Balancing selection on genomic deletion polymorphisms in humans

Authors: Alber Aqil[1], Leo Speidel[2,3], Pavlos Pavlidis[4], Omer Gokcumen[1]

Affiliations:
1. Department of Biological Sciences, University at Buffalo, Buffalo, NY. USA.
2. University College London, Genetics Institute, London, UK.
3. The Francis Crick Institute, London, UK.
4.Institute of Computer Science (ICS), Foundation of Research and Technology-Hellas, Heraklion, Crete, Greece.

Correspondence:
Omer Gokcumen, gokcumen@gmail.com

Abstract:
A key question in biology is why genomic variation persists in a population for extended periods. Recent studies have identified examples of genomic deletions that have remained polymorphic in the human lineage for hundreds of millennia, ostensibly owing to balancing selection. Nevertheless, genome-wide investigation of ancient and possibly adaptive deletions remains imperative. Here, we demonstrate an excess of polymorphisms in present-day humans that predate the modern human-Neanderthal split (ancient polymorphisms), which cannot be explained solely by selectively neutral scenarios. We analyze the adaptive mechanisms that underlie this excess in deletion polymorphisms. Using a previously published measure of balancing selection, we show that this excess of ancient deletions is largely owing to balancing selection. Based on the absence of signatures of overdominance, we conclude that it is a rare mode of balancing selection among ancient deletions. Instead, more complex scenarios involving spatially and temporally variable selective pressures are likely more common mechanisms. Our results suggest that balancing selection resulted in ancient deletions harboring disproportionately more exonic variants with GWAS associations. We further found that ancient deletions are significantly enriched for traits related to metabolism and immunity. As a by-product of our analysis, we show that deletions are, on average, more deleterious than single-nucleotide variants. We can now argue that not only is a vast majority of common variants shared among human populations, but a considerable portion of biologically relevant variants has been segregating among our ancestors for hundreds of thousands, if not millions, of years.

INTRODUCTION

The evolutionary forces that shape the allele frequency distribution of functional genetic variants remain a hotly debated issue. In humans, tens of thousands of common variants are reported to be associated with human diseases (Loos, 2020). However, the mainstream view remains that the majority of these functional genetic variants have had a negligible effect on reproductive fitness and that the frequency of these variants has fluctuated neutrally by drift over time (Bromberg et al., 2013; Dudley et al., 2012). Functional variants that have measurable fitness effects are often observed at a low frequency (Eyre-Walker, 2010). These low-frequency functional variants are considered to be in the process of being eliminated from the population by negative selection (Gibson, 2018; Lettre, 2014; Zeng et al., 2018). Nevertheless, an increasing number of studies are showing that more complex evolutionary histories (Benton et al., 2021; Mathieson and Mathieson, 2018) involving introgression from archaic hominins (McArthur et al., 2020), geography-specific adaptation (Hamid et al., 2021; Lachance and Tishkoff, 2013; Mendoza-Revilla et al., 2021), negative selection (Zeng et al., 2018), and polygenic selection (Barghi et al., 2020; Berg and Coop, 2014; Pritchard et al., 2010; Sella and Barton, 2019) may explain the allele frequencies of variants associated with complex diseases. In this context, we aim to test the hypothesis that balancing selection is a considerable force in shaping the allele frequencies of extant functional deletions in the human genome.

Balancing selection is a mode of natural selection that maintains a genomic polymorphism by overcoming the stochastic loss or fixation of one of the alleles by genetic drift (Fijarczyk and Babik, 2015; Fisher, 1922; Noonan et al., 2006). H.J. Muller was the first to discover balancing selection from his study of balanced lethals in Drosophila (Muller, 1918). Adaptive variational maintenance by balancing selection may be achieved in a number of ways. In a mechanism known as over-dominance (also called heterozygote advantage), the individual who is heterozygous for a certain variant has a higher fitness (Fisher, 1922; Wallace, 1970). In negative frequency-dependent selection, rarer variants confer higher fitness. This leads to a fluctuation of a variant's frequency in the population until an equilibrium is established, such that neither variant confers an advantage relative to the other (Smith Maynard et al., 1998; Takahashi and Kawata, 2013). Temporally varying selection, wherein the selection coefficient associated with an allele changes over time, can lead to the oscillation of this allele's frequency over time (Abdul-Rahman et al., 2021; Wittmann et al., 2017). Spatially varying selection, wherein the selection coefficient associated with an allele varies across geography, may fix this allele locally in one niche and eliminate it in another, leading to the global persistence of variation at the locus (Hedrick, 2006; Levene, 1953; Saitou et al., 2021a).

Unlike positive and negative selection, there is only a modest number of well-established instances of balancing selection (Charlesworth and Charlesworth 2016). In humans, these include polymorphisms of the *ABO* gene, which determine the A, B, and O blood groups (Ségurel et al. 2012), and polymorphisms in the major histocompatibility complex, which encodes cell-surface glycoproteins that display samples of peptides from within the cell on the cell's surface (Takahata et al. 1992). Two variants of *ERAP2,* which too is involved in the antigen-presenting pathway, have also been maintained under balancing selection (Andrés et al., 2010; Klunk et al., 2022). The classic example of recent, shorter-term balancing selection in humans is the maintenance (by over-dominance) of sickle-cell alleles at the β-goblin locus in the regions of Africa where malaria is endemic (Allison, 1954a a, 1954b b; Hedrick, 2011). Similar reasoning applies to certain □-thalassemia alleles in parts of Southeast Asia where malaria is widespread (Qiu et al. 2013). In fact, the higher fitness of heterozygotes for thalassemia alleles in malaria-struck regions was presciently predicted by Haldane in 1949 (Lederberg, 1999). In the realm of structural variants, complex copy number variation in the human salivary agglutinin

genes (Alharbi et al., 2022), a regulatory deletion upstream of APOBEC3 gene family (Gokcumen et al., 2013), and a deletion spanning *LCE3B* and *LCE3C (Pajic et al., 2016)*, which is associated with psoriasis, have been explicitly argued to be evolving in the human lineage under balancing selection.

