

Hypothesis: A modern human range expansion ~300,000 years ago had a critical role in Neandertal origins

David Reich^{1,2,3,4}

¹ Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA

² Department of Genetics, Harvard Medical School, Boston, MA, USA

³ Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA

⁴ Broad Institute of MIT and Harvard, Cambridge, MA, USA

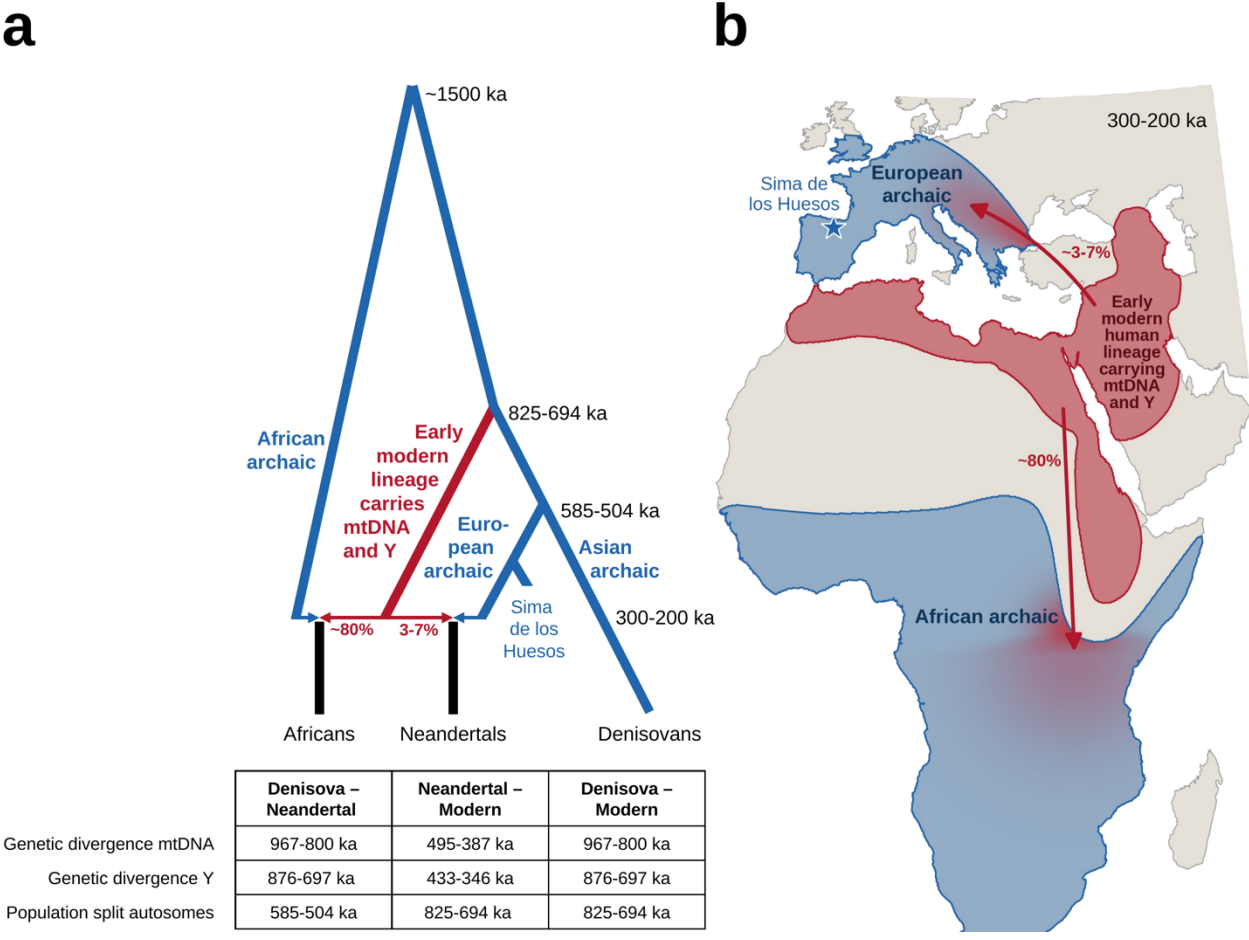
Correspondence: reich@genetics.med.harvard.edu

This paper proposes that a critical event in Neandertal evolution was a range expansion of early modern humans into Europe 300 thousand years ago (ka). As bands of early modern humans spread, they incorporated resident archaic people leading ultimately to so much admixture that the resulting population carried ~3-7% modern ancestry. I identify multiple ways in which this would lead to a high probability of these direct ancestors of Neandertals retaining early modern human lineages at both mitochondrial DNA and Y chromosomes while clustering with Denisovans genome-wide. A previous explanation for these patterns without the unique population dynamics associated with a range expansion has less than a 3% chance of matching the data. In Africa, the hypothesis also has explanatory power: if the expanding population also mixed African archaic humans, it could account for the evidence that all humans living today are admixed between lineages that separated more than a million years ago and remixed ~300 ka. This hypothesis of a bicontinental range expansion correlates with shared archaeological traits between West Eurasia and Africa that were far less present in East Asia, including Levallois point stone tools, improved hafting, use of fire and ochre, and larger and more organized campsites. Motivated by the evidence that an early modern human expansion had a role in Neandertal and modern human evolution, I identify 66 loci spanning <0.1% of the genome where people today share ancestry with Neandertals more recently than separation from Denisovans. This includes a haplotype at *FADS1/FADS2* that codes for an enzyme modifying polyunsaturated plant-derived fatty acids and may reflect adaptation to a more diversified diet, and an amino acid substitution at *ARHGAP32* modulating neural connections that may be important in modern human cognitive evolution.

A genetic puzzle

Neandertals and Denisovans were extinct populations of humans who lived across Eurasia until ~40 ka. Analysis of high-coverage genome sequences from them has led to the inference that they separated from a common archaic ancestral population 585-504 thousand years ago (ka), which in turn separated from a common ancestral population with modern humans 825-694 ka (Figure 1; ranges of dates from ref. 1). This reconstruction is consistent with the estimated genetic divergence of Denisovans and modern humans at two loci with enough informative sites to allow accurate estimates: at mitochondrial DNA (mtDNA) 967-800 ka (95% confidence intervals from ref. 1 here and in what follows), and on the Y chromosome 876-697 ka. But the estimated genetic splits of Neandertals and modern humans living today are hundreds of thousands of years younger than the inferred population separation—for mtDNA 495-387 ka^{1,2}, and for the Y chromosome 433-346 ka^{1,3}—which can only be explained by gene flow between the lineages after the inferred population split time. Multiple studies of the rest of the genome (chromosomes 1-22 and X) have found confirmatory evidence for this gene flow, inferring that Neandertals harbor 3-7% admixture^{4,5,6} from an early modern human lineage ~300-200 ka^{5,6,7}, clearly the vector for the shared mtDNA and Y lineages (genetic divergence times must be at least as old as the time since populations share ancestors, but can be hundreds of thousands of years older). A final line of evidence consistent with a ~300-200 ka gene flow from early modern humans into Neandertals is ancient DNA from an individual from Sima de los Huesos in Spain who lived 430-300 ka⁸, and who carried archaic lineage mtDNA related to Denisovans⁹ but a nuclear genome that shared drift with Neandertals¹⁰, as expected for an archaic European who lived prior to admixture with modern humans (Figure 1).

Figure 1: A bicontinental range expansion ~300 ka. (a) I start with a previously proposed model for the relationship of Neandertals, Denisovans, and modern Africans in which an early modern lineage contributed mtDNA and Y chromosomes to Neandertals and ~3-7% of the rest of the genome³. I graft onto it the deep substructure detected in people today, inferred in one study to contribute ~80% of the ancestry of present-day Africans¹¹, and inferred in several studies to have ended around ~300 ka^{7,11,12}. At bottom are estimates of divergences and population splits: range of estimated dates for the autosomes, and 95% confidence intervals for mtDNA and Y¹. (Not shown: <50 ka gene flow from Neandertals and Denisovans into non-Africans, and superarchaic gene flow to Denisovans.) (b) Schematic of what the range expansion might have looked like after ~300 ka, with an early modern human lineage in northeast Africa or the Near East expanding both into Europe and Africa and mixing with local archaic populations; ancestry within these regions would then have homogenized due to admixture among neighbors; the red-to-blue shading indicates mixing. The European archaic range is drawn based on an estimate of the later Neandertal range¹³. Coastlines are from the GLOBE 1-km digital elevation model¹⁴ via the global-land-mask package¹⁵, in an Albers equal-area projection.



Models without range expansion have a less than 3% chance of explaining these patterns

The conundrum posed by the fact that Neandertals harbor modern human-associated mtDNA and Y chromosomes but are more closely related to Denisovans on chromosomes 1-22 and X was recognized in the paper that reported the Denisovan Y chromosome sequences³. To explain it, the authors hypothesized that following admixture ~300-200 ka—which they modeled as around 5% early modern human admixture into an archaic human population in a non-geographically-explicit model—there was selection for the early modern human lineage on the entirely male line (tracked by the Y chromosome), or the female line (tracked by mtDNA), or both³. Using simulations of a non-recombining, uniparentally transmitted locus and a homogeneous population, they demonstrated that a selection coefficient of 1-2% for the genetic material from the early modern human lineage over the archaic one for either mtDNA or the Y chromosome would be sufficient to lead to probable fixation³. They proposed that such a fitness gap could have arisen through a greater accumulation of slightly deleterious mutations expected in populations of relatively smaller size (the archaic ancestors of Neandertals) than in ones of relatively larger size (the early modern ancestors of Neandertals)¹⁶. However, they did not test if the development of a fitness gap of this magnitude was consistent with biological constraints on mtDNA and the Y chromosome, or the probability of both the mtDNA and Y chromosome patterns being observed together.

To determine whether a fitness gap this large for mtDNA was likely given the biological constraints, I started with a previous framework for estimating the distribution of fitness effects of new mutations¹⁷, expanding it to include not just the approximately two-thirds of mtDNA in protein coding sequence but also the non-coding parts of mtDNA which are largely RNA genes (Supplementary Information section 1). I took into account all the sources of uncertainty I could identify, including in mutation rate, the fitness effects of mutations, and demography. I constrained

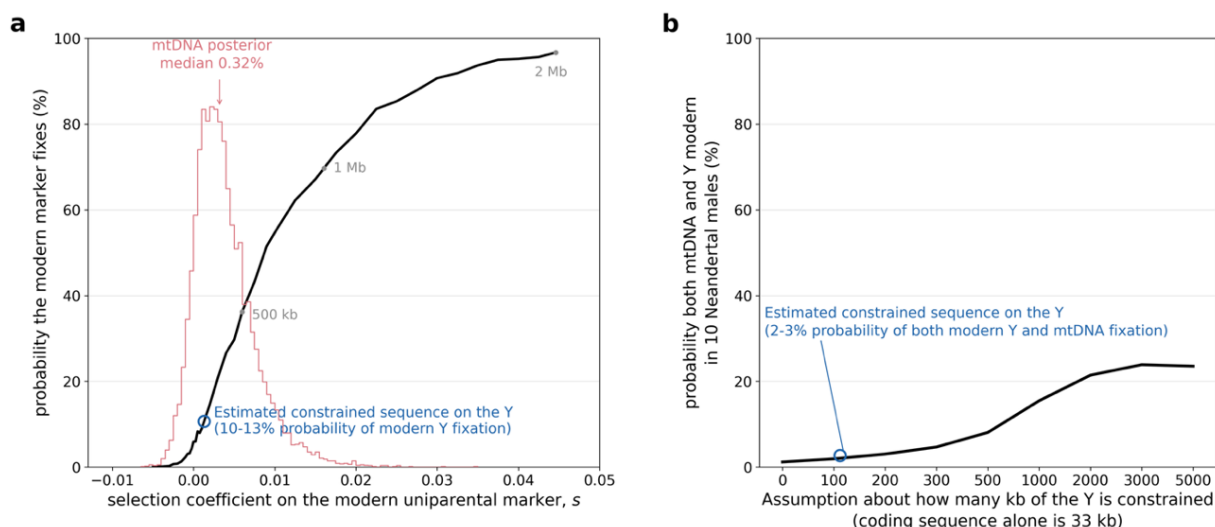
the mutation rate using the synonymous substitution counts on the modern and Denisovan mtDNA lineages, the fitness effects based on patterns of polymorphism within humans and divergence from archaic humans, and the fitness gap by the observed ratio of nonsynonymous substitutions on the two lineages (which is slightly higher on the modern than on the Denisovan side, opposite to what an excess archaic genetic load would predict). I inferred a posterior distribution on the mtDNA fitness gap with a median of 0.32% (95% of the density between [-0.2%, +1.3%]), leading to a probability of 23% of fixation of modern human mtDNA when I carried out simulations integrated over the uncertainty in the fitness gap (Figure 2a).

