Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins

Anders Eriksson[1] and Andrea Manica[1]

Evolutionary Ecology Group, Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, United Kingdom

Edited by Francisco Mauro Salzano, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, and approved July 20, 2012 (received for review January 19, 2012)

Recent comparisons between anatomically modern humans and ancient genomes of other hominins have raised the tantalizing, and hotly debated, possibility of hybridization. Although several tests of hybridization have been devised, they all rely on the degree to which different modern populations share genetic polymorphisms with the ancient genomes of other hominins. However, spatial population structure is expected to generate genetic patterns similar to those that might be attributed to hybridization. To investigate this problem, we take Neanderthals as a case study, and build a spatially explicit model of the shared history of anatomically modern humans and this hominin. We show that the excess polymorphism shared between Eurasians and Neanderthals is compatible with scenarios in which no hybridization occurred, and is strongly linked to the strength of population structure in ancient populations. Thus, we recommend caution in inferring admixture from geographic patterns of shared polymorphisms, and argue that future attempts to investigate ancient hybridization between humans and other hominins should explicitly account for population structure.

ABBA-BABA | *D* statistic | Neanderthal introgression | stepping stone model | out-of-Africa

The possibility of hybridization between anatomically modern humans and other hominins has been the focus of a heated debate spanning several decades (1–5). Four main alternative models have been put forward describing the origin of anatomically modern humans and their relationship with other hominins (6, 7). The Recent African Origin model argues that anatomically modern humans are the descendants of a recent expansion out of Africa that replaced older hominin populations without significant hybridization (8). The African Hybridization and Replacement model is similar to the Recent African Origin, but it allows for some hybridization with indigenous hominins during the expansion (9). Multiregional Evolution denies a recent African origin of modern humans, and argues that most hominins over the past 2 Myr are regional forms of a single species, with some gene flow occurring among contemporaneous populations in different areas throughout their existence (10, 11). Finally, the Assimilation model takes elements of the other models, because it accepts an important contribution of a recent African expansion, but it emphasizes assimilation rather than replacement of indigenous hominins during the expansion (7).

A pioneering study by Cann et al. (12) showed that the phylogeny of a global sample of mitochondrial sequences had the deepest divergences in Africa, and that most African and all non-African sequences clustered together in a single, more recent branch. This result was interpreted as supporting a replacement without hybridization scenario, and later work on the Y chromosome provided very similar results (13). Several studies on overall genomic diversity have also shown that African populations are more diverse than non-African ones, and that diversity declines progressively with increasing distance from Africa (14–16), as would be expected from the sequential bottlenecks experienced during an expansion without hybridization (15, 17). Though all these lines of evidence point strongly to a recent out-of-Africa (OOA) expansion providing the largest contribution to the gene pool of anatomically modern humans, they cannot rule out small levels of hybridization with other hominins, as the signal left behind in modern populations would be very difficult to detect (18). Some attempts to detect hybridization based on patterns of linkage disequilibrium have found a possible signal (19, 20), but these approaches rely on simplistic demographic models with assumptions that are not testable (21).

Our ability to directly assess the possibility of hybridization between anatomically modern humans and other hominins has been recently revolutionized by the development of techniques allowing the recovery of reliable genetic information from ancient specimens. In principle, if hybridization with another hominin only occurred within part of the range of anatomically modern humans, we would expect a recovered genome of the hybridizing hominin to share more polymorphisms with modern populations in that area compared with modern populations in other areas. Indeed, a first analysis of the Neanderthal genome revealed intriguing patterns in terms of shared polymorphisms with modern-day humans: Neanderthals are genetically more similar to present-day Eurasians than to present-day Africans (22). This asymmetry has been interpreted as evidence for hybridization between Neanderthals and anatomically modern humans during the latter's exit out of Africa (22). Given that there is no difference between Europeans and Asians in their similarity to Neanderthals, it has been argued that such hybridization would have had to happen at the very beginning of the OOA expansion, before the split between these two groups (22). Similarly, the genome from a newly discovered hominin from the Denisova caves in Siberia has been shown to share unusual polymorphisms with Australians and Melanesians (23–25), again suggesting hybridization. However, our lack of knowledge of both the geographic range of Denisovans, and of their exact taxonomic affinity to modern humans, makes it difficult to identify the exact scenario.

A possible complication in interpreting spatially heterogeneous patterns of similarity between any ancient hominin and modern human populations is that, though such differences might arise through recent hybridization, they could also, in principle, be the consequence of population structure in early

Author contributions: A.E. and A.M. designed research; A.E. and A.M. performed research; A.E. contributed new reagents/analytic tools; A.E. and A.M. analyzed data; and A.E. and A.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: am315@cam.ac.uk or aej44@cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1200567109/-/DCSupplemental.

humans (22), i.e., a case of incomplete lineage sorting (26). Despite both archaeological (27, 28) and genetic evidence (28–32) for ancient population structure in Africa, none of the genetic models used so far to study the geographic patterns of similarity has incorporated structure in Africa. Currat and Excoffier (5) used a spatial framework to investigate admixture with Neanderthals, but concentrated on the role of gene surfing during the expansion wave in affecting the distribution of polymorphisms unique to Neanderthals across Eurasia, provided admixture did happen. Their model showed that gene surfing leads to overestimates of admixture in nonspatially structured models (5); however, they assumed the ancestors of African modern humans to be unstructured and ignored the common ancestry between Neanderthals and modern humans. To fill this gap, we build a spatially explicit framework of the shared history of anatomically modern humans and Neanderthals (Fig. 1 and *Materials and Methods*), which allows us to investigate the extent to which ancient population structure might drive spatial differences in genetic overlap between an ancient hominin and modern-day populations. We restrict our analysis to Neanderthals, because we have a reasonably clear idea of their ancient distribution and ancestral relationship with anatomically modern humans, but the conclusions of our study apply generally to any hominin contemporary with early anatomically modern humans in Eurasia.