So far, most systematic investigations into balancing selection in modern humans have focused primarily on genes (DeGiorgio et al., 2014; Soni et al., 2022) and on single nucleotide variants (SNVs) (Siewert and Voight, 2020); (Bitarello et al., 2018; Siewert and Voight, 2017). Additionally, some studies have focused exclusively on "long-term" balancing selection wherein variants have been maintained in the human lineage since before the split from the chimpanzee clade (Leffler et al., 2013). Others have focused on short-term or population-specific balancing selection (Hedrick, 2011; Qiu et al., 2013). We set out to identify potential targets of balancing selection that are structural in nature and that may not have been captured by earlier studies. Thus, we concentrate our efforts on autosomal deletion polymorphisms (> 50bp) that have been maintained in the human lineage since before the split, approximately ~700,000 years ago, of anatomically modern humans (AMHs) from the lineage that led to both Neanderthals and Denisovans (henceforth, collectively referred to as archaic hominins). In this study, we will use the term "ancient polymorphisms" to refer to such polymorphisms. Focusing on such "medium-term" balancing selection will likely allow us to capture more potential targets than could an exclusive study of either "long-term" or "short-term" balancing selection. Moreover, deletions are interesting in the context of selection: since a given deletion affects more nucleotides than a single nucleotide variant (SNV), if a defined genomic window is indeed of adaptive importance, deletions may have more profound functional consequences (Conrad et al., 2010; Saitou and Gokcumen, 2020). Such functional outcomes may translate into non-trivial selection coefficients either for or against the deletion. Additionally, deletions are relatively easier to both genotype and analyze than are other structural variants, making an evolutionary analysis involving deletions tractable.

RESULTS AND DISCUSSION

AMH exhibits a greater proportion of ancient polymorphisms than expected under adaptive neutrality

Older polymorphisms may be more likely than newer ones to exhibit signatures associated with balancing selection because they have survived stochastic fixation or elimination for extended periods. It is, therefore, possible that a certain proportion of human polymorphisms that are older than the AMH-archaic split (~700,000 years), *i.e.,* ancient polymorphisms, (**Figure 1A**) have been maintained by balancing selection. We tested this hypothesis by comparing the proportion of ancient polymorphisms segregating in AMHs to a neutrally expected distribution of this proportion. If this proportion is significantly higher in the observed data than under the neutrally simulated data, and if we can reject other plausible explanations for difference, we can conclude that some of the polymorphisms older than 700,000 years may have been maintained by balancing selection.