To determine whether a fitness gap this large for the Y chromosome was likely, I could not use the same direct approach as for mtDNA, because the fraction of the Y chromosome that is coding is on the order of 0.1%, providing too little data to constrain the inference (a previous estimate of Y chromosome load was not usable for my purposes, as its authors attribute it largely to palindromic repeats, which gene conversion between near-identical copies protects from irreversible accumulation of deleterious mutations¹⁸). As an alternative, I estimated the amount of constrained sequence in the single-copy part of the Y chromosome by analyzing conservation scores in an alignment of mammalian genomes¹⁹. I identified ~112 kb of conserved sequences, and by assuming that the distribution of selection coefficients for newly arising mutations is the same as for coding bases elsewhere in the genome (or weaker, translating to a faster rate of accumulation of slightly deleterious mutations), I inferred a median 0.09-0.17% fitness gap, leading to 10-13% probability of fixation of modern human Y chromosomes (Supplementary Information section 1).

Combining mtDNA and the Y chromosome, my simulations infer a 2-3% probability of modern human lineage fixation at both loci (Figure 2b). This is in fact an overestimate, because

hybridization is expected to pull down the frequencies of functionally important alleles from the minority admixing population due to “Dobzhansky-Muller Incompatibilities” (DMIs): co-evolved variants that only work when they are on the same genetic background causing compromised reproduction in hybrids that do not carry both²⁰. DMIs appear to have affected many loci in the ~300-200 ka mixture^{4,6}, and since they often involve mtDNA²¹, it is reasonable to expect them to have pushed down the probability of modern mtDNA fixation. I was not able to identify substitutions on both sides of an interface between mtDNA and nuclear interacting proteins that are often involved in a mtDNA DMI²² (Supplementary Information section 2), but since mitochondrial DMIs are not always mediated by such interactions²⁰, 2-3% is conservative.

Figure 2: Archaic-modern fitness gap is highly unlikely to have been sufficient to retain modern DNA at both uniparentally inherited markers. (a) To estimate the fitness gap between the archaic and the early modern ancestors of Neandertals, I conditioned on the observed patterns of polymorphism and divergence in mtDNA and estimates of the demographic parameters. I infer a 23% probability of retaining modern human lineage mtDNA integrating over the uncertainty (red). For the Y chromosome, I show the point estimate of the fitness gap. **(b)** The low probability of the data is evident from considering both loci together. I show the probability that 10 sampled Neandertals 50 ka all harbor early modern human mtDNA and Y chromosomes, similar to real data. This is 2-3% at the estimated 112 kb of constrained sequence in the non-palindromic Y.



Geographically explicit scenarios predict the data with high probability

Studies of diverse species have shown that when a population moves into territory occupied by a previously established group, large-scale incorporation of local genes occurs if there is even a small rate of interbreeding at the wavefront of the expansion. By the time the expansion finishes, the genetic composition can be largely local²³, consistent with the observation that Neandertals carried only a small proportion (~3-7%) of early modern human lineage ancestry genome-wide. A previous study considered how range expansion might have affected patterns of introgression in admixtures between modern humans and Neandertals, focusing not on the ~300-200 ka event, but instead on the ~50 ka admixture in which modern humans bearing Upper Paleolithic technology expanded into Europe²⁶. It inferred a <0.1% per generation rate of gene flow from Neandertals into expanding modern humans at locations where they overlapped, based on the absence of Neandertal ancestry in later European mtDNAs. The larger introgression proportion in the earlier ~300-200 ka event would imply a higher interbreeding rate between groups or a slower expansion, likely reflecting more modest biological or cultural gaps between the interacting groups.

The geographic explicitness of a range expansion results in population genetic dynamics that do not occur in non-geographically-explicit models²³. I hypothesized that this could help explain retention of modern ancestry not only in the autosomal genome, but also at the uniparentally inherited markers. In human foragers and great apes, dispersal events to join neighboring groups are nearly always mediated by one sex^{24,25}. In a range expansion, this would be expected to lead to the marker associated with the non-dispersing sex being retained by the expanding population, shielded from the archaic ancestry flooding the rest of the genome. In contrast, sex-bias does not help to explain the data for the previously proposed non-geographically-explicit model of a ~5% pulse of modern human flow into Neandertal ancestors³, since the lineage associated with the sex

that mates across populations would still have a small probability of fixing due to drift (Extended Data Figure 1.) Retention of modern human ancestry at the other uniparentally inherited marker is also more plausible in a range expansion, since at the beginning modern human lineages at both mtDNA and Y chromosomes are in the majority, so what is necessary to explain the data is to block archaic lineages from coming in; I show four plausible scenarios that do this in what follows.

To identify if there were plausible scenarios whereby both uniparentally transmitted markers could be retained with high probability during a range expansion, I developed two computer simulation frameworks capturing different features of the geography, timing, and population sizes in Neandertal ancestors before and after the ~300-200 ka mixture. In the “deme model” (inspired by previous modeling of range expansions^{23,26}), individuals live in a grid of usually 12x12 squares each of which can include individuals of both groups (archaic, or early modern human). Modern individuals have a reproductive advantage over archaic ones that live in the same deme, and occasionally interbreed across groups within a deme with the group identity of the offspring determined by either mothers or fathers. In the “roving band model”, there are on the order of 70 bands of typical size 25 scattered across a landscape that I match to the geography of the area of Europe where Neandertals lived¹³. Bands are 100% archaic or 100% early modern human, diffuse across the landscape each generation, and modern human bands have reproductive advantages over archaic ones when near enough to compete for the same resources. Individuals occasionally switch groups (males join their partner’s band in some sets of simulations, females join their partner’s band in other sets of simulations). Supplementary Information section 3 gives full details.

The results replicate the qualitative behavior of previously reported simulations of range expansions^{23,26}. The reproductive advantage of early modern humans at the wavefront of expansion

amplifies the archaic DNA they take on through admixture, so even after the first hundred generations more than half of the expanding population's ancestry is archaic. By the time the wavefront sweeps across the landscape, there is typically only a small fraction of early modern human lineage genes left (Figure 3a, animated in the Supplementary Video, Extended Data Figure 2). These patterns are robust to a wide range of assumptions (Extended Data Figure 3).

For each mechanism I hypothesized could retain the other uniparentally inherited marker, I tested a range of parameters (744 for the deme model, 567 for the band model) (Extended Data Figure 4). For parameter combinations matching the chromosome 1-22 ancestry (3-7% early modern human), and a plausible time for range expansion completion (100-1000 generations), I tabulated the rate of fixation of early modern human lineage mtDNA and Y chromosomes in 10 simulated Neandertal males (Supplementary Table 1). Four scenarios match the data with high probability (Figure 3b, Supplementary Information section 3).

The first scenario I found only works when males are the sex that switches groups. "Haldane's Rule"²⁷ states that when two populations have been isolated for long enough that first generation hybrids have fragile reproductive phenotypes (due to DMIs: pairs of co-evolved alleles that do not work if they aren't inherited together), it is almost always the heterogametic sex that is affected (males in mammals), blocking introgression of archaic Y chromosomes. If the fragile male reproductive phenotype is reduced fertility, this is not an effective mechanism for retaining early modern human Y (Extended Data Figure 4); This is because selection in hybrids against both partner alleles in the DMI rapidly ends the block to Y introgression. However, the other frequent manifestation of DMIs—hybrids of the heterogametic sex (males in mammals) having a skew toward mostly having children of the homogametic sex (daughters in mammals), a pattern seen in

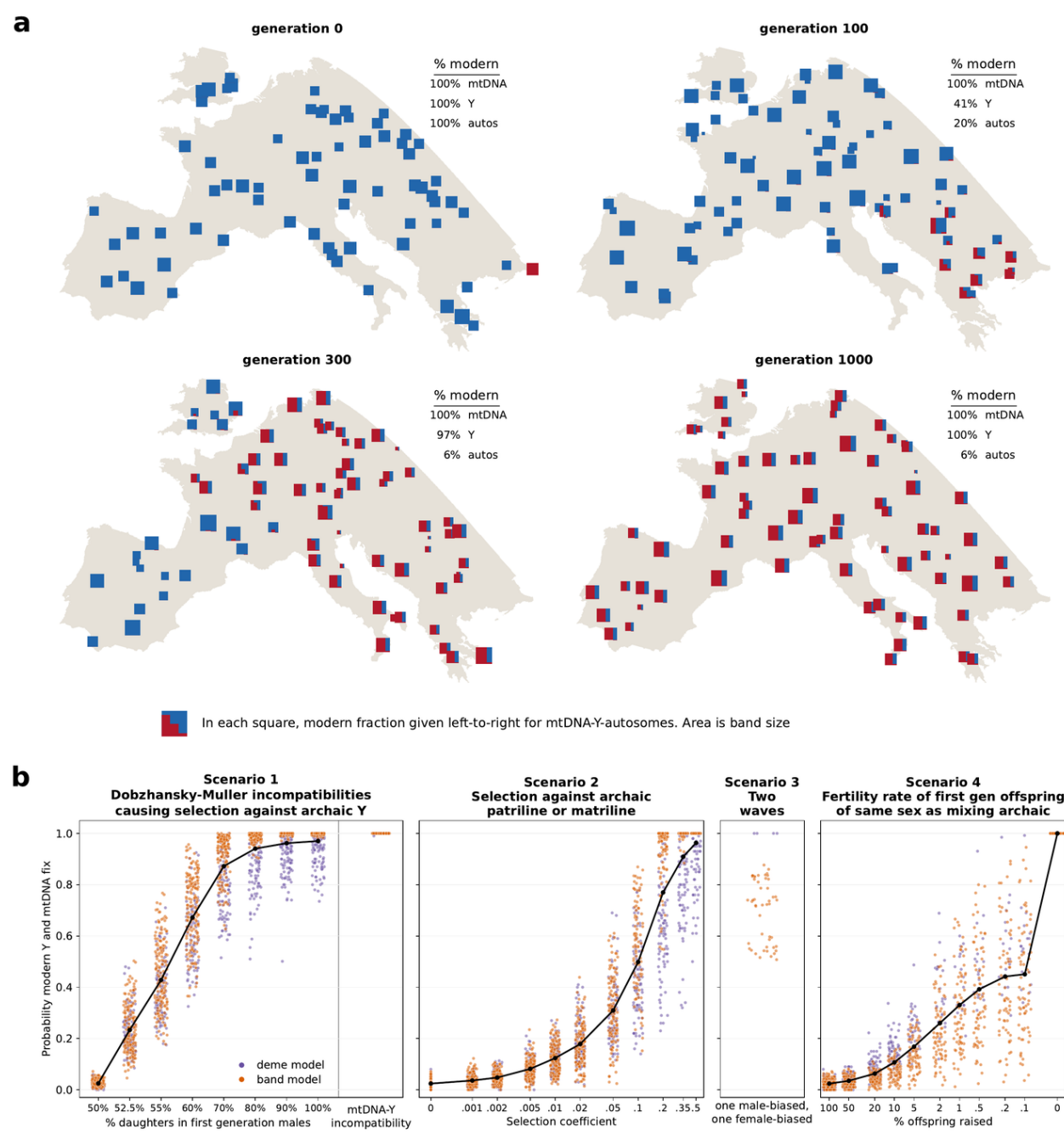
mice and fruit flies^{27,28,29,30,31}—effectively prevents archaic Y introgression (Figure 3b, Extended Data Figure 4). Daughter-skew arises due to selfish mutations on the X that cause men to overproduce sperm carrying X chromosomes, which are normally masked by suppressor alleles on the Y. But in hybrids, archaic Y lineages not carrying the suppressor are outcompeted by early modern ones carrying it. I show that a daughter skew as little as 60% is enough to make it likely to retain the modern Y chromosome (Figure 3b, Supplementary Information section 3).

The second scenario that provides a high probability match to the mtDNA and Y chromosome patterns works for both directions of sex-bias. Human cultures often track clan membership generations into the past³², and in such a context the prestige associated with modern patriline or matriline can translate to increased fertility. In an expanding early human population impacted by hybridization by local archaic humans, it would not be surprising if individuals who traced their ancestry to early modern humans on the entirely male line (prestigious patriline) or female line (prestigious matriline) would have been more successful at competing for mates and therefore have a higher reproductive rate. My simulations show that if the fertility of modern patriline or matriline within modern groups was higher by roughly 9-14%, it would be sufficient to drive the observations (Figure 3b, Supplementary Information section 3).