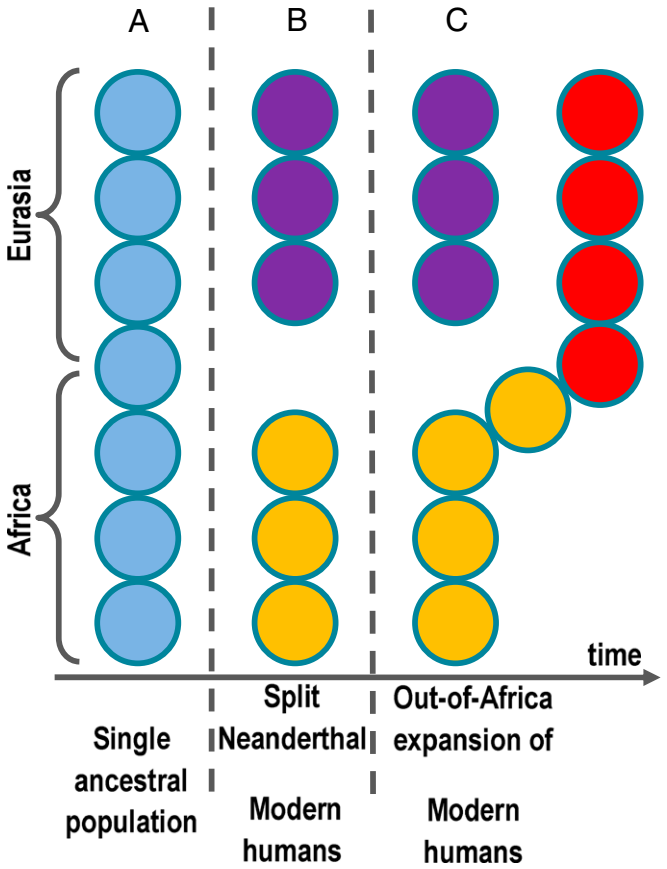


Fig. 1. Schematic representation of the spatially structured model representing the history of Neanderthals and anatomically modern humans. We start with a connected string of demes representing the common ancestor (in light blue), spanning Africa and Eurasia (*A*). Following the split (*B*), the northern part of the range becomes Neanderthals (purple), and the African part of the range eventually becomes modern humans (in orange). From this African range, modern humans later expand to colonize Eurasia (in red) (*C*). This figure only provides the logic of the model; for details, including the actual number of demes, see *Materials and Methods*.

Results

In our model, Africa and Eurasia are represented as a linear string of identical populations (demes), each of which can exchange migrants only with its adjacent neighbors. The simulation starts at 500 kya, with all demes occupied by the common ancestor of Neanderthals and anatomically modern humans (Fig. 1*A*). The connection between the African and Eurasian branches was truncated at 320 kya to represent the speciation of Neanderthals outside Africa (Fig. 1*B*). The populations in the African branch, which eventually turns into modern humans, were allowed to reexpand, eventually colonizing a set of Eurasian demes parallel to the ones occupied by Neanderthals (but without any admixture; Fig. 1*C*). As in Gronau et al. (33), the timings of the transition to modern demography, and of the subsequent expansion out of Africa, were free parameters in the model. We explored dates ranging from 100 to 200 kya for the transition to modern demography, and from 40 to 80 kya for the expansion. The population dynamics of modern humans (with parameters accounting for population size, growth rate, and colonization and migration rates to adjacent demes) was allowed to differ between this expansion and the ancestral phase. This part of the model is analogous to previous spatial models (14, 34–36) used to reconstruct the OOA expansion of anatomically modern humans. Our model differs from models previously used to investigate the similarity between Neanderthal and present-day humans in that it explicitly includes population structure within Africa and Eurasia, whereas previous models considered continents as unstructured compartments within which individuals mix freely.

We parameterized our model by comparing the patterns of genetic variability in modern populations predicted by the model with the observed quantities in the Human Genetic Diversity Project of the Centre d'Etude du Polymorphisme Humain (HGDP-CEPH) (37, 38), which comprises 51 populations covering the whole globe and typed for a large number of neutral microsatellites. This dataset has been used extensively to reconstruct the expansion out of Africa of anatomically modern humans (14, 15, 34, 39), and it provides a sensible yardstick to define the range of possible ancient demographics that would be compatible with the global distribution of genetic diversity in present-day humans. We used an approximate Bayesian computation (ABC) framework (40) to assess the likelihood of combinations of demographic parameters over a broad range of values, using time to most recent common ancestor (TMRCA) between major continents as summary statistics (see *Materials and Methods* for details and Fig. S1 for a schematic representation of the analysis). The demographic parameters describing the OOA expansion (Fig. S2) were compatible to those found in earlier studies using similar spatial models (14, 35): a rapid expansion with relatively small founding populations in each deme, and high migration rate between neighboring demes after the initial colonization. The ancient population demography was characterized by small carrying capacities and high migration rate. The time of transition to modern demography was compatible to recent estimates based on whole-genome sequences (33). Times of expansion in the explored range between 40 and 80 kya provided generally good fits, without any particular time being favored by the model.

To explore how different measures of the similarity between Neanderthal and different populations of present-day humans were affected by realistic levels of population structure, we used the best 0.05 percentile of demographic parameter combinations (referred to as demographic scenarios in the rest of the text), weighted by their likelihood as estimated by ABC. We considered two measures originally developed in the paper reporting the

draft of the Neanderthal Genome (22). Because both approaches lead to similar conclusions, we present the results from the *D* statistics [supplementary online material (SOM) 15 in ref. 22] in the main text, whereas the results from the degree of matching between ancient and derived SNPs in candidate regions (SOM 17 in ref. 22) are reported in the supplementary text. The *D* statistics follow the following rationale. Consider four sequenced genomes: one each from two modern populations (P_1 and P_2); one from an ancient hominin, such as Neanderthal (N); and one from an outgroup, such as a chimpanzee (O). We confine our analysis to a subset of the genome that we can align perfectly across the four samples. We now focus on positions where P_1 and P_2 differ from each other, and for which N and O also differ. We call allele "A" the variant found in O, and allele "B" the other variant found in N. For the ordered set {P1,P2,N,O}, we can then have either pattern ABBA or BABA. Pattern ABBA represents a situation where the first modern population is more similar to the ancestral individual than the second population, whereas BABA represents the opposite case. The *D* statistic is the difference in the number of sites with either pattern (no. of ABBA − no. of BABA) divided by the total number of sites being considered (no. of ABBA + no. of BABA). Under a scenario of no admixture between the ancestral population and the two modern populations (and assuming unstructured populations), we expect *D* to be equal to zero. Conversely, if admixture occurred between the ancient hominin and one of the modern populations, *D* will deviate from zero.