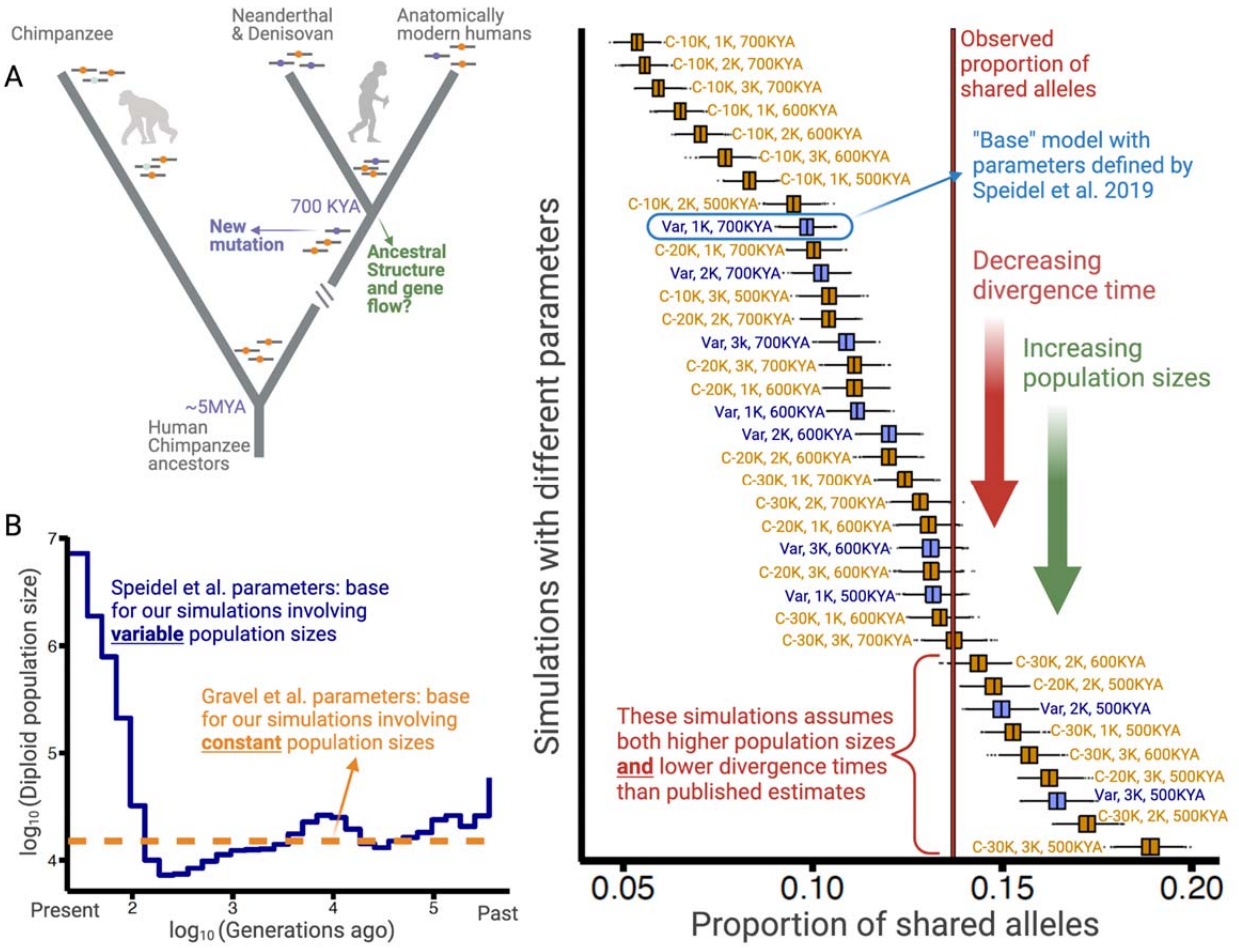

Figure 1. Excess of ancient polymorphisms segregating in AMHs. A. A schematic representation of derived "ancient" variants (purple) that emerged before the AMH-archaic hominin divergence (and after hominin-chimp divergence), and have remained polymorphic in the AMH lineage. The ancestral variants are indicated as orange, and the derived chimpanzee-specific variants are indicated in light blue. **B.** The Speidel et al. and Gravel et al. simulation parameters. Speidel et al. provide parameters that involve varying population sizes for the YRI population. **C.** Expected distribution of the proportion of ancient polymorphisms in YRI under different models. Each distribution is labeled with three parameters in the form (*AMH-N_e, Archaic-N_e, time since archaic-AMH divergence*). The simulations where we used variable effective population size published by Spiedel et al. are indicated by blue color and labeled "Var". The simulations where *AMH-N_e* is constant are shown in orange, and provide the population size used. The vertical line represents the empirical proportion of ancient polymorphisms in YRI.

For this test, we focused on 28,291 randomly chosen SNVs (minor allele count > 1) in the Yoruba (YRI) population (1000 Genomes Project Consortium et al., 2015); see **Supplementary material** for a discussion of our rationale behind using SNVs rather than deletion polymorphisms, the variant class of our interest). We focused on random SNVs instead of all the SNVs in order to mitigate the biases that would be introduced due to linkage. A variant was classified as ancient if the derived allele was shared, by common descent, with at least one of the four high-coverage archaic hominin genomes (three Neanderthals and one Denisovan) (Mafessoni et al., 2020; Meyer et al., 2012; Prüfer et al., 2017, 2014). We found that the ancient SNVs make up 13.7% (3,894 SNVs) of the total. Note that we removed the recurrent SNVs from our analysis using a linkage-disequilibrium based approach (see **Methods**). To compare the proportion of ancient SNVs against neutral expectations, we used *ms* (Hudson, 2002) to

produce 2,000 runs of 20,000 neutrally simulated variants in the Yoruba population (see **Methods** for details). Thereupon, in each run, we calculated the proportion of variants shared with archaic hominins, producing a distribution of the expected proportion of ancient polymorphisms under neutrality.

To ensure that our analysis is not biased by the idiosyncrasies of any particular model, we performed these simulations using 36 distinct models. The models vary by three parameters: N_e of Yoruba/AMH, N_e of archaic hominins, and the time of divergence between the AMH and the archaic hominin lineage. The N_e for humans can be either constant (ranging from 10,000 to 30,000) or varying over time (Speidel et al., 2019) (**Figure 1B**). N_e of archaic hominins ranges from 1,000 to 3,000; and the divergence time ranges from 500 to 700 kya (Bergström et al., 2021; Mafessoni et al., 2020; Meyer et al., 2012; Prüfer et al., 2014). In the main text, we focus on the model with variable AMH-N_e (Speidel et al., 2019), using the well-accepted archaic hominin N_e of 1,000 and a divergence time of 700 kya. We refer to this as the "base model".

Using this model, we find that the entire simulated distribution of the proportion of ancient polymorphisms lies to the left of the empirical value of 13.7% (**Figure 1C**). These results hold for all other models with realistic sets of parameters. Even when we change any single parameter to unrealistic levels (e.g., either AMH-N_e = 30,000, or divergence time = 500 kya), we still observe an excess of ancient polymorphisms. Neutral models can explain the empirical proportion of ancient polymorphisms only when at least two parameters are tweaked in a less realistic direction (e.g., both AMH-N_e = 30,000 and divergence time = 500 kya). Therefore, we conclude that the high proportion (13.7%) of ancient polymorphisms cannot be explained by realistic neutral scenarios.

There is a possibility that we are observing an empirical excess of ancient polymorphisms owing to a high incidence of recurrent mutations in the AMH and archaic hominin lineages that remained undetected by the linkage-disequilibrium-based approach we used to identify them. Since CpG sites may be particularly prone to recurrent mutations, we calculated the proportion of empirical ancient polymorphisms again using only A⇄T SNVs. This analysis yielded a proportion of 14.24%, not much different from the previously calculated 13.7%. Moreover, if recurrence was the main cause of the observed excess of ancient polymorphisms, we would expect this excess (relative to the simulated distribution) to be more pronounced among polymorphisms with low derived allele frequencies. To test if this is the case, we repeated our analysis using the base model for simulations, dividing the empirical and simulated SNVs into derived allele frequency bins (**Figure 1—figure supplement 1**). We observed that the excess of ancient polymorphisms is, in fact, most pronounced at high derived allele frequency. Both these results combined suggest that our results are not biased due to undetected recurrent mutations.