The third scenario that can explain the data is two successful range expansions. In one, males are the sex that interbreeds across populations leading to retention of early modern human mtDNA. In the other that precedes or succeeds it, females interbreed across populations leading to retention of early modern human Y chromosomes (Figure 3b). Cultural adaptation in humans can be rapid, and two waves with different social structures, on the order of 10,000 years apart, seem plausible.

Figure 3: Range expansion scenarios that retain early modern mtDNA and Y with high probability.

(a) Example of an early modern human expansion into Europe in a “roving band” model with baseline demographic parameters and the Y retained due to a 16% higher reproductive rate for prestigious male patriline. The landscape is seeded by a single early modern human band of size 25 (the remaining 69 are archaic), and bands diffuse every generation and exchange male migrants mixing modern (red) and archaic (blue) lineages. Each square has three vertical bars showing the modern proportion for mtDNA (left), Y (middle), and autosomes (right). Next to each plot is a count of the percent modern genetic ancestry in bands with a modern affiliation at that generation; because only males switch bands, a band’s mtDNA bar is either fully modern or fully archaic. Modern ancestry at the Y and autosomes decreases because of archaic gene flow (but recovers at the Y due to advantage of prestigious male patriline). (b) Simulations of a wide range of parameter combinations for both the deme model (purple) and roving band model (orange) show four scenarios through which modern ancestry can be retained at both mtDNA and Y with high probability.



A fourth scenario that could explain the data is if first-generation offspring of the same sex as their archaic parent were not raised. There are ethnographic examples in forager societies today of offspring of different sexes being incorporated into the two parental populations differently depending on their sex^{33,34}, but I view this as the least anthropologically plausible of the four scenarios, as my simulations in Figure 3b show that the differential treatment of males and females would have to be more than a hundred-fold for this process to explain the data.

Genetic substructure in people today suggests that the range expansion also impacted Africa

I now highlight an independent genetic inference that has not previously been co-considered with the Neandertal evidence, and also points to a major range expansion ~300 ka. Multiple studies have inferred that the ancestors of all living humans including sub-Saharan Africans were deeply substructured into lineages that diverged more than a million years ago^{7,11,12,35,36,37}. Several studies using independent methods have now dated the mixture of these divergent lineages to form the ancestors of modern humans to approximately 300 ka^{7,11,12}, although a caveat is that the inference is not stable across populations plausibly reflecting multiple layers of mixture in Africa^{11,36}, and some studies have preferred models of persistent substructure³⁷ consistent with a view of human ancestors as a deeply structured metapopulation³⁸. Nevertheless, the coincidence of the ~300 ka inference over multiple studies is significant, as it increases the weight of evidence for the theory that an expanding early modern human lineage impacted both Europe and Africa around this time.

Taken together, these results—now corroborated by an independent study that analyzed DNA sequence motifs that are bound by the *PRDM9* gene determining where in the genome recombinations occur⁷—imply a view in which before ~300 ka, there were at least three major lineages that contributed to high coverage sequenced human genomes (Figure 1). The first

contributed to modern humans living today but not to Eurasian archaic humans (Denisovans and Neandertals), likely reflecting African archaic humans. The second contributed to Denisovans and Neandertals but not to modern Africans, likely representing Eurasian archaic humans. The third—showing evidence of activity of the “*PRDM9-A*” allele, which Denisovans lack—is the expanding early modern human lineage represented in admixed form in Neandertals and modern humans living today, but not in Denisovans: a successful early modern human population. In Europe, the expansion of this population led to large-scale incorporation of local archaic genes because barriers to producing viable offspring were few, giving rise to Neandertals who ultimately harbored 3-7% ancestry from the expanding early modern human lineage. In Africa, the mixture was with a more divergent lineage, which meant that there was less permeability of the expanding population to admixture with the local archaic group. Thus, the final population which became our own ancestors had a much higher proportion of ancestry from the expanding early modern human lineage¹¹.

A range expansion affecting Africa and West Eurasia correlates to archaeological evidence

Is the genetic evidence for major movements of people linking Neandertals in Europe and early modern humans in Africa ~300 ka, to the exclusion of Denisovan ancestors in East Asia, mirrored by archaeological data? Indeed it is. The genetically inferred expansion follows the time when a tool production technique (Levallois points) based on prepared stone cores spread across a wide area of Africa where it was used by early modern humans (known there as the Middle Stone Age and starting by at least 400 ka^{39,40}), and to the Middle East and Europe where it was used by Neandertals (known there as the Middle Paleolithic, starting by at least 480-300 ka⁴¹). There is evidence that these technologies were used much less intensively or not at all in East Asia where the closest cousins of early modern humans and Neandertals were Denisovans. In China, the sole strong claim of Levallois in the Middle Paleolithic is at the site of Guanyindong Cave dating to

~170-80 ka⁴², but this is not a confident inference as another study refutes this interpretation⁴³. Levallois has been documented in the Altai Mountains⁴⁴, but this region was proximate to West Eurasia and the technology could have spread from there through contact with Neandertal populations which are well-documented in the Altai. Similarly, evidence of Levallois in India⁴⁵ ~250-200 ka could have been spread by modern human populations from the Near East.

The theory that the spread of Levallois stone tools may be a material marker for early human migrations was suggested decades ago as the “Mode 3 Hypothesis,” with the authors writing: “Neanderthal and modern lineages share a common ancestor... 350,000 and 250,000 years ago rather than in the earlier Middle Pleistocene; this ancestral population, which developed mode 3 technology (Levallois/Middle Stone Age), dispersed across Africa and western Eurasia in a warmer period prior to independent evolution towards Neanderthals and modern humans.”⁴⁶ This proposal has been questioned based on evidence that prepared core techniques have been invented on multiple occasions including on the order of one million years ago in Africa⁴⁷, and reconstructions suggesting that Levallois itself may have been invented independently⁴⁸. Whether or not Levallois points were independently invented, the period preceding the ~300 ka range expansion also corresponds to at least four other shared material culture features that increase the plausibility of a spread of an early modern human lineage from a core region to broader Europe and Africa around this time: improved hafting, expansion in the use of controlled fire⁴⁹, use of ochre, and larger and more organized campsites⁵⁰. All of these developments became widespread both in West Eurasia and in Africa and were much less present or entirely absent in East Asia, providing independent archaeological evidence for connections. The archaeological record from this period is thin and the accuracy of date estimates too poor to determine if the timings of appearance of each these cultural traits were coincidental enough to point to a range expansion.

But taken together, these results increase the weight of evidence for the hypothesis¹⁰ that some of these shared material culture features may have spread through the demographic expansion of a successful early modern human group that has been inferred here entirely through genetic analysis.

Catalog of mutations that may contribute to distinctively early modern human biology

Previous studies have compiled lists of genetic variants present in nearly all living people, but absent in Neandertals and Denisovans, under the hypothesis that these are likely to be enriched in alleles that contributed to unique modern human innovations⁵¹. But the evidence presented here indicates that these catalogs are likely to screen out an important class of mutations: ones that evolved in the early modern human population prior to its bi-continental range expansion, and that may have supported behaviors shared between Neandertals and African modern humans like the intensive use of Levallois point tools, ochre, fire, and hafting; and the propensity to use larger and more organized campsites.

To search for candidate mutations that are recently shared between modern humans and Neandertals and absent in Denisovans, I analyzed the output of an Ancestral Recombination Graph inferred by the ARGweaver-D software⁴. This process identified 0.06% of chromosomes 1-22 where 16 African haplotypes and 4 Neandertal haplotypes are estimated to share a common ancestor more recently than 625 ka, while the 2 Denisovan haplotypes share an ancestor with them only more than 700 ka ago (Supplementary Information section 4). Altogether I identified 1,576 one kilobase windows that merge into 141 loci and in turn fall into 66 clusters (Extended Data Table 1). I estimate that 77% of these windows are genuine instances of shared ancestry of the sampled African and Neandertal haplotypes after their separation from the population ancestral to Denisovans, based on the excess over what I obtain when I repeat the experiment with the

Neandertal and Denisovan labels exchanged. Ten of these clusters are visualized in Extended Data Figure 5. One includes the genes *FADS1* and *FADS2*, where modern humans and Neandertals share a common ancestor around 544 ka. This locus has been under repeated selection in the last hundreds of thousands of years, allowing more efficient conversion of polyunsaturated fatty acids that typically come from plants⁵², and raising the possibility of selection for a more diversified diet during the early modern human expansion. A second overlaps *ARHGAP32* which modulates the structure of neural connections, and contains Glu1489Asp, one of six amino-acid-changing substitutions where nearly all sampled modern humans and Neandertals share an ancestor <700 ka while Denisovans are much more distantly related (Table 1 and Extended Data Table 2).

Table 1. Nonsynonymous substitutions where modern humans and Neandertals share a common ancestor <625 ka and Denisovans are more distantly related

chr	hg19 pos.	closest gene	substitution	
3	47,162,659	<i>SETD2</i>	His1156Arg	300
3	47,165,351	<i>SETD2</i>	Val259Ile	
3	182,994,744	<i>MCF2L2</i>	Asn593Ser	301
9	84,605,717	<i>SPATA31D1</i>	Leu111Arg	
11	128,840,599	<i>ARHGAP32</i>	Glu1489Asp	302
14	58,862,763	<i>TOMM20L</i>	Asp30Asn	303

I tested if the catalog is enriched for particular categories of gene function⁵³, thereby identifying nine groupings of genes (of 536 tested) largely associated with protein catabolism that showed significant enrichment even after correction for multiple hypothesis testing (False Discovery Rate q-values from 0.001 to 0.006) (Extended Data Table 3). A notable enrichment is seen for genes involved in proteasome-mediated ubiquitin-dependent protein catabolism, suggestive of viral interactions⁵⁴. Virus-interacting proteins account for an estimated 30% of adaptive amino acid substitutions in the mammalian-conserved part of the human proteome⁵⁵, and so are an intriguing candidate for being involved in early-modern-human-specific evolution in this period.

Discussion

The hypothesis of an early modern human range expansion ~300 ka that impacted Neandertal ancestors is not new to this study: multiple studies have provided evidence of early modern human gene flow into Neandertals^{2,3,4,5,6}, and since the gene flow occurred in a geographic context, it was almost certainly a range expansion. What is new to this study is the demonstration that the only plausible way to explain the data is the unique population genetic dynamics associated with a geographically explicit range expansion—large-scale incorporation of local genes into a population that maintained coherence from generation to generation as it expanded into a new territory—which means that the range expansion would have been a plausible vector for spreading culture even if the ultimate proportion of ancestry from the early modern human ancestors was small (~3-7%). In Europe, Neandertal ancestors formed as this expanding early modern human population incorporated a stream of local archaic people, retaining early modern human mtDNA and Y chromosomes while having most of the rest of their genome replaced by archaic genes. In Africa, the ancestors of living humans formed as this expanding population also incorporated local archaic genes. By establishing a vector of major movements connecting West Eurasia and Africa, these results increase the weight of evidence that at least some of the distinctive shared archaeological patterns between West Eurasia and Africa—including use of Levallois points, controlled use of fire, use of ochre, improved hafting, and larger and more organized campsites—spread by the bi-continental, large-scale early modern human range expansion hypothesized here.

Materials and Methods

Details of the methods used in this paper are presented in the Supplementary Information.

Acknowledgments

This research was inspired by a presentation Janet Kelso gave at the Boston Evolutionary Genomics Supergroup in September 2022, in which she proposed the puzzle addressed here. The work benefited from comments either in-person or through public forums by Bridget Alex, Mira Bernstein, Amy Clark, Trevor Cousins, Rob Foley, David Gokhman, John Hawks, Alexander Kim, Hrushikesh Loya, Dan Lieberman, Iain Mathieson, Simon Myers, Svante Pääbo, Nick Patterson, Molly Schumer, Ashwin Sivakumar, Pontus Skoglund, Daniel Tabin, Shamam Waldman, Tianyi Wang, and Richard Wrangham. I am grateful to Melissa Hubisz and Adam Siepel for sharing the output of their ancestral recombination graph reconstruction of relationships among 8 Africans, 2 Neandertals, and a Denisovan, and to Leo Speidel for sharing his reconstruction for individuals from the 1000 Genomes project. I acknowledge support from the National Institutes of Health (grants HG012287 and HG013987), the Howard Hughes Medical Institute.