When the Neanderthal genome was compared with genomes from modern populations, a significant deviation from zero was detected for *D* when comparing Africans to Europeans and Africans to Asians, but not when comparing Europeans to Asians (22). This pattern could imply admixture between Neanderthals and the ancestor of both Europeans and Asians before their split. Population structure can confound results from the *D* statistic (21), but the key question is whether realistic levels of population structure could generate the patterns observed in the Neanderthal genome. When we used our demographic scenarios (described above) to generate synthetic genomes for both Neanderthal and modern human populations, we found that the predicted levels of *D* for the different pairwise comparisons fitted very well the observed patterns. For Africans vs. Europeans and Africans vs. Asians, the distributions of predicted *D* values are tightly centered around the corresponding observed values; for European vs. Asians, we recover the expected negative values, even though the predicted distribution is centered around values slightly closer to zero than the observed one (Fig. 2). The same conclusion holds for any choice of African and OOA populations (Fig. S3), which is agreement with the finding that the Aborigines and Eurasian genomes show similar overlap with the Neanderthal genome (25). Similar conclusions were reached using the alternative measure presented in SOM 17 of Green et al. (22) (*SI Text* and Fig. S4).

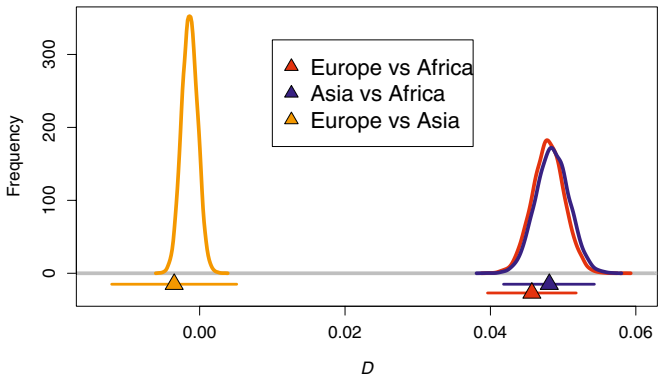


Fig. 2. Observed and simulated *D* statistics. *D* statistics for Europe vs. Africa (red), Asia vs. Africa (blue), and Europe vs. Asia (orange) in the original data (22) (triangles, with a 95% confidence interval) and as predicted by our spatial model (lines, representing the distributions of possible values).

Discussion

For both measures of shared polymorphisms between Neanderthals and modern humans, the patterns observed in the data could be generated by our spatial model that included ancient population structure of a strength that is compatible with the modern distribution of genetic diversity, even in the complete absence of hybridization. For the *D* statistics, the patterns produced by our spatial model were virtually identical to the ones observed in the data. For the candidate region approach, we also obtained a very good fit to the data; indeed, our spatial model provided a more realistic description of the data than the original nonspatial model that included hybridization in Green et al. (22), which failed to produce the correct level of matching for ancestral tag SNPs (see *SI Text* for details).

Currat and Excoffier (5) had already shown that models lacking spatial realism tend to overestimate the level of hybridization by ignoring gene surfing during the human expansion into Eurasia; including ancient structure in Africa further undermines the case for hybridization. Our conclusion was robust to the use of different measures of shared polymorphism, clearly indicating the limitation of looking for polymorphisms that are only shared between a single ancient genome and a restricted group of modern populations. In addition, our results show the extent to which ignoring spatial population structure causes models to underestimate the geographic differences expected in the absence of hybridization. Therefore, the issues raised in this paper should also apply to claims of admixture based on the distribution of linkage disequilibrium (19, 41).

Since the Neanderthal genome was published, another ancient genome has been recovered from a previously undescribed hominin from the Denisova caves (23). Comparisons with a wide range of modern populations revealed an excess of shared polymorphisms between this hominin and Australians and Melanesians, which has been interpreted as a sign of recent admixture (23, 24). Lack of knowledge on the past distribution and exact taxonomic affinity of Denisovans prevents us from quantifying the role ancient population structure played in creating similarities between modern humans and this hominin. However, our Neanderthal results suggest that ancient population structure could be an important determinant of spatial patterns of genetic overlap in this case also.

In conclusion, we urge caution in inferring recent admixture from geographic differences in genetic overlap between ancient hominins and modern-day populations. Though we do not claim that anatomically modern humans never admixed with other hominins, our results imply that current evidence for such admixture events is inconclusive at best. Future tests, to be convincing, will need to show that the genetic patterns used to invoke hybridization cannot be explained by population structure, for which there is both genetic (28–31) and archaeological evidence (27, 28). A key step toward such evidence will be the addition of ancient genomes of modern humans and other hominins from multiple locations, because they will allow reconstruction of ancient population structure and its possible effects on the genetic patterns in modern humans.

Materials and Methods

Demographic Model. We represented the joint history of anatomically modern humans and Neanderthals using a spatially explicit stepping-stone model (Fig. 1). Before the modern human/Neanderthal split, the common ancestor of the two hominins occupied a linear string of demes (separated by 100 km) stretching from south sub-Saharan Africa into Eurasia (Fig. 1*A*; 130 demes in total). These demes contained K_0 individuals and exchanged

m_0K_0 migrants with each neighboring deme each generation. At 320 kya, future modern humans and Neanderthal split and evolved separately; this was represented by introducing a barrier, at deme 70 (in North Africa), to migration between Africa and the Eurasian branch of the stepping-stone model (Fig. 1*B*).