Next, we consider possible non-adaptive explanations for the observed excess of ancient polymorphisms. First, we considered scenarios invoking structure in the population that was ancestral to both AMHs and archaic hominins, while allowing gene flow between the latent subgroups within the ancestral population (**Figure 1**; see **Methods** for details). We found that the excess of ancient polymorphisms can be explained by structuring the ancestral population

into 3 distinct subpopulations, such that the fraction of each subgroup formed by the migrants of each of the other subgroups, every generation, is below 0.0075% (**Figure 1—figure supplement 2A-B**). However, the allele frequency spectrum for SNVs simulated with ancient population structure significantly deviates from the observed allele frequency spectrum in that the former overestimates the intermediate/common variants (**Figure 1—figure supplement 2C**). Therefore, invoking such structure to explain the excess of ancient polymorphisms may be unrealistic. Another possible explanation comes from the evidence of introgression from early modern human ancestors into Neanderthals to the exclusion of Denisovans (Posth et al., 2017). Such admixture can increase the apparent proportion of ancient polymorphisms due to elevated allele sharing with Neanderthals. However, we do not observe a higher proportion of derived allele sharing (by common descent) with Denisovans than with Neanderthals. In fact, the proportion of derived alleles shared with the Denisovan (9.75%) and the Altai Neanderthal (10.28%) is higher than the proportion shared with Chagyrskaya and Vindija Neanderthals (6.14% each), which is incompatible with such introgression as the prime cause of excessive ancient polymorphisms in AMHs. We note that the differential allele sharing with the Denisovan and Altai Neanderthal on one hand, and the Vindija and Chagyrskaya Neanderthal on the other would be an interesting subject for future studies.

Overall, based on our current knowledge of ancient interactions and demographic history, our analyses implicate balancing selection as a possible cause of the excess of observed ancient polymorphisms. Next, we focus on deletion polymorphisms segregating in AMHs, categorize them based on their evolutionary histories, and test whether ancient deletion polymorphisms are enriched for targets of balancing selection.

Categorizing human deletions based on their evolutionary history

Having established that AMHs exhibit an excess of ancient polymorphisms that cannot be explained solely by non-adaptive causes, we identify ancient deletion polymorphisms among AMHs. Since the vast majority of deletions in AMHs are derived relative to chimpanzees (**Supplementary material**), this could be accomplished by identifying AMH deletions that are shared with archaic hominins by common descent.

In this analysis, AMHs are represented by the YRI (Yoruba), CEU (Utah residents with Northern and Western European ancestry), and CHB (Han Chinese in Beijing) from 1000 Genomes Project Phase-3 dataset (1000 Genomes Project Consortium et al., 2015); and archaic hominins are represented by the four available high-coverage (~30X) archaic hominin genomes (Mafessoni et al., 2020; Meyer et al., 2012; Prüfer et al., 2017, 2014). Our choice of AMH populations was guided by our wish to both sample from different regions, and use relatively well-studied populations. We genotyped all AMH deletions in the archaic hominin genomes using a read depth based pipeline (**Supplementary File 1**). We considered a deletion "shared" if it was identified in at least one of the four archaic genomes. For our analysis, we used only the deletions with an allele count greater than 1 in YRI, CEU, and CHB combined. Additionally, we retained only 4,863 human deletion polymorphisms that are in linkage-disequilibrium (LD, $r^2 >$ 0.9) with at least one SNV (**Supplementary File 2**). We imposed this LD requirement because SNVs in LD with the deletion can enable us to distinguish the shared deletions that are introgressed or recurrent from those that are shared by common descent.

We found that 575 (11.8%) AMH deletions were shared with archaic hominins, *i.e.,* identified in at least one archaic hominin genome (**Figure 2**). We identified 53 instances of independent emergence (*recurrent deletions*) in archaic hominin and AMH lineages, wherein no SNV that is in LD with the deletion in AMHs accompanied the deletion in archaic hominins. In parallel, we identified 92 deletions that were *introgressed* from archaic hominins into AMHs: the SNVs in LD

with these deletions were present in previously identified introgressed haplotypes (Taskent et al., 2017; Vernot and Akey, 2014). By this process of elimination, we found that 430 (8.8% of the total) shared deletions are ancient polymorphisms, *i.e.,* they are shared with archaic hominins by common descent and thus emerged at least ~700,000 years ago.

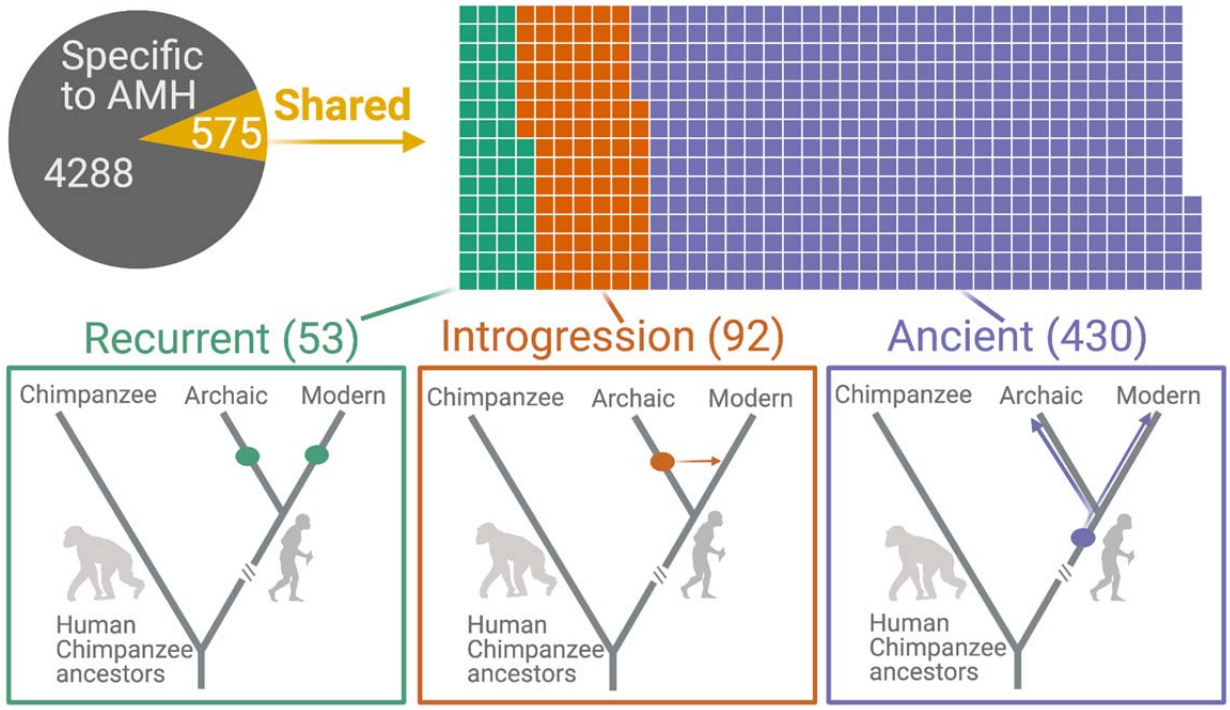

Figure 2 - Deletions in AMHs that are shared with archaic hominins. The top panel shows the categorization of deletion polymorphisms as AMH-specific, recurrent (green), introgressed (orange), or ancient (purple). The evolutionary histories of shared deletions are summarized schematically in the bottom panel.