Use of artificial intelligence tools

Many of the analyses in this study are the product of a dialogue with artificial intelligence tools (Claude Sonnet, Opus, and Fable models, sometimes implemented within Cursor Integrated Development Environment). The AI tools wrote the code and generated figures using Python, and I checked the robustness of the results. For the supplementary notes but not for the main manuscript, some of the text was composed by the AI and I edited it.

Conflict of Interest Statement

356 I declare no competing interests.

357

358 **Data Availability Statement**

359 The quantitative simulation results are available as Supplementary Table 1, and the results of the
360 scan for loci and variants that may have provided advantages to early modern humans are available
361 as Supplementary Tables 2 and 3.

362

363 **Code Availability Statement**

364 The scripts used to generate the analyses and figures are available as four zipped files, one for each
365 Supplementary Information section.

Supplementary Information Guide

Supplementary Information sections

(SI 1) Biological constraints mean that in the absence of a range expansion, fixation of early modern human lineage mtDNA and Y chromosomes in Neandertals is unlikely

(SI 2) No evidence for a Dobzhansky-Muller Incompatibility between Denisovan mtDNA and modern human nuclear DNA

(SI 3) Modern human range expansions predict fixation of modern mtDNA and Y in Neandertals

(SI 4) A catalog of mutations that may have been important in uniquely modern human biology

Analysis Scripts (separate zipped folder with a README for each SI)

(1) SI1codebase.zip

(2) SI2codebase.zip

(3) SI3codebase.zip

(4) SI4codebase.zip

Supplementary Video

(Supplementary Video) Animated version of Figure 3a

Supplementary Tables

(Supplementary Table 1) Simulation results: Excel file with 15 tabs providing all the simulation results and details on the simulation setup

(Supplementary Table 2) gzipped tsv consisting of 2.44 million rows and the results of the scan for loci that may have provided advantages to early modern humans

(Supplementary Table 3) tsv with 1,223 rows: sites in candidate introgressed regions that are high-frequency-derived in Neandertals and modern humans and absent in Denisovans

Extended Data Table 1: Catalog of candidate loci for early modern human adaptation. 141 runs of 1 kb autosomal grid points (points within 5 kb of one another merged into a single locus) where 16 sampled modern human and 4 Neandertal haplotypes derive from a common ancestor ≤ 625 ka while 2 Denisovan haplotypes diverge ≥ 700 ka based on parsing of an ARGweaver-D ancestral recombination graph (ARG). I pool loci within 100 kb to form 66 clusters (gray striping). Each TMRCAs is the posterior median with its 90% central interval. “P(<700 ka)” is the posterior probability that the modern-plus-Neandertal coalescence is below the 700 ka threshold at the gridpoint within the locus where this quantity is highest. “Relate percentile” is the inferred time to the most recent common ancestor compared to the autosome-wide distribution in a larger panel of 216 West African Yoruba haplotypes. “P(real)” from the 1 kb grid point in each locus showing the strongest posterior signal of enrichment.

chr	interval (Mb)	span (kb)	P(Real)	T _{Afr} , ka (90% CI) (Argweaver)	T _{Afr} percentile (Relate)	T _{Afr+Nea} , ka (90% CI) (Argweaver)	P(T _{Afr+Nea} <700 ka) (Argweaver)	T _{Afr+Den} , ka (90% CI) (Argweaver)	genes within 5 kb
1	16.804-16.806	3	0.73	521 (271-925)	9	596 (431-1125)	0.80	1400 (813-2437)	<i>CROCCP3</i> (non-coding)
1	16.815	1	0.73	550 (424-804)	4	575 (428-987)	0.80	1400 (813-2437)	<i>CROCCP3</i> (non-coding)
1	35.762-35.775	14	0.77	565 (434-874)	3	584 (464-909)	0.88	1479 (983-2416)	<i>ZMYM4</i> (+2 non-coding)
1	35.821-35.824	4	0.67	600 (558-684)	3	600 (558-684)	0.97	1500 (986-2674)	<i>ZMYM4</i> (+2 non-coding)
1	35.831-35.835	5	0.67	603 (558-690)	3	603 (558-690)	0.97	1535 (960-3049)	<i>ZMYM4</i> (+1 non-coding)
1	35.859	1	0.67	550 (424-804)	4	600 (516-836)	0.84	1400 (980-1879)	<i>ZMYM4</i>
1	155.212-155.215	4	0.56	589 (430-1091)	8	589 (430-1091)	0.80	862 (629-1167)	<i>GBA</i> , <i>FAM189B</i>
2	22.123-22.124	2	0.40	538 (276-1047)	13	600 (432-1171)	0.69	1981 (1541-3668)	<i>AC067959.1</i> (non-coding)
2	81.377-81.415	39	0.81	543 (388-841)	90	555 (413-874)	0.99	895 (644-1221)	none within 5 kb; <i>AC105342.1</i> at 6 kb
2	107.414	1	0.40	550 (424-804)	10	600 (516-836)	0.84	1500 (996-2518)	<i>ST6GAL2</i>
2	162.102-162.110	9	0.74	526 (273-966)	4	581 (429-1033)	0.90	1172 (654-1837)	<i>AC009299.2</i> , <i>AC009299.3</i> (non-coding)
2	162.117-162.127	11	0.66	494 (262-741)	29	598 (469-989)	0.85	1236 (754-1826)	none within 5 kb; <i>AC009299.2</i> at 6 kb
2	172.735-172.738	4	0.43	527 (273-967)	3	606 (475-1027)	0.74	1488 (932-2717)	<i>SLC25A12</i>
2	195.306	1	0.48	550 (280-1139)	51	575 (372-1155)	0.74	1600 (1180-2598)	none within 5 kb; <i>AC106883.1</i> at 13 kb
2	195.330	1	0.46	556 (282-1185)	99	581 (374-1201)	0.72	7563 (3850-12545)	none within 5 kb; <i>AC106883.1</i> at 10 kb
2	195.339-195.344	6	0.74	521 (271-925)	97	560 (367-1040)	0.90	1413 (884-2345)	none within 5 kb; <i>AC106883.1</i> at 19 kb
2	230.808	1	0.55	556 (425-850)	11	613 (517-882)	0.79	825 (720-930)	<i>FBXO36</i> (+1 non-coding)
2	231.981-231.995	15	0.59	599 (474-996)	9	599 (474-996)	0.85	1429 (975-2106)	<i>PSMD1</i> , <i>HTR2B</i> (+1 non-coding)
3	27.164-27.165	2	0.77	566 (427-918)	68	566 (427-918)	0.88	1700 (1196-3518)	<i>NEK10</i>
3	27.180-27.200	21	0.79	571 (490-725)	36	573 (492-729)	0.97	2308 (1704-4662)	<i>NEK10</i>
3	40.959-40.969	11	0.79	569 (427-947)	72	580 (446-953)	0.90	1631 (1114-3201)	none within 5 kb; <i>RP11-520A21.1</i> at 43 kb
3	47.157-47.174	18	0.79	557 (413-888)	2	564 (428-899)	0.91	1147 (664-1768)	<i>SETD2</i>
3	47.180-47.212	33	0.83	550 (364-973)	5	569 (370-1104)	0.92	1233 (733-1906)	<i>SETD2</i> (+1 non-coding)
3	47.218	1	0.61	581 (374-1201)	1	613 (382-1385)	0.63	1200 (656-1831)	<i>KIF9-AS1</i> (non-coding)
3	47.224-47.229	6	0.66	600 (380-1339)	1	600 (380-1339)	0.65	1250 (660-2052)	<i>KIF9-AS1</i> (non-coding)
3	47.236-47.266	31	0.81	531 (284-982)	1	563 (368-1063)	0.90	1212 (696-1884)	<i>KIF9</i> (+1 non-coding)
3	47.272-47.295	24	0.81	520 (271-920)	0	564 (369-1074)	0.90	1236 (753-1781)	<i>KIF9</i> (+2 non-coding)
3	47.305-47.315	11	0.71	557 (358-1008)	4	592 (431-1110)	0.80	1219 (702-1807)	<i>KIF9</i>
3	47.326-47.353	28	0.87	508 (267-839)	2	544 (362-935)	0.97	1153 (691-1686)	<i>KIF9</i> , <i>KLHL18</i>
3	47.441-47.478	38	0.79	465 (253-654)	3	577 (428-998)	0.90	1500 (1004-2584)	<i>PTPN23</i> , <i>SCAP</i>
3	47.522-47.531	10	0.69	477 (257-700)	0	602 (485-962)	0.84	1496 (877-3074)	<i>SCAP</i>
3	47.550-47.554	5	0.73	561 (426-886)	0	589 (472-912)	0.88	1770 (1185-3376)	<i>ELP6</i>
3	83.244-83.351	108	0.86	285 (216-343)	4	557 (437-794)	0.97	1212 (810-1892)	none within 5 kb; <i>CYP51P1</i> at 387 kb
3	150.865-150.889	25	0.83	508 (267-853)	10	566 (427-932)	0.97	1607 (1075-3084)	<i>MED12L</i>
3	182.984-182.992	9	0.83	558 (412-902)	96	558 (412-902)	0.92	1828 (1213-3002)	<i>MCF2L2</i> , <i>B3GNT5</i>
3	183.003-183.004	2	0.67	600 (516-836)	99	600 (516-836)	0.84	4000 (3161-11973)	<i>MCF2L2</i> , <i>B3GNT5</i>
4	44.455-44.460	6	0.62	539 (361-887)	1	539 (361-887)	0.90	1608 (1126-2975)	<i>KCTD8</i> (+1 non-coding)
4	54.153-54.157	5	0.64	580 (479-823)	2	580 (479-823)	0.88	785 (623-1024)	<i>SCFD2</i>
4	85.821-85.854	34	0.69	498 (264-762)	1	587 (483-859)	0.88	1395 (910-2135)	<i>WDFY3</i>
4	85.864	1	0.40	506 (266-817)	2	600 (474-1004)	0.74	1500 (996-2518)	<i>WDFY3</i>
4	85.892-85.909	18	0.78	486 (277-730)	4	554 (366-999)	0.92	1137 (651-1716)	<i>WDFY3</i> (+1 non-coding)