At time t_{modern}, while modern humans were still confined to Africa, the demographic parameters were allowed to change from ancestral to modern: carrying capacities changed from K_0 to K, and migration rates from m_0 to m. At t_{exit}, the barrier preventing the exit out of Africa was removed, and modern humans spread to colonize Eurasia. This process was represented by connecting the African branch of the stepping-stone model to a new branch with initially empty demes, stretching across Eurasia and the American continent (Fig. 1*C*, 260 demes). During the spread, full demes sent a fraction c of colonizers to adjacent empty demes. Population size grew linearly at the rate rK per generation, until it reached the new carrying capacity K. Each pair of adjacent occupied demes exchanged mN_{min} migrants per generation, where N_{min} represents the smaller of the two population sizes. Thus, when the new branch was connected to Africa, humans spread across it, taking approximately $(1 - c)/r$ generations to fill up a deme and send colonizers to the next empty deme. Once all demes had reached carrying capacity, population size stayed constant (at K) until present.

Parameterizing the Model. To parameterize the model, we fitted the parameters (m, c, r, K, m_0, K_0, t_{modern}, and t_{exit}) to estimates of average TMRCA within and between populations (see Fig. S1 for a schematic representation). TMRCAs were calculated from the mean square difference of repeat counts in di- and trinucleotide microsatellite markers (41), genotyped in over 1,000 individuals from 51 populations in the HGDP-CEPH panel (37, 38). Dinucleotide markers where calibrated using the mutation rate of Dib et al. (42), $\mu = 1.52 \times 10^{-3}$ single-step mutations per 27 y TMRCA of trinucleotide markers were scaled to match the average TMRCA of the dinucleotide markers.

The predicted TMRCA for a given parameter combination was calculated as follows: we first ran the demographic model described in the previous section, and then generated 100 gene genealogies for 10 individuals in each of the 51 populations corresponding to the HGDP-CEPH populations in our data (placed according to the deme corresponding to the distance from a location in sub-Saharan Africa, calculated using shortest distances on land, as in refs. 15 and 43). We traced gene genealogies backward in time, generation by generation, assuming diploid, random mating within each colonized deme, and with migration probabilities to neighboring demes given by the demographic model.

We fitted our model in the ABC framework, using the ABC-GLM algorithm implemented in the ABCtoolbox software (44). We generated six summary statistics from the average TMRCA between continents. We treated Europe and Central Asia as one continent (Eurasia), and East Asia as a separate continent. Because Oceania only has two populations (both in Papua New Guinea), we included these populations in the East Asian set. Our summary statistics are thus $T_{Africa,Eurasia}$, $T_{Africa,EastAsia}$, $T_{Africa,America}$, $T_{Eurasia,EastAsia}$, $T_{Eurasia,America}$, and $T_{America,America}$.

We started by randomly sampling 2.2 million parameter values from the following ranges: $m \in [10^{-6}, 0.33]$, $c \in [10^{-4}, 0.33]$, $r \in [0.01, 1]$, $K \in [10, 10^5]$, $K_0 \in [10, 10^5]$, $m_0 \in [10^{-6}, 0.33]$, $t_{modern} \in [100, 200]$ (kya), and $t_{exit} \in [40, 80]$ (kya). All parameters (with the exception of t_{modern} and t_{exit}) were log-transformed to ensure an adequate exploration of the large ranges of possible values. We further imposed (through rejection sampling) the constraint $cK < K/2$ (cannot send out more colonists than individuals). Finally, we used ABC to estimate the likelihood of the 0.05% best-fitting scenarios (corresponding to 1,115 scenarios) and to generate parameter posterior distributions (Fig. S2).

Analysis of the *D* Statistic for Allele Sharing Between Humans and Neanderthals. We attempted to match the sample design and analysis of SOM 15 of Green et al. (22) as closely as possible within our framework. For each of the parameter combinations selected by the rejection step of ABC, we generated 10 million unlinked SNPs. When generating samples from our models, we took 10 individuals each from deme 10 (Africa), deme 90 (Europe), and deme 130 (East Asia). We placed the Neanderthal on the northern branch of the stepping-stone model (which became separated from the African at the split between humans and Neanderthals) at location that matches the distance between the origin in sub-Saharan Africa and the Vindija cave in Croatia [where the Neanderthal sample used in Green et al.'s (22) analysis was found]. For each SNP, the gene genealogy of the sample was generated as described in the previous section. For each of the scenarios selected by ABC, we then generated mutations on the gene genealogy according to the Jukes–Cantor model (assuming a split with chimpanzees 6 Mya) with mutation rate 2.5×10^{-8} per gamete per generation (45), and calculated the D statistic for each pair of populations. For each of the three comparisons (Europe vs. Africa, Asia vs. Africa, and Europe vs. Asia), we estimated the corresponding distributions of D values from 10,000 bootstrap samples. Finally, to see how the D statistic depends on the distance from sub-Saharan Africa, we repeated the analysis with 10 individuals from every 10th deme (every 1,000 km). Fig. S3 shows the average D between all pairs of populations, as a function of distance from the origin of the stepping-stone model.

Analyses of Candidate Regions for Gene Flow from Neanderthals. To be able to directly compare our model to the real data and simulations in Green et al. (22), we closely followed the analyses described in SOM 17 of that paper. See *SI Text* for details.

ACKNOWLEDGMENTS. We thank S. Lycett, A. Vail, and three anonymous referees for useful comments on the manuscript. This work was supported by the Leverhulme Trust and the Biotechnology and Biological Sciences Research Council.

