179718164	179718283	179718420	179718491	179718492	179718621	179718977	179719030	179719130	179719260	179727869	179727870	179727922	179728023	179728221	179728254	179728314	179728525	179728711	179728796	179728857	179729161	179729221	179729231	179729281	179729444	179729451	179729541	179736287	179736363	179736378	179736411	179736422	179736437	179736564	179736762	179736898	179736963	179737002	179737179	179737276	179737345	179737389	179737426	179737436	179737452	179737532	179737646	179737685	179737719	179738071	179738125
Biaka Ht	n	A	C	G	T	G	C	A	C	C	A	T	A	C	A	A	A	G	C	A	C	A	C	A	A	T	G	C	T	A	A	G	G	A	G	C	C	C	T	C	T	—	G	C	G	C	A	C	A	G	C	G	C	C	T
B1	1	C	.	T	.	.	N	N	N	N	N	.	.	T	G	G	.	.	.	T	.	.	.	G	.	.	.	.	C	G	.	.	C	G	.	.	.	.	.	.	A	GCCA	T	A	.	T	.	T	.	.	T	.	.	.	C
B2	1	C	.	T	.	.	.	.	N	.	.	.	.	T	G	G	.	.	.	T	.	.	.	G	.	.	.	.	C	.	.	.	C	G	.	.	.	.	.	.	A	GCCA	T	A	.	.	.	T	.	.	T	.	.	A	C
B3	1	C	.	T	.	.	.	.	N	.	.	.	.	T	G	G	.	.	.	T	.	.	.	G	.	.	.	.	C	.	.	.	C	G	.	.	.	.	.	.	A	GCCA	T	A	.	T	.	T	.	.	T	.	.	.	C
B4*	9	.	T	.	.	.	.	G	.	.	G	.	.	.	.	.	G	.	.	.	T	.	.	.	—	.	.	.	.	.	G	A	.	.	A	.	.	.	C	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
B5†	1	.	T	.	.	.	.	G	T	.	G	.	.	.	.	.	G	.	.	.	.	.	.	.	—	.	.	.	.	.	G	A	.	.	A	.	.	.	C	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
B6	1	.	T	.	.	.	.	G	.	.	G	.	.	.	.	.	G	.	.	.	.	.	.	.	—	.	.	.	.	.	G	A	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
B7	4	.	T	.	.	.	T	G	.	.	G	.	.	.	.	.	G	.	.	.	.	.	.	.	—	.	.	T	.	.	.	A	.	.	.	.	.	T	.	G	.	.	.	.	A	.	G	.	.	A	.	A	T	.	.
B8	3	.	.	.	.	.	.	G	.	—	G	.	.	.	.	.	G	.	.	.	.	.	.	.	—	.	A	T	.	.	.	A	.	.	.	.	.	T	.	.	.	.	.	.	A	.	G	.	.	.	.	A	T	.	.
B9†	1	.	.	.	.	.	.	G	.	N	G	.	.	.	.	.	G	.	.	.	.	.	.	.	—	.	.	T	.	.	.	A	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
B10	1	.	.	.	.	.	.	G	.	—	G	.	.	.	.	.	G	.	.	.	.	G	.	.	—	.	.	.	.	.	G	A	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
B11*	1	.	.	.	.	.	.	G	.	—	G	.	.	.	.	.	G	A	.	.	.	.	.	.	—	C	.	.	.	.	.	A	.	.	.	.	T	.	C	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
B12*	3	.	.	.	.	.	.	G	.	—	G	.	.	.	.	.	G	.	T	.	.	.	.	.	—	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B13	2	.	T	.	C	A	.	G	.	.	G	.	C	.	.	.	G	.	.	.	.	.	—	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	C	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	N
B14	1	.	T	.	.	.	.	G	.	.	G	C	.	.	.	.	G	.	.	.	.	.	—	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
Bonobo		A	C	N	T	G	C	A	C	C	A	T	A	C	A	A	A	G	C	A	C	A	C	A	A	T	G	C	T	A	A	G	G	A	G	C	C	C	T	C	T	—	G	C	G	C	C	C	A	G	C	G	C	C	T
Chimp		A	C	G	T	G	C	A	C	C	A	T	A	C	A	A	A	G	C	A	C	A	C	A	A	T	G	C	T	A	A	G	G	A	G	C	C	C	T	C	T	—	G	C	K	C	C	C	A	G	C	G	C	C	T
Gorilla		A	C	G	T	G	C	G	C	C	A	T	A	Y	R	A	T	G	C	A	C	A	C	A	A	T	G	A	T	A	A	G	G	A	G	C	C	C	T	C	T	—	G	C	G	C	A	C	A	G	C	G	C	C	T
Orang		A	C	G	T	G	T	G	C	C	A	T	A	N	N	N	N	G	C	A	T	A	C	A	—	T	G	C	T	A	A	G	G	A	G	C	C	C	T	C	T	—	G	C	G	C	A	C	A	G	T	G	C	C	T

*Shared with Mandenka and San.
†Shared with Mandenka.

Table S4. Probability of $l_b \geq 37$ for the Biaka data as a function of ρ

ρ/kb	Pr ($l_b \geq 37$)
0.00	0.271
0.25	$5.8 * 10^{-3}$
0.50	$3.7 * 10^{-4}$
0.75	$3.0 * 10^{-5}$
1.00	$8. * 10^{-6}$
1.25	$1.4 * 10^{-6}$
1.50	$<< 10^{-6}$

lb refers to the maximum numbers of pairwise congruent sites (1).

1. Wall JD, Lohmueller KE, Plagnol V (2009) Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Mol Biol Evol* 26:1823–1827.