4	85.918-85.928	11	0.77	494 (262-750)	6	541 (362-909)	0.92	1509 (956-2921)	<i>WDFY3-AS2</i> (non-coding)
4	139.977-139.998	22	0.67	370 (230-473)	15	561 (435-849)	0.96	1431 (894-2500)	<i>ELF2</i> (+1 non-coding)
4	168.893	1	0.35	613 (517-882)	21	613 (517-882)	0.79	1500 (996-2518)	<i>RP11-31019.1</i> (non-coding)
5	23.548-23.555	8	0.46	597 (474-984)	51	597 (474-984)	0.85	789 (623-1032)	none within 5 kb; <i>PRDM9</i> at 19 kb
6	27.503-27.509	7	0.40	533 (295-966)	41	602 (474-1010)	0.74	1589 (1107-2859)	none within 5 kb; <i>HNRNPA1P1</i> at 11 kb
6	27.530-27.544	15	0.64	569 (464-799)	33	569 (464-799)	0.96	1419 (952-2130)	<i>XXbac-BPG34I8.3</i> (non-coding)
6	27.560-27.604	45	0.76	542 (434-712)	28	555 (455-738)	0.97	1350 (831-2101)	<i>RNU6-471P, TRNAI6</i> (non-coding)
6	27.639	1	0.40	600 (474-1004)	42	600 (474-1004)	0.74	1500 (955-2727)	none within 5 kb; <i>RPI-15D7.1</i> at 18 kb
6	27.754-27.755	2	0.69	597 (378-1293)	16	597 (378-1293)	0.72	1012 (641-1443)	none within 5 kb; <i>RSL24D1P1</i> at 5 kb
6	42.771-42.810	40	0.83	525 (330-873)	19	534 (359-879)	1.00	1346 (774-2360)	<i>GLTSCR1L</i>
6	83.303-83.332	30	0.64	531 (369-788)	1	563 (454-805)	0.97	1833 (1306-3334)	none within 5 kb; <i>TPBG</i> at 222 kb
6	83.338-83.342	5	0.40	600 (474-1004)	3	600 (474-1004)	0.74	2300 (1636-5156)	none within 5 kb; <i>TPBG</i> at 257 kb
6	83.348-83.355	8	0.48	596 (473-964)	3	596 (473-964)	0.88	1607 (1098-3083)	none within 5 kb; <i>UBE3D</i> at 247 kb
7	97.737-97.740	4	0.82	553 (394-910)	42	572 (398-1048)	0.90	1450 (821-2897)	<i>LMTK2</i>
7	109.185-109.197	13	0.79	559 (420-881)	3	559 (420-881)	0.90	1798 (1336-3413)	<i>AC073071.1</i> (non-coding)
7	142.876	1	0.54	613 (433-1217)	51	613 (433-1217)	0.66	1000 (640-1420)	<i>TAS2R39</i>
7	142.892-142.895	4	0.64	559 (315-1145)	47	575 (372-1155)	0.74	947 (636-1322)	none within 5 kb; <i>TAS2R39</i> at 10 kb
8	81.158	1	0.48	550 (280-1139)	24	575 (372-1155)	0.74	1575 (967-3417)	<i>MIR5708</i> (non-coding)
8	85.006-85.007	2	0.65	550 (364-971)	95	600 (474-1004)	0.74	1062 (645-1535)	<i>RP11-120121.2</i> (non-coding)
8	86.117	1	0.65	601 (464-1046)	26	601 (464-1046)	0.73	1100 (764-1436)	<i>E2F5</i>
8	86.304-86.321	18	0.70	597 (515-808)	27	597 (515-808)	0.97	1050 (666-1588)	<i>CA3</i>
9	84.505-84.524	20	0.76	580 (473-854)	49	581 (476-854)	0.97	916 (633-1268)	<i>RNA5SP287, SNORD95</i> & 1 other
9	84.565-84.631	67	0.98	349 (250-495)	10	384 (320-502)	1.00	1007 (641-1486)	<i>SPATA31D1</i> (+3 non-coding)
9	84.641-84.642	2	0.96	363 (310-405)	2	363 (310-405)	1.00	787 (623-1029)	none within 5 kb; <i>SPATA31B1</i> at 31 kb
9	84.659-84.679	21	0.75	536 (360-887)	83	536 (360-887)	0.97	832 (627-1111)	<i>SPATA31B1</i> (non-coding)
9	84.687-84.708	22	0.71	552 (369-980)	11	552 (369-980)	0.97	852 (628-1147)	<i>DDX10P2</i> (non-coding)
9	84.726-84.727	2	0.71	503 (349-663)	20	503 (349-663)	0.97	725 (618-914)	none within 5 kb; <i>RP11-388B24.4</i> at 19 kb
9	84.737-84.747	11	0.62	577 (378-1150)	11	577 (378-1150)	0.84	897 (632-1229)	<i>RP11-388B24.4</i> (non-coding)
9	84.780-84.784	5	0.60	564 (386-1014)	2	579 (424-1023)	0.80	820 (626-1089)	none within 5 kb; <i>RP11-388B24.4</i> at 33 kb
9	84.792-84.799	8	0.68	529 (357-821)	18	543 (396-828)	0.97	770 (622-997)	none within 5 kb; <i>RP11-388B24.4</i> at 45 kb
9	84.813-84.814	2	0.57	538 (360-879)	22	563 (426-895)	0.90	762 (621-983)	none within 5 kb; <i>RP11-388B24.4</i> at 66 kb
9	84.820	1	0.53	575 (372-1155)	15	600 (422-1213)	0.68	900 (632-1236)	none within 5 kb; <i>RP11-15B24.5</i> at 68 kb
9	84.899-84.918	20	0.73	575 (470-820)	60	586 (472-892)	0.88	1697 (1196-3229)	<i>RP11-15B24.5</i> (non-coding)
9	135.801	1	0.40	601 (506-878)	29	601 (506-878)	0.81	725 (618-914)	<i>TSC1</i>
10	4.737-4.741	5	0.59	555 (414-882)	49	558 (415-901)	0.91	1535 (1052-2588)	none within 5 kb; <i>LINC00704</i> at 17 kb
10	109.119-109.132	14	0.79	537 (276-1048)	97	564 (369-1072)	0.92	1021 (662-1426)	none within 5 kb; <i>RNA5SP326</i> at 89 kb
10	109.251-109.254	4	0.57	563 (436-812)	25	600 (516-836)	0.84	800 (716-884)	none within 5 kb; <i>RNA5SP326</i> at 30 kb
11	1.020-1.023	4	0.77	569 (435-910)	29	569 (435-910)	0.90	1325 (842-1861)	<i>MUC6</i>
11	14.555	1	0.58	556 (282-1185)	53	600 (380-1339)	0.65	1000 (640-1420)	<i>PSMA1</i>
11	14.566-14.573	8	0.69	577 (359-1167)	1	602 (425-1182)	0.76	1378 (836-2328)	<i>PSMA1</i>
11	14.579-14.587	9	0.73	585 (470-891)	1	585 (470-891)	0.88	1378 (898-2018)	<i>PSMA1</i>
11	57.372-57.381	10	0.67	588 (440-995)	37	602 (474-1011)	0.74	1533 (982-3145)	<i>SERPING1</i>
11	61.055-61.071	17	0.58	581 (468-869)	75	581 (468-869)	0.97	792 (624-1038)	<i>VWCE, DDB1</i>
11	61.541-61.548	8	0.71	481 (258-718)	11	588 (438-1048)	0.90	1156 (653-1774)	<i>MYRF, TMEM258</i>
11	61.554-61.597	44	0.87	436 (253-620)	1	544 (399-835)	0.97	1583 (1064-2948)	<i>MYRF, TMEM258, FEN1, FADS2, FADS1</i>
11	66.822-66.857	36	0.68	554 (380-961)	9	555 (380-967)	0.92	819 (626-1087)	<i>SYT12, RHOD</i>
11	66.870-66.907	38	0.73	567 (427-934)	3	567 (427-934)	0.97	891 (631-1220)	<i>KDM2A</i>
11	66.915-66.920	6	0.54	595 (431-1125)	5	595 (431-1125)	0.80	875 (630-1190)	<i>KDM2A</i>
11	66.928-66.968	41	0.81	544 (423-775)	4	544 (423-775)	0.97	1020 (642-1464)	<i>KDM2A, AP001885.1</i>
11	66.974-66.987	14	0.67	538 (422-726)	4	538 (422-726)	0.97	758 (621-975)	<i>KDM2A</i> (+1 non-coding)
11	66.996-67.009	14	0.69	571 (449-877)	1	571 (449-877)	0.90	890 (631-1217)	<i>KDM2A</i>
11	67.022	1	0.49	613 (517-882)	35	613 (517-882)	0.79	900 (632-1236)	<i>KDM2A</i>
11	67.054	1	0.62	613 (517-882)	26	613 (517-882)	0.79	1300 (796-1855)	<i>ADRBK1, ANKRD13D</i>

11	128.831-128.850	20	0.87	490 (261-743)	6	524 (356-805)	0.97	1538 (1023-2764)	<i>ARHGAP32</i>
12	3.767	1	0.64	600 (432-1171)	58	600 (432-1171)	0.69	1175 (734-1616)	<i>EFCAB4B</i>
12	53.852-53.857	6	0.55	565 (426-911)	11	591 (469-949)	0.88	1442 (847-2681)	<i>PCBP2</i> (+1 non-coding)
12	74.343-74.370	28	0.59	499 (316-695)	2	565 (476-734)	0.97	1863 (1363-3337)	none within 5 kb; <i>RP11-711C17.1</i> at 30 kb
12	85.564-85.581	18	0.46	594 (480-927)	11	597 (486-929)	0.85	1826 (1351-3585)	<i>LRR1Q1</i>
12	85.612-85.618	7	0.59	577 (470-833)	45	601 (513-848)	0.85	1525 (1024-2658)	<i>LRR1Q1</i>
12	85.630-85.668	39	0.65	582 (490-805)	7	584 (492-806)	0.97	1574 (1039-2702)	<i>LRR1Q1</i> (+1 non-coding)
13	77.631-77.640	10	0.67	580 (477-833)	4	603 (516-847)	0.84	1609 (1198-2614)	<i>MYCBP2</i>
13	77.729-77.741	13	0.73	598 (537-744)	7	598 (537-744)	0.97	1703 (1240-3126)	<i>MYCBP2</i>
13	77.748	1	0.67	569 (501-674)	20	600 (558-684)	0.97	1425 (817-2667)	<i>MYCBP2</i>
13	77.761-77.765	5	0.72	484 (259-687)	30	595 (515-801)	0.95	1544 (1161-2249)	<i>MYCBP2</i>
13	77.783-77.788	6	0.67	600 (516-836)	43	604 (516-851)	0.84	1696 (1357-2675)	<i>MYCBP2</i>
14	21.360-21.362	3	0.65	600 (516-836)	46	600 (516-836)	0.84	1054 (645-1534)	<i>RNASE3</i> (+1 non-coding)
14	21.379-21.386	8	0.75	566 (419-924)	43	566 (419-924)	0.97	1048 (644-1516)	<i>RP11-219E7.4</i> , <i>RP11-84C10.3</i> & 1 other
14	21.392-21.407	16	0.87	415 (246-534)	5	498 (348-687)	1.00	877 (630-1193)	<i>RP11-219E7.4</i> , <i>RP11-84C10.3</i> & 1 other
14	58.701-58.702	2	0.62	516 (269-886)	9	606 (475-1027)	0.74	1162 (653-1752)	<i>C14orf37</i> , <i>ACTR10</i>
14	58.711-58.726	16	0.59	362 (234-500)	1	591 (443-1043)	0.88	1467 (865-2939)	<i>C14orf37</i> , <i>PSMA3</i>
14	58.743-58.746	4	0.65	200 (200-200)	1	578 (429-1010)	0.80	1300 (796-1855)	<i>C14orf37</i> , <i>PSMA3</i> (+1 non-coding)
14	58.757-58.758	2	0.59	231 (205-257)	2	600 (432-1171)	0.69	1300 (796-1855)	<i>C14orf37</i> , <i>AL132989.1</i> (+3 non-coding)
14	58.773-58.802	30	0.66	375 (235-488)	9	560 (414-904)	0.92	1596 (1064-3114)	<i>ARID4A</i>
14	58.813	1	0.46	506 (266-817)	0	581 (419-1075)	0.76	1700 (1196-3518)	<i>ARID4A</i>
14	58.826-58.834	9	0.64	541 (416-779)	2	545 (423-781)	0.97	2181 (1593-3288)	<i>ARID4A</i> (+1 non-coding)
14	58.845-58.849	5	0.64	469 (254-627)	1	553 (425-827)	0.90	1550 (1005-2977)	<i>ARID4A</i> (+1 non-coding)
14	58.857-58.862	6	0.83	451 (249-578)	0	487 (344-613)	1.00	1392 (803-2563)	<i>TOMM20L</i> (+2 non-coding)
14	58.872-58.892	21	0.71	462 (252-614)	10	513 (352-727)	0.97	1666 (1160-3287)	<i>TOMM20L</i> , <i>TIMM9</i> , <i>KIAA0586</i>
14	102.184-102.185	2	0.65	588 (420-1121)	12	588 (420-1121)	0.78	1300 (796-1855)	<i>RP11-1029J19.5</i> (non-coding)
15	72.713-72.720	8	0.78	536 (276-1036)	30	581 (374-1201)	0.84	1256 (756-1861)	<i>RP11-1007O24.3</i> (non-coding)
15	72.763-72.769	7	0.73	580 (471-852)	9	580 (471-852)	0.88	1236 (699-1952)	<i>ARIH1</i> (+2 non-coding)
15	72.797-72.804	8	0.79	508 (351-694)	6	586 (514-747)	0.97	1388 (889-2135)	<i>ARIH1</i>
15	72.859-72.867	9	0.72	588 (491-838)	1	588 (491-838)	0.88	1128 (769-1494)	<i>ARIH1</i>
15	74.952-74.962	11	0.55	581 (429-1025)	22	588 (441-1029)	0.80	830 (626-1106)	<i>EDC3</i>
15	74.972-74.975	4	0.54	586 (482-847)	27	586 (482-847)	0.88	806 (625-1063)	<i>EDC3</i>
15	74.983	1	0.44	613 (475-1050)	30	613 (475-1050)	0.71	825 (626-1098)	<i>EDC3</i>
15	75.738-75.805	68	0.86	515 (272-875)	3	550 (368-963)	0.97	1041 (643-1517)	<i>SIN3A</i> , <i>PTPN9</i> (+2 non-coding)
15	78.310-78.311	2	0.67	600 (558-684)	13	600 (558-684)	0.97	1313 (882-1773)	<i>TBC1D2B</i>
17	62.513-62.532	20	0.80	562 (368-1035)	0	574 (405-1043)	0.91	1516 (1011-2695)	<i>CEP95</i> (+1 non-coding)
17	62.610-62.634	25	0.87	567 (386-1041)	50	567 (386-1041)	0.97	1353 (803-2192)	<i>SMURF2</i>
17	67.847	1	0.67	575 (428-987)	8	600 (474-1004)	0.74	1400 (813-2437)	<i>AC003051.1</i> (non-coding)
17	74.063-74.070	8	0.83	545 (385-864)	4	556 (390-933)	0.92	1563 (1001-3119)	<i>SRP68</i> , <i>ZACN</i> , <i>GALR2</i>
18	42.233-42.253	21	0.62	603 (558-690)	96	603 (558-690)	0.97	1697 (1189-2999)	<i>RP11-456K23.1</i> (non-coding)
18	76.991	1	0.40	575 (428-987)	21	600 (474-1004)	0.74	1700 (1196-3518)	<i>ATP9B</i>
18	76.998-77.004	7	0.52	602 (474-1010)	83	602 (474-1010)	0.74	1514 (1023-2610)	<i>ATP9B</i>
18	77.034-77.046	13	0.65	541 (383-827)	15	555 (425-835)	0.90	1583 (1068-3009)	<i>ATP9B</i>
18	77.053-77.081	29	0.66	579 (468-855)	31	579 (468-855)	0.89	1538 (1055-2613)	<i>ATP9B</i>
20	32.635-32.653	19	0.76	532 (358-845)	1	545 (373-895)	0.92	1317 (795-2147)	<i>RALY</i> (+1 non-coding)
20	46.186-46.198	13	0.65	600 (474-1004)	18	602 (474-1009)	0.74	1031 (643-1488)	<i>NCOA3</i> (+1 non-coding)
22	29.723-29.726	4	0.63	575 (384-1124)	1	600 (432-1171)	0.69	906 (633-1247)	<i>APIB1</i> (+1 non-coding)