1. Herrera KJ, Somarelli JA, Lowery RK, Herrera RJ (2009) To what extent did Neanderthals and modern humans interact? *Biol Rev Camb Philos Soc* 84:245–257.
2. Hofreiter M (2011) Drafting human ancestry: What does the Neanderthal genome tell us about hominid evolution? Commentary on Green et al. (2010). *Hum Biol* 83:1–11.
3. Currat M, Excoffier L (2004) Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biol* 2:e421.
4. Trinkaus E, et al. (2003) An early modern human from the Peştera cu Oase, Romania. *Proc Natl Acad Sci USA* 100:11231–11236.
5. Currat M, Excoffier L (2011) Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. *Proc Natl Acad Sci USA* 108:15129–15134.
6. Stringer C (2002) Modern human origins: Progress and prospects. *Philos Trans R Soc B Biol Sci* 357:563–579.
7. Aiello LC (1993) The fossil evidence for modern human origins in Africa—a revised view. *Am Anthropol* 95:73–96.
8. Stringer CB, Andrews P (1988) Genetic and fossil evidence for the origin of modern humans. *Science* 239:1263–1268.
9. Brauer G (1989) The evolution of modern humans: A comparison of the African and non-African evidence. *The Origins and Dispersal of Modern Humans: Behavioural and Biological Perspectives*, eds Mellars P, Stringer CB (Edinburgh Univ Press, Edinburgh).
10. Hawks J, et al. (2000) An Australasian test of the recent African origin theory using the WLH-50 calvarium. *J Hum Evol* 39:1–22.
11. Wolpoff MH, Hawks J, Frayer DW, Hunley K (2001) Modern human ancestry at the peripheries: A test of the replacement theory. *Science* 291:293–297.
12. Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325:31–36.
13. Underhill PA, Kivisild T (2007) Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* 41:539–564.
14. Liu H, Prugnolle F, Manica A, Balloux F (2006) A geographically explicit genetic model of worldwide human-settlement history. *Am J Hum Genet* 79:230–237.
15. Prugnolle F, Manica A, Balloux F (2005) Geography predicts neutral genetic diversity of human populations. *Curr Biol* 15:R159–R160.
16. Li JZ, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319:1100–1104.
17. Manica A, Amos W, Balloux F, Hanihara T (2007) The effect of ancient population bottlenecks on human phenotypic variation. *Nature* 448:346–348.
18. Relethford JH (2008) Genetic evidence and the modern human origins debate. *Heredity (Edinb)* 100:555–563.
19. Plagnol V, Wall JD (2006) Possible ancestral structure in human populations. *PLoS Genet* 2:e105.
20. Hammer MF, Woerner AE, Mendez FL, Watkins JC, Wall JD (2011) Genetic evidence for archaic admixture in Africa. *Proc Natl Acad Sci USA* 108:15123–15128.
21. Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Mol Biol Evol* 28:2239–2252.
22. Green RE, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328:710–722.
23. Reich D, et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468:1053–1060.
24. Reich D, et al. (2011) Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am J Hum Genet* 89:516–528.
25. Rasmussen M, et al. (2011) An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334:94–98.
26. Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Mol Biol Evol* 5:568–583.
27. Gunz P, et al. (2009) Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. *Proc Natl Acad Sci USA* 106:6094–6098.
28. Scheinfeldt LB, Soi S, Tishkoff SA (2010) Colloquium paper: Working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. *Proc Natl Acad Sci USA* 107(Suppl 2):8931–8938.

29. Campbell MC, Tishkoff SA (2010) The evolution of human genetic and phenotypic variation in Africa. *Curr Biol* 20:R166–R173.
30. Henn BM, et al. (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc Natl Acad Sci USA* 108:5154–5162.
31. Harding RM, McVean G (2004) A structured ancestral population for the evolution of modern humans. *Curr Opin Genet Dev* 14:667–674.
32. Blum MGB, Jakobsson M (2011) Deep divergences of human gene trees and models of human origins. *Mol Biol Evol* 28:889–898.
33. Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet* 43:1031–1034.
34. Ramachandran S, et al. (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci USA* 102:15942–15947.
35. Deshpande O, Batzoglou S, Feldman MW, Cavalli-Sforza LL (2009) A serial founder effect model for human settlement out of Africa. *Proc Biol Sci* 276:291–300.
36. DeGiorgio M, Jakobsson M, Rosenberg NA (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 USA* 106:16057–16062.
37. Cann HM, et al. (2002) A human genome diversity cell line panel. *Science* 296: 261–262.
38. Rosenberg NA, et al. (2002) Genetic structure of human populations. *Science* 298: 2381–2385.
39. Fagundes NJR, et al. (2007) Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci USA* 104:17614–17619.
40. Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.
41. Eriksson A, Manica A (2011) Detecting and removing ascertainment bias in microsatellites from the HGDP-CEPH Panel. *G3 (Bethesda)* 1:479–488.
42. Dib C, et al. (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152–154.
43. Manica A, Prugnolle F, Balloux F (2005) Geography is a better determinant of human genetic differentiation than ethnicity. *Hum Genet* 118:366–371.
44. Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* 11:116.
45. Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297–304.

Supporting Information

Eriksson and Manica 10.1073/pnas.1200567109

SI Text

Analyses of Candidate Regions for Gene Flow from Neanderthals. The original publication of the draft Neanderthal genome (1) included two different metrics to describe the degree of shared polymorphism between Neanderthals and different modern human populations. We investigated the effect of ancient population structure on both metrics. The effect on the *D* statistics [supplementary online material (SOM) 18 in ref. 1] is described in the main text. Here we describe the effect on the metric presented in SOM 17 of ref. 1, which focuses on ancestral and derived SNPs in sections of the human genome that have large times to most recent common ancestor (TMRCA).

SI Materials and Methods

We have attempted to replicate the analysis of Green et al. (1) as closely as possible within our model (see Fig. S5 for an outline of the main steps). The analysis compares the Neanderthal genome to African (AFR) and out-of-Africa (OOA) samples (individuals with European or Asian ancestry). Candidate regions were found by looking for regions with high TMRCA in the OOA sample compared with the TMRCA of the AFR sample. TMRCA was calculated from sequence data in 50-kb bins, sliding across the genome in steps of 10 kb, by first constructing an unweighted pair group method with arithmetic mean (UPGMA) tree (2) for the bin and then taking the average of the number of mutations from each tip of the tree to its root (the tips of the tree correspond to the observed sequences). Let S_T denote the ratio of TMRCA for the OOA sample to the TMRCA for the AFR sample for a given 50-kb bin. A region was then classified as "candidate" if six consecutive bins all had S_T values in the top 0.5% of all S_T values in the genome. Hence, candidate regions are 100 kb long.