To confirm that our pipeline for identifying ancient deletions **(Figure 2—figure supplement 1)** has high accuracy, we estimated the ages of deletions, based on the ages of SNVs in LD. We used two methods in parallel: 1) Human Genome Dating (Albers and McVean, 2020); and 2) Relate (Speidel et al., 2019) **(see Methods)**. If both our genotyping pipeline and categorization of shared deletions (as recurrent, introgressed, or ancient) are sound, we should expect that ***Age****(human-specific)* ≈ ***Age****(recurrent)* < ***Age****(introgressed)* < ***Age****(Ancient).* Both methods yielded the expected pattern of ages across the categories of deletions **(Figure 3, Figure 3—figure supplement 1)**. We found that the median age for ancient deletions, using Relate, is ~1 million years. About 15% (63) of these deletions are older than 2 million years. In contrast, the median age of non-ancient deletions is ~239,000 years. As such, we infer that both our genotyping pipeline and deletion categorization approach are sound. Counterintuitively, a small number of "ancient" deletions have very recent dates. This may be due to instances of recent soft sweeps involving some deletions leading to an increased length of the associated haplotype and an artificial decrease in age. Secondly, some ancient deletions may have low frequencies, which too, creates a downward bias in age. Lastly, this may be due to miscategorization of non-ancient deletions as ancient. Next, we test if ancient deletions are enriched for targets of balancing selection relative to non-ancient deletions.

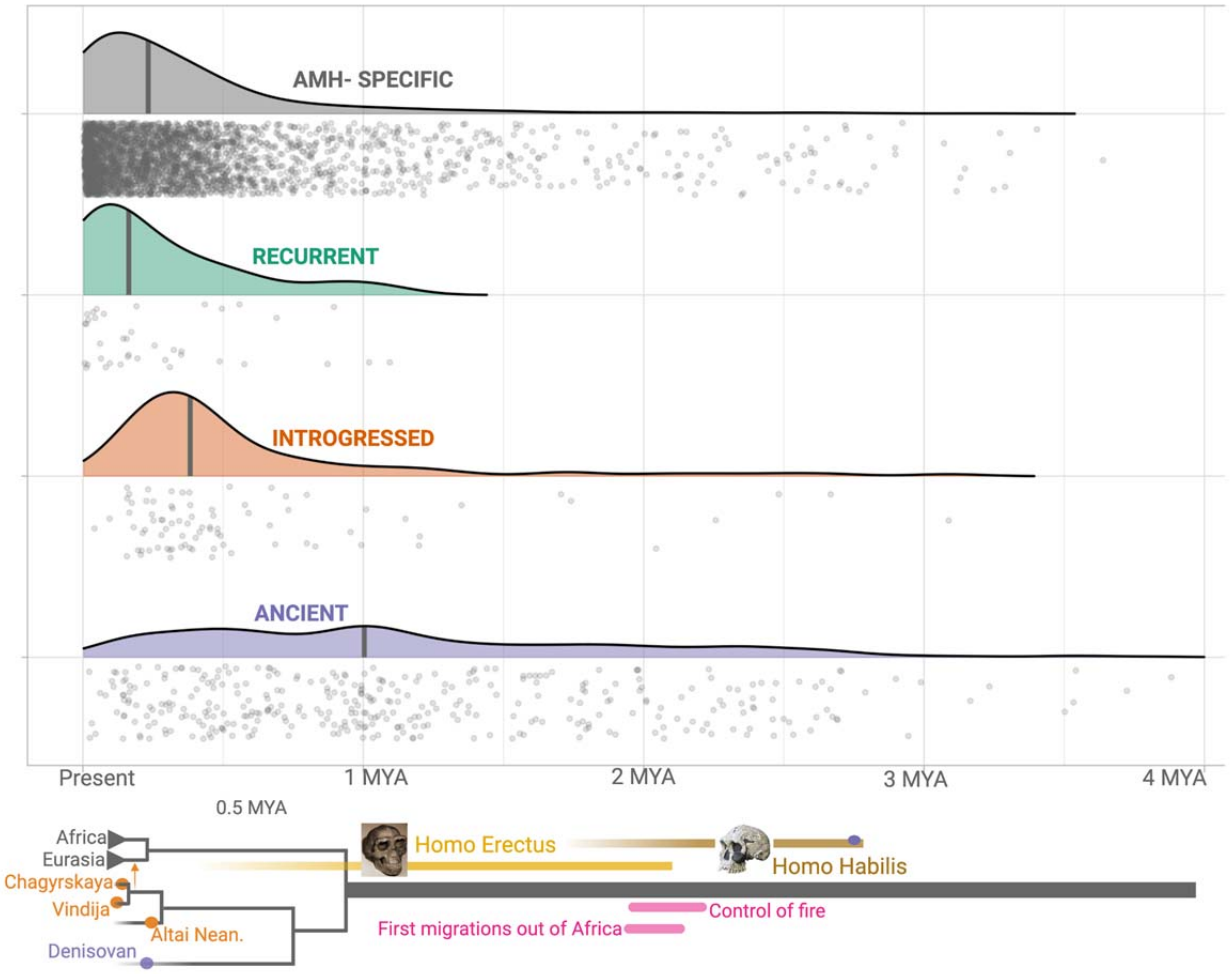

Figure 3. Age estimates of the haplotypes harboring polymorphic deletions. The x-axis shows the age estimates, obtained using Relate, for the deletions. For orienting the reader regarding the age of these variants, we provide below a schematic phylogeny representing recent human evolution.