Extended Data Table 2. Candidate variants of high predicted functional significance. The 6 amino-acid-changing substitutions in Table 1 of the main text are a subset of the 9 variants here. CADD measures predicted deleteriousness by combining functional annotations into one score on the PHRED scale of $-10 \times \log_{10}(\text{percentile})$, for this table running from 10.9 (top 10%) through 20.2 (top 1%). Haplotype columns give derived over total count for 5,008 modern human haplotypes, up to 10 haplotypes from high-coverage Neandertals, and up to 4 haplotypes from high-coverage Denisovans. Gene column gives the (nearest) gene body. P(real) is evaluated at the substitution itself rather than across the locus around it, and is 0.00 at *MCF2L2* and *TOMM20L* because the modern-plus-Neandertal coalescence inferred at those exact positions is older than the 625 ka catalog ceiling even though the surrounding locus qualifies; both loci are at P(real) = 0.83 in Extended Data Table 1.

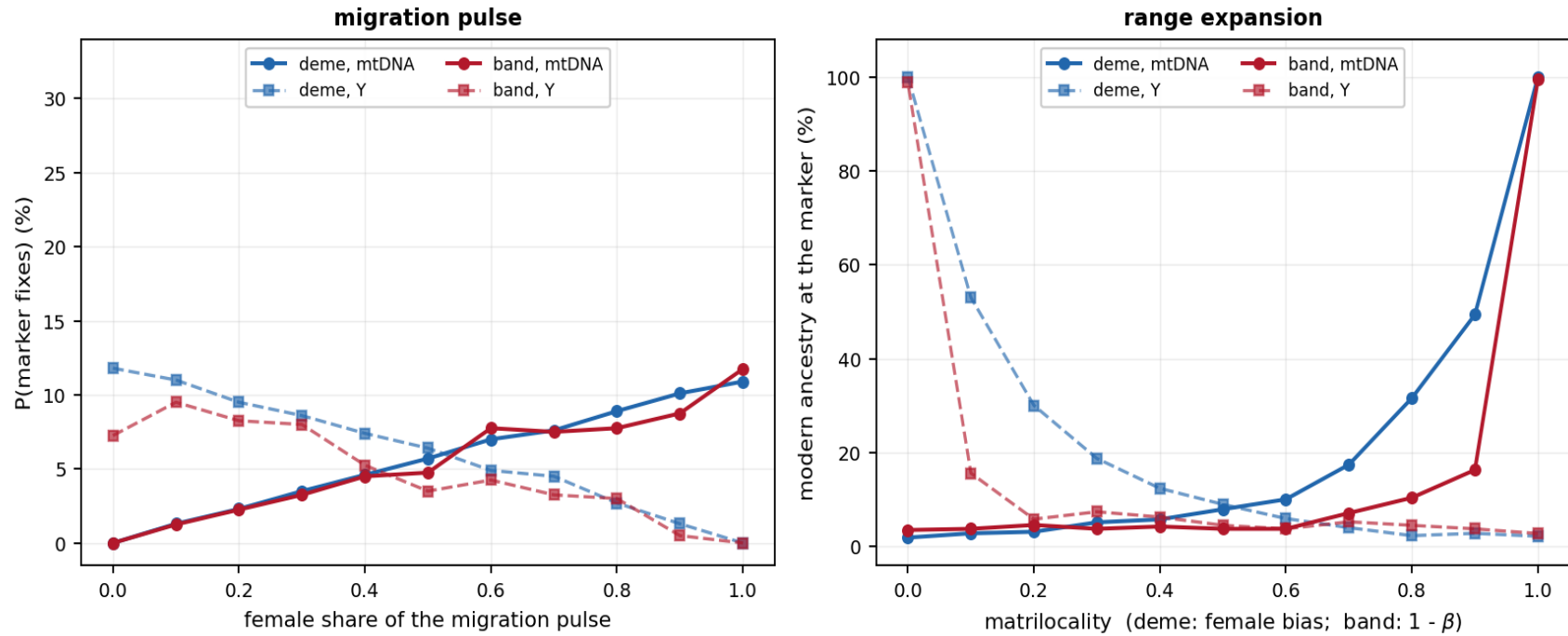
chr	position (hg19)	gene	functional evidence (6 protein-coding changes highlighted)	CADD	P(real)	moderns	Nea	Den
2	162,101,812	<i>TANK</i> (9 kb)	conserved, DNase cluster, CADD top 1%	20.2	0.71	5008/5008	10/10	0/4
3	47,162,659	<i>SETD2</i>	His1156Arg	19.2	0.76	5008/5008	10/10	0/4
3	47,165,351	<i>SETD2</i>	Val259Ile	16.7	0.79	5008/5008	10/10	0/4
3	182,994,744	<i>MCF2L2</i>	Asn593Ser, splice region, conserved element	16.2	0.00	4872/5008	10/10	0/4
6	42,788,926	<i>GLTSCR1L</i>	5' untranslated, conserved, DNase cluster	14.7	0.75	4970/5008	10/10	0/2
9	84,605,717	<i>SPATA31D1</i>	Leu111Arg	10.9	0.97	5000/5008	10/10	0/2
11	67,053,028	<i>ADRBK1</i>	3' UTR, conserved, TFBS cluster	15.8	0.52	5008/5008	10/10	0/4
11	128,840,599	<i>ARHGAP32</i>	Glu1489Asp, 3' UTR, conserved, DNase cluster	18.3	0.87	5008/5008	10/10	0/4
14	58,862,763	<i>TOMM20L</i>	Asp30Asn, conserved, DNase & TFBS clusters	16.7	0.00	5008/5008	8/8	0/2

Extended Data Table 3. Gene set most strongly enriched for recent modern human - Neandertal ancestors

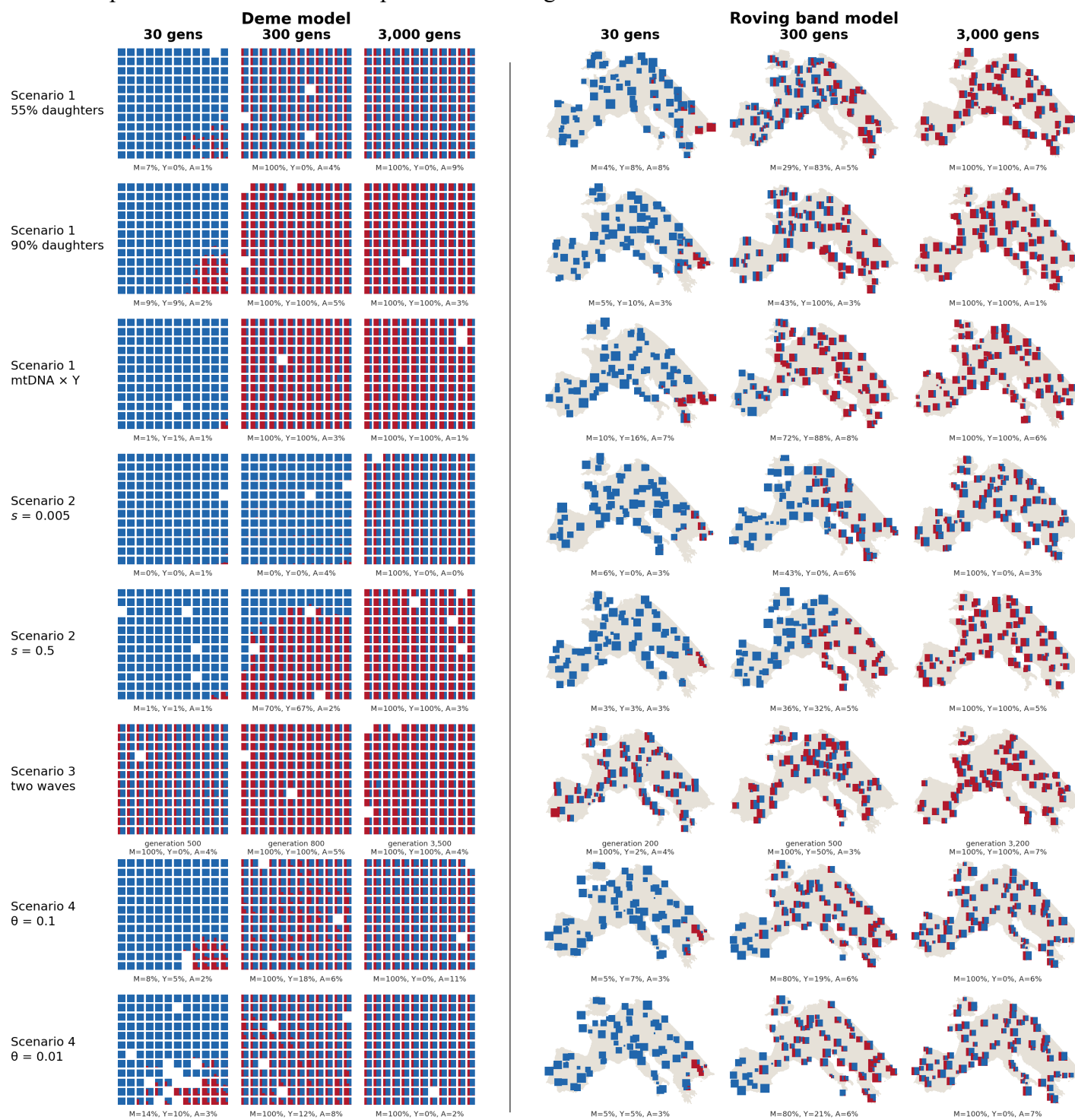
Gene set (“Gene Ontology” biological process category)	genes	catalog	expected	p-value (nominal)	q-value (FDR)	p-value (10⁵ perm.)
macromolecule catabolic process	960	22	7.3	2.5×10^{-6}	0.0010	$< 1 \times 10^{-5}$
response to type I interferon	85	7	0.6	3.6×10^{-6}	0.0010	$< 1 \times 10^{-5}$
proteasome-mediated ubiquitin-dependent catabolic process	397	13	3.0	9.5×10^{-6}	0.0016	$< 1 \times 10^{-5}$
ubiquitin-dependent protein catabolic process	541	15	4.1	1.4×10^{-5}	0.0016	$< 1 \times 10^{-5}$
modification-dependent protein catabolic process	551	15	4.2	1.7×10^{-5}	0.0016	$< 1 \times 10^{-5}$
modification-dependent macromolecule catabolic process	553	15	4.2	1.8×10^{-5}	0.0016	$< 1 \times 10^{-5}$
cellular response to type I interferon	78	6	0.6	2.7×10^{-5}	0.0021	4×10^{-5}
proteasomal protein catabolic process	477	13	3.6	6.4×10^{-5}	0.0043	2×10^{-5}
protein catabolic process	722	16	5.5	1.0×10^{-4}	0.0062	5×10^{-5}

Notes: “genes”: size of the gene set (of a total of 14,515 analyzable). “catalog”: the number of genes within the 141 loci whose regulatory domain overlaps the gene set. “expected”: expectation for the catalog count from a hypergeometric distribution. “p-value (nominal)”: hypergeometric test over genes for enrichment. “q-value (FDR)”: Benjamini-Hochberg false discovery rate across all 536 categories with 10-1,000 annotated genes and at least two catalog genes. “p-value (10⁵ perm.)”: a test that relocates all 66 clusters to blocks of equal length and similar mean B-statistic, and reports the fraction of permutation replicates with the observed enrichment. The nine significant categories are nested descriptions of protein catabolism and two of type I interferon response with overlapping genes, so these are correlated findings.