Within each candidate region, tag SNPs were identified as follows. First, a UPGMA tree was constructed for the joint AFR and OOA sample. In this tree, tag SNPs are mutations on the root lineages, identified using parsimony, such that they separate a clade containing only lineages from the OOA sample from a joint OOA/AFR clade (Fig. S4, step 2). Finally, the tag SNPs were classified with respect to whether the OOA allele matches the Neanderthal (M for match or N for nonmatch), and whether the Neanderthal allele is derived (D) or ancestral (A), giving four categories: AM, AN, DM, and DN (Fig. S2, step 3).

For each of the parameter combinations selected by ABC (*Materials and Methods*), we generated 150,000 100-kb regions, corresponding to approximately five full genomes for each individual and parameter combination. We attempted to match the sample design of Green et al. (1) as closely as possible within our model. When generating sample sequences from our models, we took 24 individuals from randomly chosen demes in the range 70–120 (corresponding to Europe), 24 individuals from demes 130–170 (East Asia), and 23 samples from Africa. Because the American African sample used by Green et al. (1) was admixed with 80% African and 20% European ancestry, for each 46 gamete in the African sample we picked it with probability 0.8 from the range 1–10 (sub-Saharan Africa) and with probability 0.2 from a random deme in the European range. For the ABBA-BABA analysis, we placed the Neanderthal on the northern branch of the stepping-stone model (which became separated from the African at the split between humans and Neanderthals) in the deme corresponding to the distance between their origin in sub-Saharan Africa and the Vindija cave in Croatia, where the Neanderthal sample used in Green et al.'s (1) analysis was found. We simulated unlinked 100-kb regions by generating the gene genealogy of the sample as described in the previous section. We then generated mutations in the gene genealogy according to the Jukes–Cantor model (assuming a split with chimpanzees 6 Mya) with mutation rate 2.5×10^{-8} per gamete per generation (3). Using the procedure described above, we identified candidate regions and their tag SNPs from the simulated samples. Finally, we generated the distribution of fraction of matching for derived (DM vs. DN) and ancestral (AM vs. AN) as follows. First, for each of the selected demographic scenarios we used bootstrapping to generate the corresponding distributions of fraction of matching tag SNPs (for ancestral and derived separately) in a sample of 166 randomly chosen tag SNPs. Second, we averaged the distributions over the demographic scenarios weighted by likelihood as estimated by ABC (*Materials and Methods*).

SI Results

In the real data, there is an excess of matching vs. nonmatching positions (90.5% of derived SNPs are matching, and 69.5% of ancestral SNPs are matching, as obtained from table 5 in ref. 1). Green et al. (1) used a model with no population structure within continents to represent the split of Neanderthals from ancestors of anatomically modern humans and the latter's subsequent spread out of Africa. With no admixture between Neanderthals and modern humans, this model failed to capture the excess of matching to nonmatching (15.6% of derived and 50.7% of ancestral predicted as matching). However, introducing 1–4% of admixture just after modern humans' spread out of Africa raised the fraction of matching to 77.5% among derived SNPs (but gave only 38.3% matching of ancestral SNPs).

When we considered the best 0.05% of our parameter combinations, weighted by the likelihoods estimated by ABC, our predictions were compatible with the observed patterns. Indeed, the most likely proportions of matching for both derived and ancestral SNPs were closer to the observed values than the predictions of the models with and without admixture presented in Green et al. (1), with 83.4% of derived and 67.1% of ancestral SNPs matching (Fig. S4 *A* and *B*, respectively).

1. Green RE, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328: 710–722.
2. Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. *Univ Kans Sci Bull* 28:1409–1438.
3. Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297–304.