Ancient deletion polymorphisms are more likely to be targets of balancing selection than are non-ancient ones

We used the std□2-statistic (Siewert and Voight, 2020) to test the hypothesis that ancient deletion polymorphisms are more likely than are non-ancient deletions to be targets of balancing selection. Std□2 is a measure of balancing selection, that calculates the weighted average of the number of flanking derived variants, where weights are the similarity in frequency between the core allele and the flanking variants (Siewert and Voight, 2017).

For a conceptual understanding of Std□2 (**Figure 4A**), suppose a deletion emerges and the resulting polymorphism is subject to balancing selection; the deletion will rise in frequency until it reaches a certain equilibrium frequency. New SNVs will emerge on the haplotypes carrying the deletion. Some of these SNVs will drift upward in frequency, but since these SNVs will be linked to the deletion, they too can only rise to the equilibrium frequency of the balanced deletion (Siewert and Voight, 2020, 2017). We refer to this type of drift as *Goldilocks drift*, since the linked SNVs drift upward to the "just-right" equilibrium frequency of the balanced deletion. Goldilocks drift thus leads to allelic class build-up (analogous to how hitchhiking leads to sweeps), which refers to a situation involving the fixation of many flanking variants within the set

of haplotypes carrying the deletion. Std□2 value for a core variant may be thought of as the average intensity of *Goldilocks drift* experienced by SNVs around it. Therefore, a high std□2 value for a variant implies that it is either a target of balancing selection or close to a target of balancing selection.

We observed that std□2 estimates for ancient deletions are significantly larger than those for non-ancient deletions across YRI, CEU, and CHB populations ($p < 10^{-7}$, Wilcoxon) (**Figure 4B**). These results provide empirical evidence that ancient deletion polymorphisms are enriched for targets of balancing selection. Our results are consistent with other recent studies (Soni et al., 2022) that have argued that the role of balancing selection in explaining the maintenance of common variation in the human lineage is underappreciated.

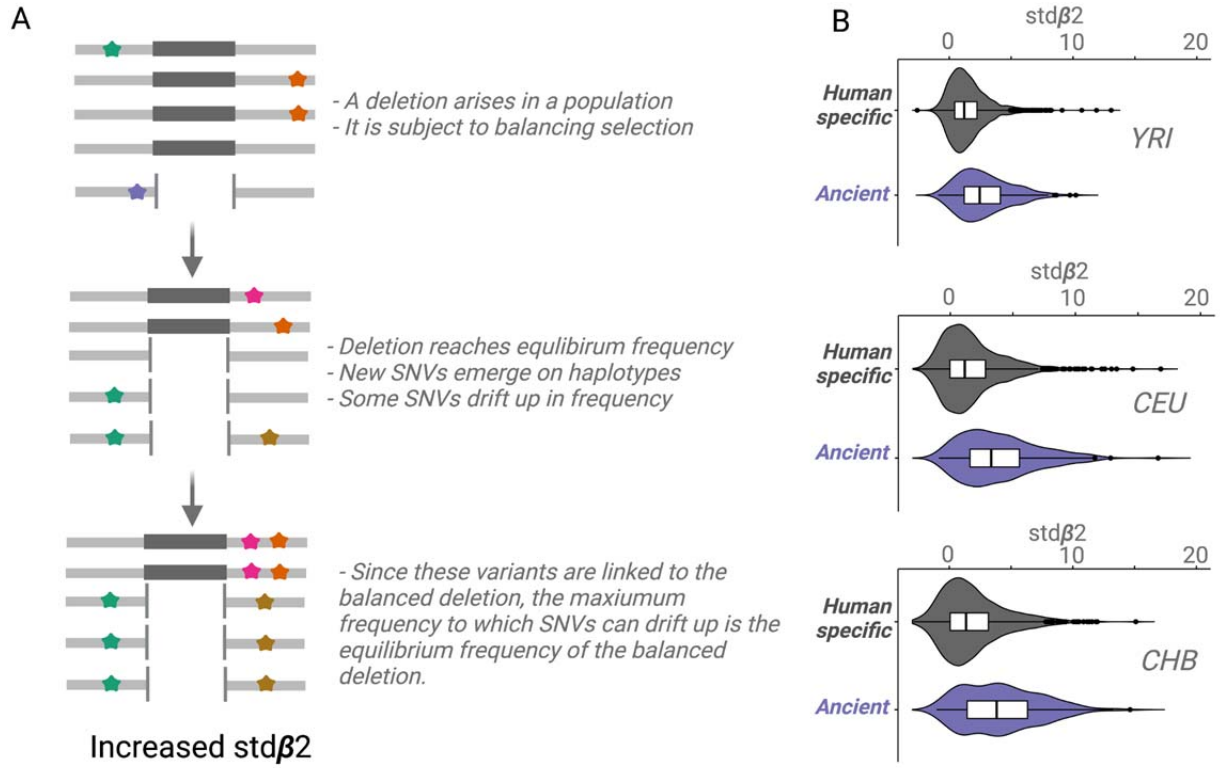

Figure 4. An empirical assessment of putative balancing selection among ancient deletions. A) The conceptual framework in which std□2 statistic works. The last step demonstrates "Goldilocks" drift (the process that results in allelic class build-up). **B**) A box plot for std□2 for AMH-specific, versus ancient deletions (frequency > 5% in respective populations). Higher std□2 values for older deletions represented in purple empirically show that older deletions are significantly enriched for targets of balancing selection. All comparisons are significant, $p < 10^{-7}$ (Wilcoxon).