Extended Data Figure 1: Sex-bias drives one uniparentally inherited marker to fixation in a range expansion, but not in a non-geographically-explicit model. 1,000 replicates per point. *Left: Non-geographically-explicit.* 5% gene flow from early modern humans into archaic humans is followed by drift. Even under complete sex bias, fixation is unlikely (~10%). *Right: Range expansion.* As the probability that the child of a cross-lineage pairing joins its mother's group nears 1, fixation of the corresponding marker becomes likely. This does not explain the pattern at the other uniparentally transmitted marker, where biological or cultural selection is needed.

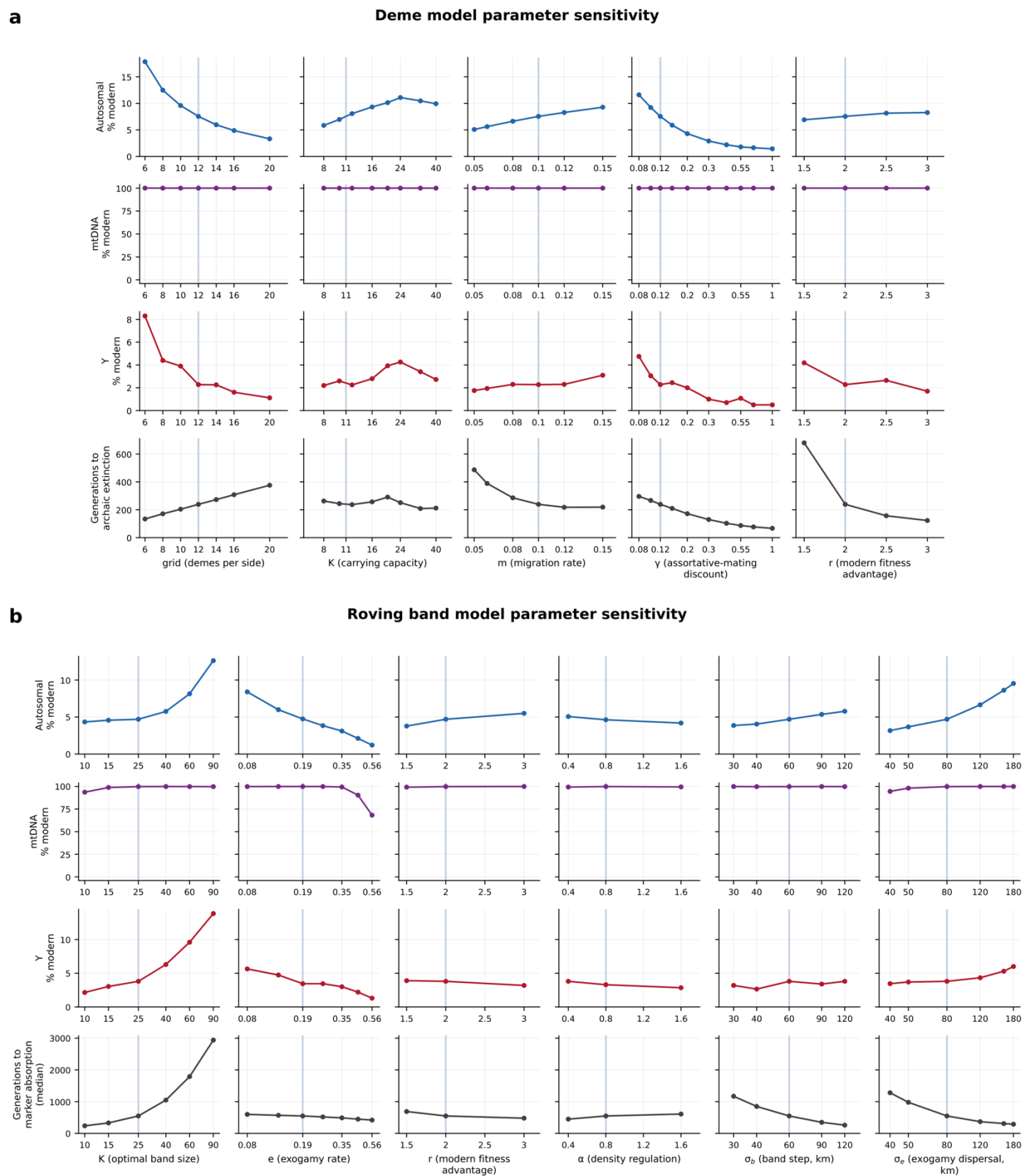


Extended Data Figure 2. Robustness across both simulation frameworks. Each row is an example of one mechanism at one strength, drawn on the left in the deme framework and on the right in the roving band framework, at 30, 300 and 3,000 generations after contact. Each square is one deme or one band, with three contiguous bars giving the modern fraction at mtDNA, the Y and the autosomes from left to right. The percentages below each panel are the whole-landscape values at that generation for the modern human subset of individuals.



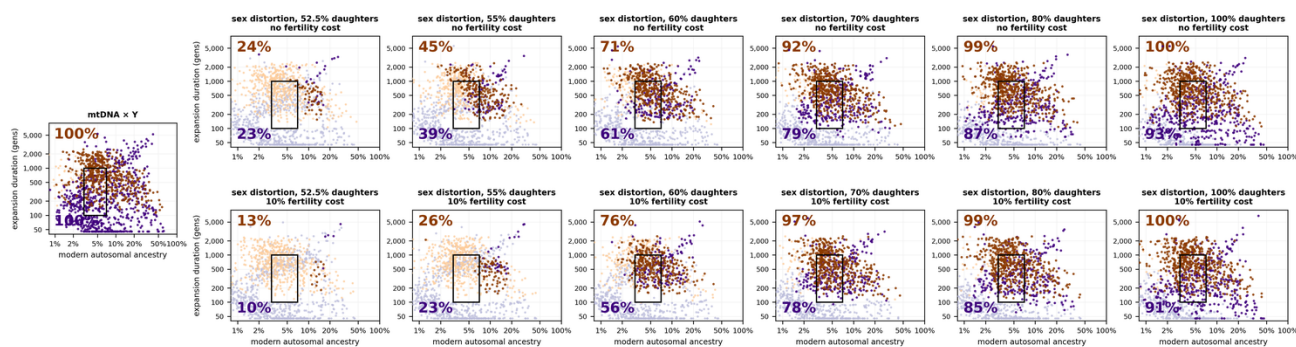
In each square, modern fraction given left-to-right for mtDNA-Y-autosomes; area is band size. M / Y / A: whole-landscape percentages.

Extended Data Figure 3. Sensitivity of deme and band frameworks to each parameter. 2,000 replicates per point in both. Light vertical line indicates the baseline value from which each axis varies the parameter.

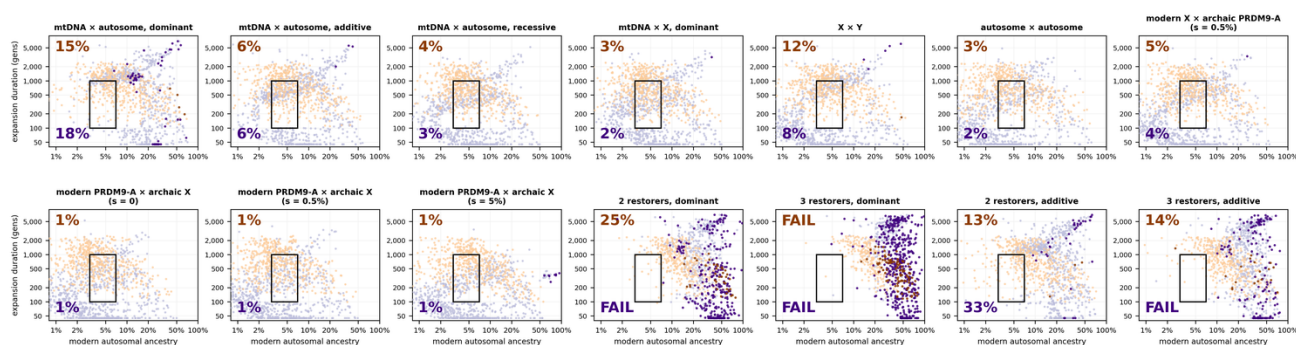


Extended Data Figure 4: Systematic survey of biological and cultural scenarios matching the data. Orange band model, purple deme model, points full grid of parameters. Black rectangle is acceptance region (modern autosomal ancestry 3-7%, expansion in 100-1,000 generations). Mean probability that all ten sampled males are modern at both mtDNA and the Y is shown for band top left, deme bottom left, averaged over admissible combinations (FAIL marks no admissible combination). Dark points are where a sample of ten males is modern at both markers with probability $\geq 50\%$. Every point inside the box is plotted in the main Figure 3b associated with the same scenario. Each cell is judged on 100 or 200 replicates per point.

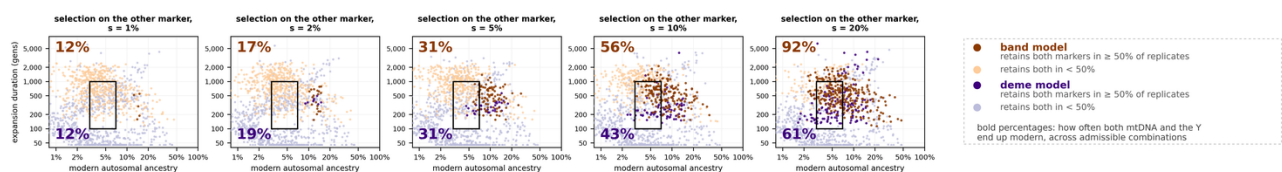
SCENARIO 1a DMI involving the Y (works effectively)



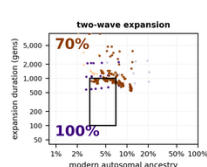
SCENARIO 1b DMI not involving the Y (does not work)



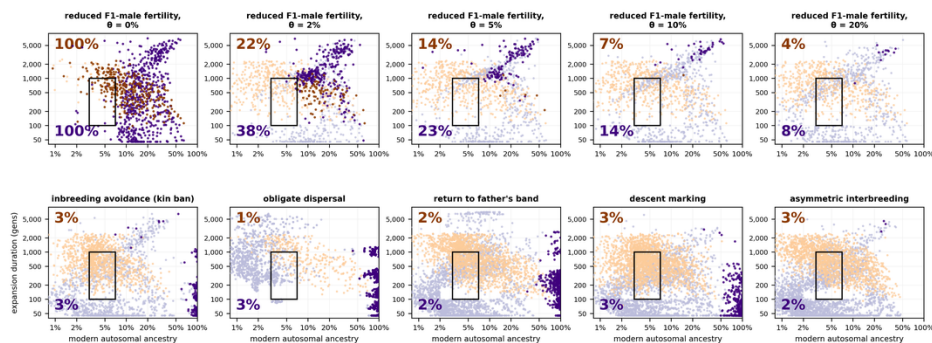
SCENARIO 2 Selection against the archaic patriline or matriline