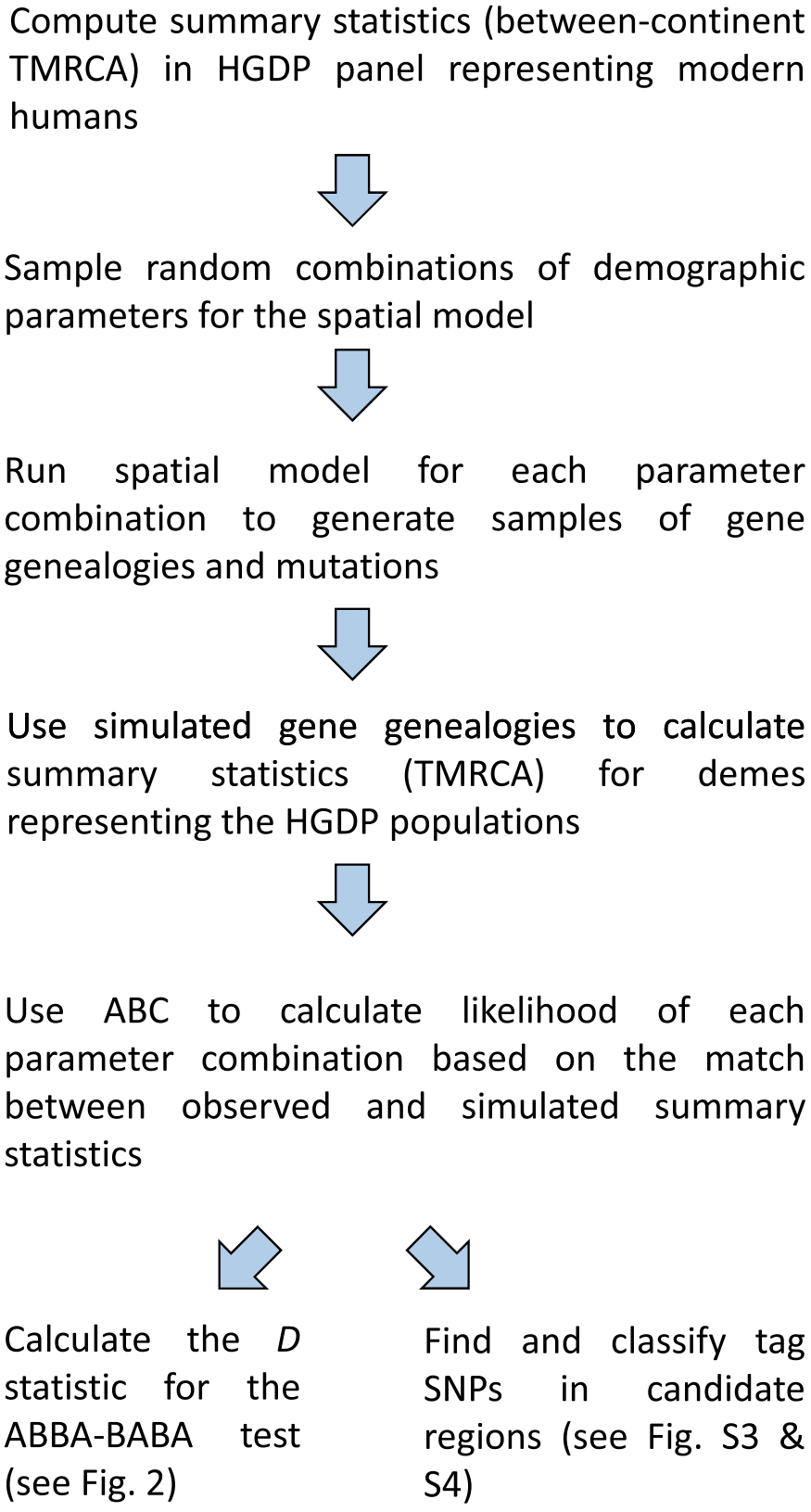


Fig. S1. Schematic representation of the logic behind model fitting.

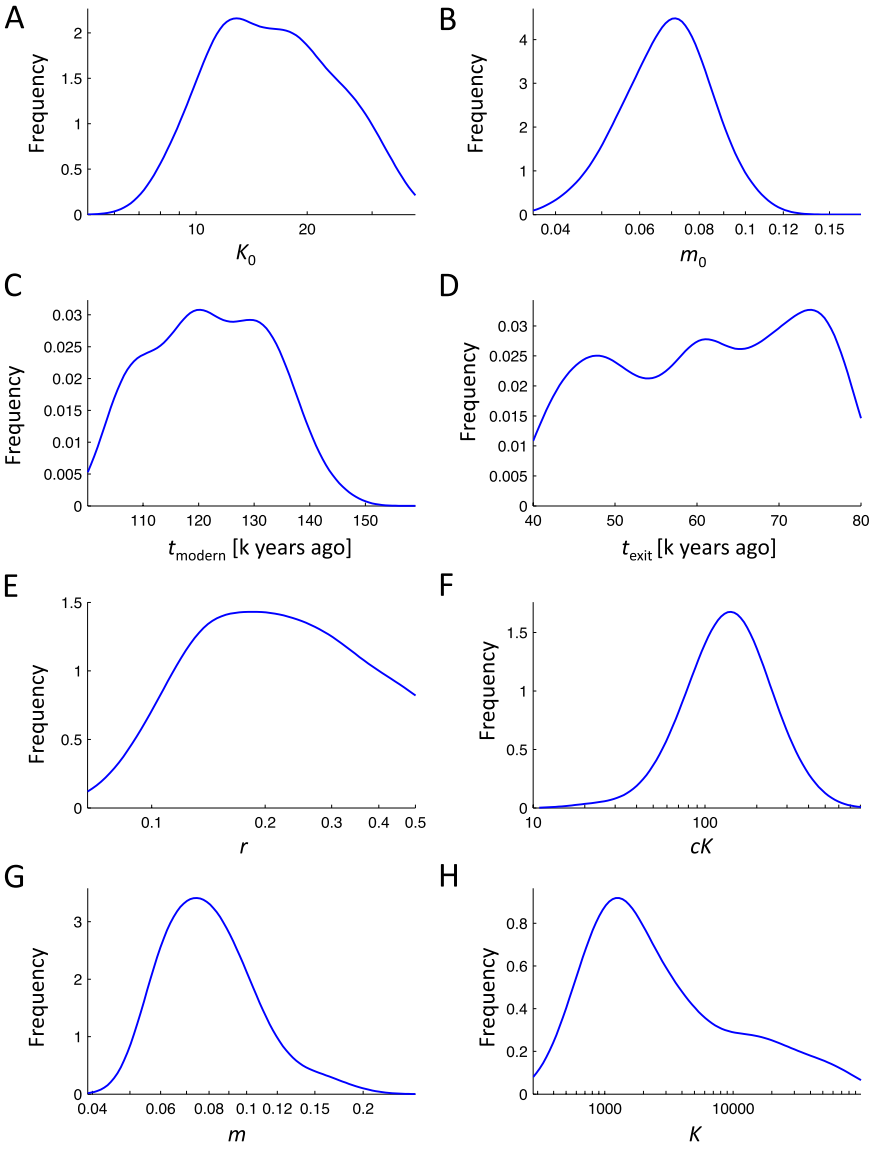


Fig. S2. Posterior distributions of the demographic parameters based on ABC. (*A*) Ancient carrying capacity, K_0; (*B*) migration rate, m_0; (*C*) timing for the transition from ancient to modern demography, t_{modern}, in kya; (*D*) timing for the expansion of modern humans out of Africa, t_{exit}, in kya; (*E*) modern growth rate, r, during the out-of-Africa expansion; (*F*) modern effective founder population size, cK; (*G*) modern migration rate, m; and (*H*) modern carrying capacity, K.