Previous genome-wide balancing selection scans focused on either individual genes or single nucleotide variants (SNVs). Consequently, we do not expect to find many ancient deletions that have previously been reported as targets of balancing selection. Nevertheless, we investigated whether the exons in any of the genes that have been reported as targets of balancing selection in DeGiorgio et al. (2014) or Soni et al. (2022) overlap with ancient deletions. We found no overlaps. We also found that 77 common (> 5% in YRI, CEU, and CHB combined) ancient deletions were in LD ($r^2 > 0.9$) with SNVs that had high (in the 95th percentile) Std□2 (Siewert and Voight, 2020) values in YRI, CEU, and CHB. This is unsurprising since this is the measure we used to show that ancient deletion polymorphisms are enriched for balancing selection

targets. Interestingly, one of the ancient deletions with a high associated Std□2 value overlaps a candidate region for balancing selection previously identified using the non-central deviation (*NCD*) method (Bitarello et al., 2018). This 433 bp deletion (esv3607090), which is 2 million years old and common across populations, deletes part of an intron of the *STK32A* gene. This could be an interesting subject for future studies. Regardless, a vast majority of common ancient deletions (73%) were not reported previously as balancing selection candidates and thus are novel targets for future studies.

Phenotypic relevance of ancient deletion polymorphisms

Selection can only act on a region of the genome by means of the phenotypic function it confers. It follows then that any adaptively maintained ancient polymorphisms must be functional. If an appreciable proportion of ancient deletion polymorphisms have evolved under balancing selection and more recent deletion polymorphisms have not, we should expect ancient deletions to be enriched for functional effects. To avoid biases introduced by different proportions of rare variants among ancient versus non-ancient deletions, we focus only on deletion with frequency > 5% in AMHs. For both ancient and non-ancient deletions, we investigated 1) whether a deletion intersects exons and 2) whether any of the SNVs in LD with it are associated with UK BioBank GWAS traits with $p < 10^{-8}$ (http://www.nealelab.is/uk-biobank/**; Figure 5A; see Methods**). We did not observe a significant increase in either the proportion of exonic (ancient=5.6%; non-ancient=3.6%) or GWAS-associated (ancient=19.1%; non-ancient=16.4%) ancient deletion, relative to non-ancient deletions (**Figure 5B**). However, when we classify a deletion as functional more conservatively, *i.e.,* it both intersects an exon and has a GWAS association (*i.e.,* one of the SNVs in LD with it has a GWAS association), we observe a 4.2-fold enrichment (p=0.003) of functional variants among ancient deletions (**Figure 5B**). In fact, out of the 18 common deletions (frequency > 5%) that both intersect genes and have GWAS associations, 9 (50%) are ancient. Further, we observed a phenotypic enrichment among ancient deletions for some GWAS trait categories: a 12.5-fold enrichment ($p=10^{-5}$) of traits related to inflammatory response and a 2.8-fold enrichment (p=0.003) of traits related to metabolism (**Figure 5C**).

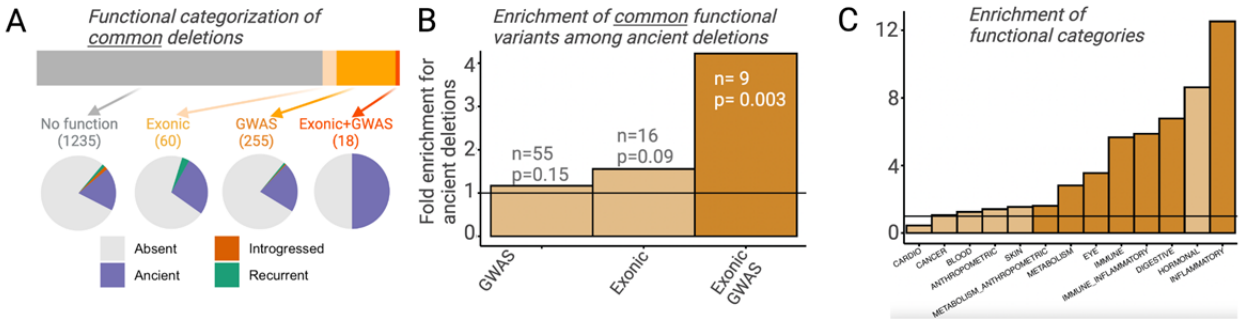

Figure 5. Functional enrichment among ancient deletions. A. Functional categorization of common deletions. Within each category, the proportions of deletions falling under different evolutionary categories are shown in pie-charts. **B.** Permutation-based analysis of enrichment of functionality among ancient deletions, relative to non-ancient deletions. The black horizontal line indicates the expected ratio of 1.0. For each definition of functionality, the number of functional ancient deletions, and the p-value associated with the enrichment are provided. **C.** Permutation-based enrichment analysis for different phenotypic categories (based on GWAS) among ancient deletions, relative to non-ancient deletions. The black horizontal line indicates the expected ratio of 1.0. Dark orange indicates a statistically significant deviation from the expected ratio of 1.0. Light orange means no significant deviation from the expected ratio of 1.0.

A focused literature review and analysis of functional effects associated with some of the ancient deletions revealed multiple mechanisms through which they affect function (**Figure 6A**).

Firstly, whole gene deletions may affect the function of entire environment-interacting gene families. We found two ancient whole gene deletions: esv3587563 (deleting *LCE3B* and *LCE3C*) and esv3600896 (deleting *UGT2B28*). The members of the *LCE3* and *UGT2B* gene families mediate immune response and steroid metabolism, respectively; genes from both families likely evolved under adaptive forces (de Guzman Strong et al., 2010; Pajic et al., 2016; Starr et al., 2021; Xue et al., 2008). The functional consequence of whole gene deletions is, of course, loss of function of the deleted genes. In addition, esv3587563 is associated with an increase in the expression of *LCE3A (de Guzman Strong et al., 2010; Pajic et al., 2016),* while esv3600896 is associated with an increase in the expression of *UGT2B11.* Thus, we propose that whole-gene deletions of members of environment-interacting gene families may lead to the functional "fine-tuning" of the entire gene family.