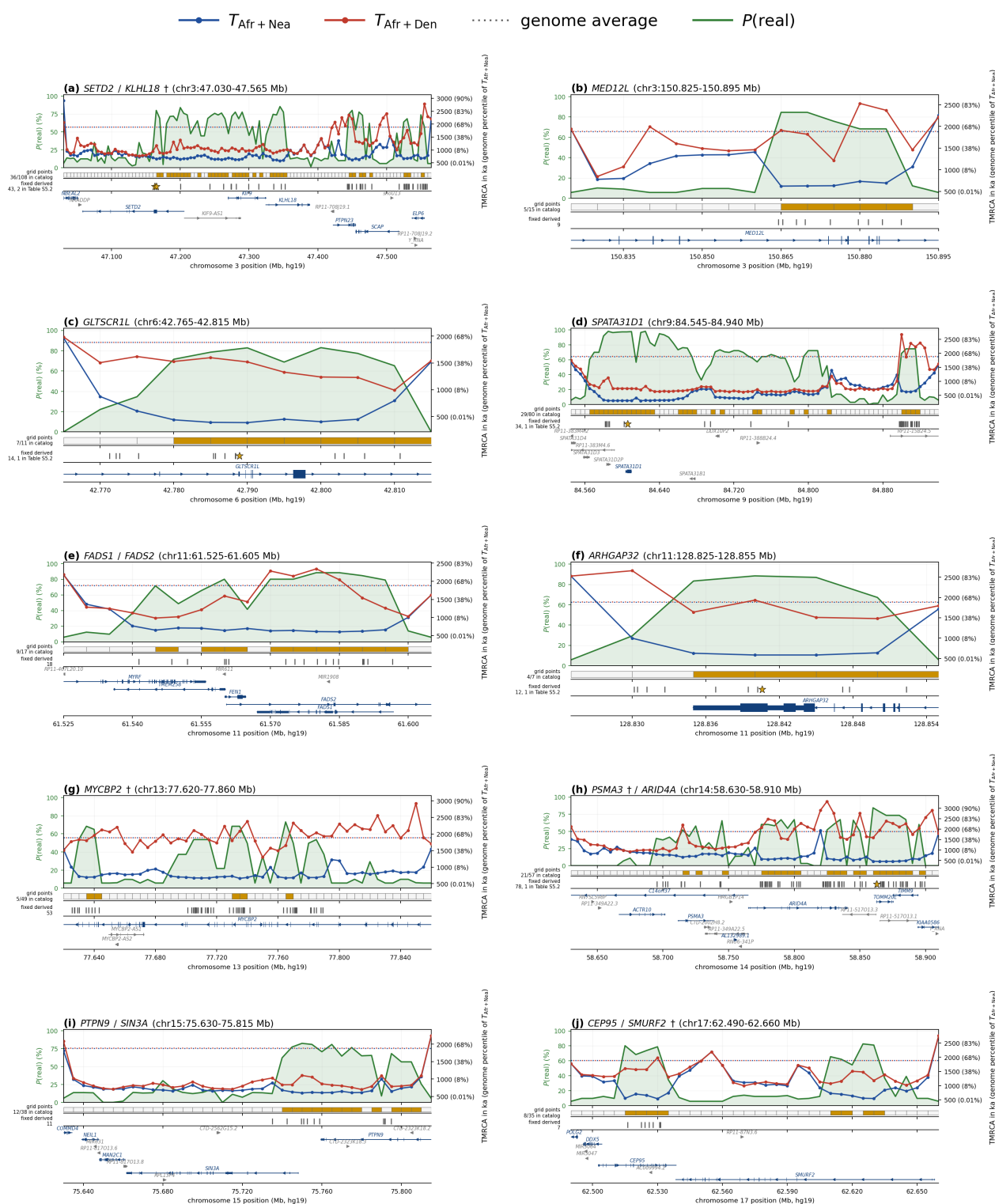
SCENARIO 3 Two waves, matrilineal and patrilineal



SCENARIO 4 Fertility of the first-generation offspring of the archaic parent's sex



Extended Data Figure 5: Ten notable genomic regions where present-day humans and Neandertals share an extraordinarily recent ancestor making them candidates for harboring evolutionary adaptations important to early modern humans. The estimated probability of modern humans and Neandertals sharing a common ancestor more recently than the population split from Denisovans is shown on the left axis as $P(\text{real})$. The right axis shows the mean time to the most recent common ancestor of both Africans and Neandertals $T_{\text{Afr}+\text{Nea}}$, and Africans and Denisovans $T_{\text{Afr}+\text{Den}}$ (parentheses give percentile in the genome-wide distribution of $T_{\text{Afr}+\text{Nea}}$; dotted lines genome-wide averages). Below the graphs, I show the grid points that qualify for inclusion in the catalog, fixed derived sites (starred if predicted to be highly functional), and GENCODE v19 genes with coding sequence drawn thickly.



References

- ¹ Peyrégne S et al. (2025) A high-coverage genome from a 200,000-year-old Denisovan. *bioRxiv* 2025.10.20.683404.
- ² Posth C et al. (2017) Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neandertals. *Nat Commun.* 8, 16046.
- ³ Petr M et al. (2020) **Error! Bookmark not defined.** The evolutionary history of Neanderthal and Denisovan Y chromosomes. *Science* 369, 1653-6.
- ⁴ Kuhlwilm M et al. (2016) Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* 530, 429-433.
- ⁵ Hubisz MJ, Williams AL, Siepel A (2020) Mapping gene flow between ancient hominins through demography-aware inference of the ancestral recombination graph. *PLoS Genetics* 16, e1008895.
- ⁶ Harris DN et al. (2023) Diverse African genomes reveal selection on ancient modern human introgressions in Neandertals. *Current Biology* 33, 4905-16.
- ⁷ Loya H, Hinch AG, Palamara PF, Speidel L, Myers SR (2026) Genome-wide genealogies reveal deep admixtures forming modern humans *bioRxiv* 2026.04.17.719197.
- ⁸ Arsuaga JL et al. (2024) How the Sima de los Huesos was won. *Anat Rec* 307, 2225-45.
- ⁹ Meyer M et al. (2014) A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* 505, 403-6.
- ¹⁰ Meyer M et al. (2016) Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* 531, 504-7.
- ¹¹ Cousins T, Scally A, Durbin R (2025) A structured coalescent model reveals deep ancestral structure shared by all modern humans. *Nat Genet* 57, 856-864.
- ¹² Fan S et al. (2023) Whole-genome sequencing reveals a complex African population demographic history and signatures of local adaptation. *Cell* 186, 923-939.
- ¹³ Yaworsky PM, Nielsen ES, Nielsen TK (2024) The Neanderthal niche space of Western Eurasia 145 ka to 30 ka ago. *Sci Rep* 14, 7788.
- ¹⁴ GLOBE Task Team et al. (Hastings, D. A., Dunbar, P. K. et al., eds.) *The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0*. NOAA National Geophysical Data Center, Boulder, Colorado (1999).
- ¹⁵ Karin, T. *Global Land Mask* (v1.0.0). Zenodo <https://doi.org/10.5281/zenodo.4066722> (2020).
- ¹⁶ Muller HJ (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1, 2-9.
- ¹⁷ Piganeau G, Eyre-Walker A (2003) Estimating the distribution of fitness effects from DNA sequence data: implications for the molecular clock. *Proc. Natl Acad. Sci. USA* 100, 10335-10340.
- ¹⁸ Wilson Sayres MA, Lohmueller KE, Nielsen R (2014) Natural selection reduced diversity on human Y chromosomes. *PLoS Genet.* 10, e1004064.
- ¹⁹ Christmas MJ et al. (2023) Evolutionary constraint and innovation across hundreds of placental mammals. *Science* 380, eabn3943.
- ²⁰ Langdon QK et al. (2024) Swordtail fish hybrids reveal that genome evolution is surprisingly predictable after initial hybridization. *PLoS Biol* 22, e3002742.
- ²¹ Robles NV et al. (2025) Admixture mapping reveals evidence for multiple mitonuclear incompatibilities in swordtail fish hybrids. *Molecular Ecology* 34, e70106.
- ²² Burton RS, Barreto FS (2012) A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Molecular Ecology* 21, 4942-57.

- ²³ Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: Massive introgression by local genes. *Evolution* 62, 1908-20.
- ²⁴ Greenwood PJ (1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, 28, 1140-62.
- ²⁵ Marlowe FW (2004). Marital Residence among Foragers. *Current Anthropology*, 45, 277-84.
- ²⁶ Currat M and Excoffier L (2004) Modern humans did not admix with Neandertals during their range expansion into Europe. *PLoS Biol.* 2, e421.
- ²⁷ Haldane JBS (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics* 12, 101-9.
- ²⁸ Frank SA (1991) Divergence of meiotic drive-suppression systems as an explanation for sex-biased hybrid sterility and inviability. *Evolution* 45, 262-7.
- ²⁹ Hurst LD, Pomiankowski A (1991) Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics* 128, 841-858.
- ³⁰ Cocquet J, Ellis PJI, Mahadevaiah SK, Affara NA, Vaiman D, Burgoyne PS (2012) A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. *PLoS Genetics* 8, e1002900.
- ³¹ Phadnis N, Orr HA (2009) A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* 323, 376-9.
- ³² Fortes M (1953) The structure of unilineal descent groups. *American Anthropologist* 55, 17-41.
- ³³ Verdu P et al. (2013) Sociocultural behavior, sex-biased admixture, and effective population sizes in Central African Pygmies and non-Pygmies. *Mol. Biol. Evol.* 30, 918-37.
- ³⁴ Joiris DV (2003) The framework of central African hunter-gatherers and neighbouring societies. *African Study Monographs Suppl.* 28, 57-79.
- ³⁵ Plagnol V, Wall JD (2006) Possible Ancestral Structure in Human Populations. *PLoS Genet.* 2, e105.
- ³⁶ Durvasula A, Sankararaman, S (2020) Recovering signals of ghost archaic introgression in African populations. *Sci. Adv.* 6, eaax5097.
- ³⁷ Ragsdale AP et al. (2023) A weakly structured stem for human origins in Africa. *Nature* 617, 755-763.
- ³⁸ Scerri EML et al. (2018) Did Our Species Evolve in Subdivided Populations across Africa, and Why Does It Matter? *Trends Ecol Evol.* 33, 582-594.
- ³⁹ Shipton C (2022) Predetermined Refinement: the Earliest Levallois of the Kapthurin Formation. *Journal of Paleolithic Archaeology* 5, 4.
- ⁴⁰ Gill JP, Ashton N, Wilkinson KN, Gasparyan B, Adler DS (2025) The shape of technology to come: An examination of evolutionary relationships between bifacial and core technologies at the Lower-Middle Palaeolithic boundary across regions in Eurasia. *J Hum Evol.* 205, 103702.
- ⁴¹ Moncel MH et al. (2020) Early Levallois core technology between Marine Isotope Stage 12 and 9 in Western Europe. *J Hum Evol* 139, 102735.
- ⁴² Hu Y et al. (2019) Late Middle Pleistocene Levallois stone-tool technology in southwest China. *Nature* 565, 82-85.
- ⁴³ Li F et al. (2019) A refutation of reported Levallois technology from Guanyindong Cave in south China, *National Science Review*, 6, 1094-6.
- ⁴⁴ Shunkov MV, Kozlikin MB (2023) The earliest Paleolithic assemblages from Denisova Cave in the Altai. *Archaeology, Ethnology & Anthropology of Eurasia.* 51, 18-32.
- ⁴⁵ Anil D et al. (2022) An Early Presence of Modern Human or Convergent Evolution? A 247 ka Middle Palaeolithic Assemblage from Andhra Pradesh, India. *Journal of Archaeological Science: Reports.* 45, 103565.

-
- ⁴⁶ Mirazón Lahr M, Foley RA (1998) Towards a theory of modern human origins: Geography, demography, and diversity in recent human evolution. *Am. J. Phys. Anthropol.*, 107, 137-76.
- ⁴⁷ Chazan M (2015). "Technological Trends in the Acheulean of Wonderwerk Cave, South Africa." *African Archaeological Review* 32, 701-28.
- ⁴⁸ Adler DS et al. (2014) Early Levallois technology and the Lower to Middle Paleolithic transition in the Southern Caucasus. *Science* 345, 1609-13.
- ⁴⁹ Davis R et al. (2026) Earliest evidence of making fire. *Nature* 649, 631-7.
- ⁵⁰ Kuhn SL (2020). *The Evolution of Paleolithic Technologies* (1st edition). Routledge.
- ⁵¹ Green RE et al. (2010) A Draft Sequence of the Neandertal Genome. *Science* 328, 710-22.
- ⁵² Mathieson S, Mathieson I (2018) *FADS1* and the timing of human adaptation to agriculture. *Molecular Biology and Evolution* 35, 2957-70.
- ⁵³ McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology* 28, 495-501.
- ⁵⁴ Luo H (2016) Interplay between the virus and the ubiquitin-proteasome system: molecular mechanism of viral pathogenesis. *Current Opinion in Virology* 17, 1-10.
- ⁵⁵ Enard D, Cai L, Gwennap C, Petrov DA (2016) Viruses are a dominant driver of protein adaptation in mammals. *eLife* 5, e12469.