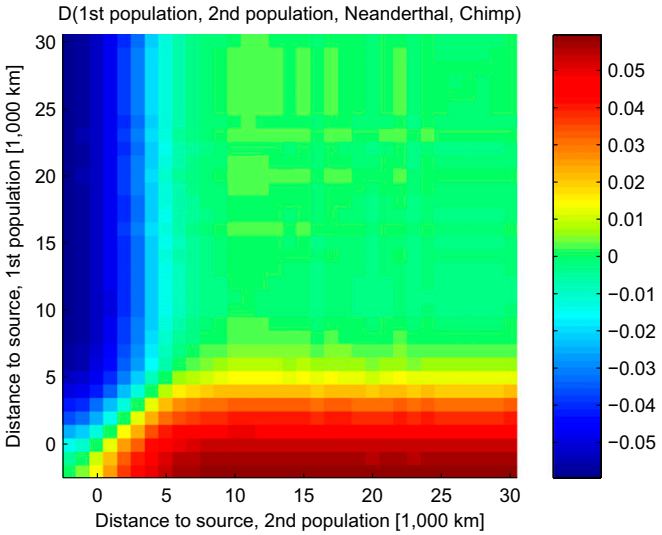


Fig. S3. *D* values for every pair of populations, averaged over the best-fitting scenarios.

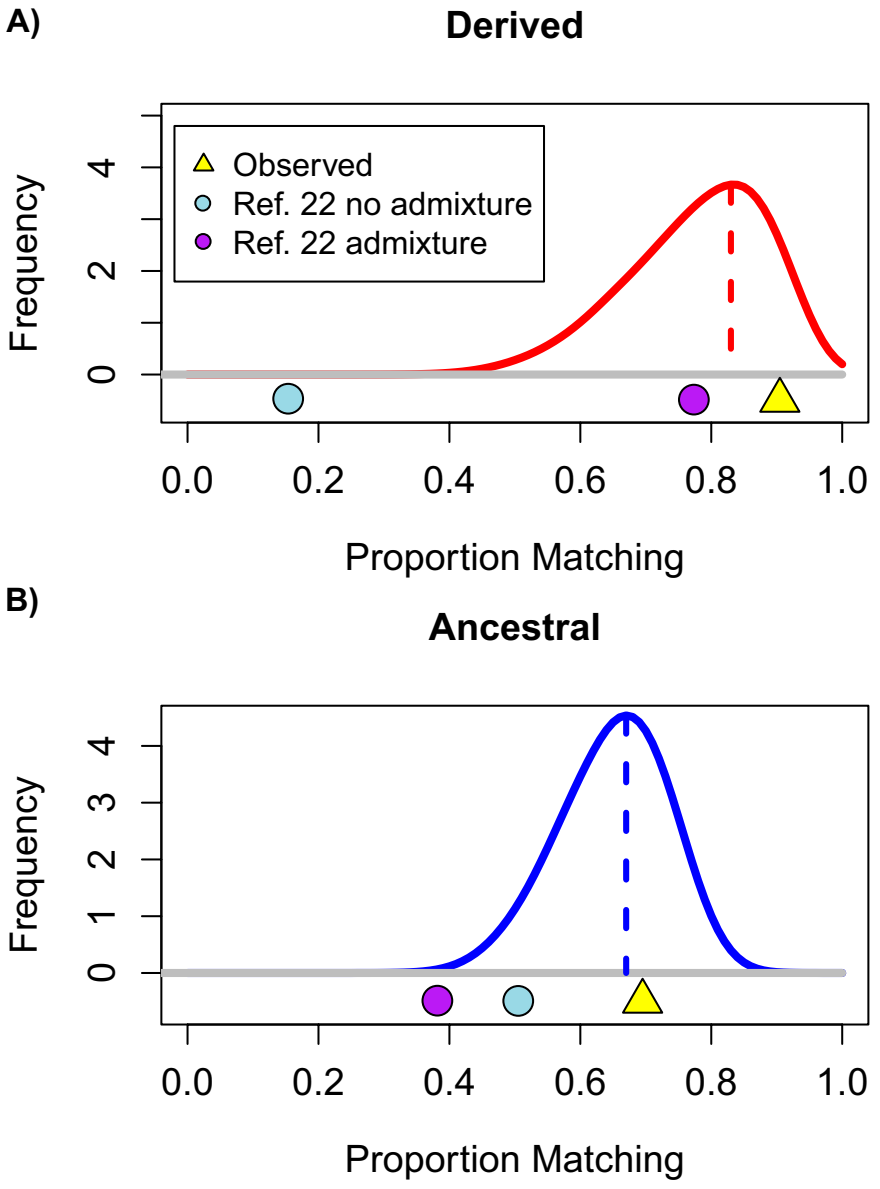
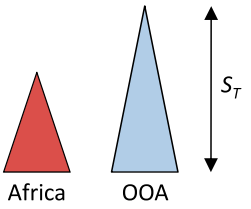


Fig. S4. The proportion of Neanderthal alleles matching those of Eurasians, separated into derived (*A*) and ancestral (*B*). In each panel we show the actual data (yellow triangle) and the proportions predicted by the Green et al. (1) model without (light blue circle) and with admixture between Neanderthal and Eurasians (purple circle). Results from our spatially structured model are presented as frequency plots of predicted values (red line for derived, and blue for ancestral), with a vertical dashed line representing the most likely fit.

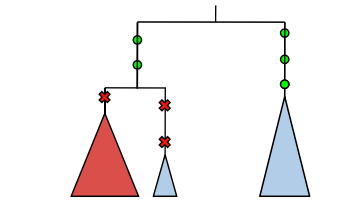
Step 1: Select candidate regions



S_T = average number of mutations to the root, estimated separately for African and OOA samples using UPGMA.

Select regions with large S_T ratio (OOA vs. Africa) as candidate regions.

Step 2: Find tag SNPs in joint UPGMA tree



Tag SNP
Non-tag SNP

Tag SNPs are SNPs on root lineages separating an OOA clade from a cosmopolitan clade (OOA + Africa).

Step 3: Classify tag SNPs

Neanderthal matches Chimpanzee?	Neanderthal matches OOA? Yes	No
Yes	AM	AN
No	DM	DN

Classify tag SNPs by comparing the Neanderthal allele to the OOA and the Chimpanzee alleles

Fig. S5. Schematic representation of the candidate region approach.