Secondly, we revealed dozens of potentially adaptive ancient deletions that both mediate gene regulation and are associated with human traits. For example, we found multiple ancient deletions that are proximal to the *HLA* locus, associated with immune-related phenotypes, and affect the expression levels of nearby genes. One such example, the deletion esv3608584, is noteworthy within the context of balancing selection because it affects the expression of different *HLA* genes in opposite directions (**Figure 6B**). As such, this deletion may lead to increased susceptibility to some pathogens while increasing the defenses against others. Further, we observed that the ancient deletions in the *HLA* locus also lead to the expression of different isoforms of HLA genes. Using the GTeX database, we found at least four other instances where ancient deletions lead to the expression of different isoforms, including deletions affecting the *HLA-DRB1-6, HLA-DOB, SIRPB1, GHR,* and *CYP3A43* genes. We recently showed that the ancient deletion of the third exon of the growth hormone receptor gene leads to the expression of a smaller version of growth hormone, which may be adaptive in times of starvation (Saitou et al., 2021b). The *SIRPB1* gene encodes a glycosylated transmembrane receptor protein (Kharitonenkov et al., 1997), and its different isoforms may lead to the recognition of different pathogens. Similarly, *CYP3A43*, a member of the cytochrome p450 gene family, is involved in metabolizing external substances, and genetically determined isoforms contribute to its functional variation in humans (Agarwal et al., 2008). Thus, ancient deletions that lead to specific isoform expression may have been adaptively evolving to adjust the function of environment-interacting genes across both geography and time. It is important to acknowledge that these non-exonic deletions may not be the causal variant in the associated haplotypes. Nevertheless, the full extent of deletion polymorphisms shaping the expression levels and sculpting the isoform diversity at the genetic level remains a fascinating area of future research.

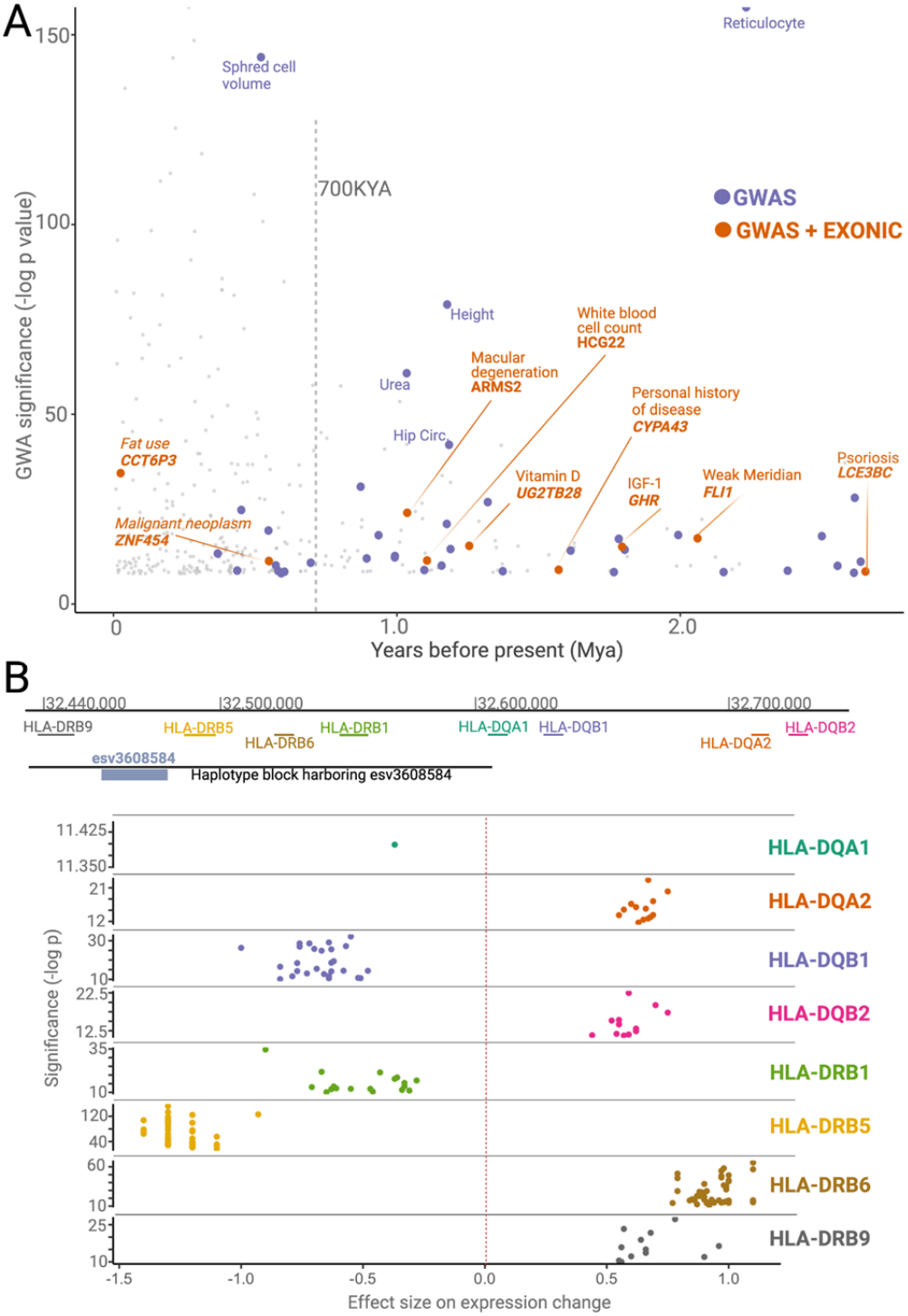

Figure 6. A. The significance levels (-log(p-value)) of phenotypic associations of deletions with GWAS traits as a function of their emergence time. Gray points indicate non-ancient deletions. Purple and orange points indicate non-exonic ancient deletions with GWAS hits and exonic ancient deletions with GWAS hits, respectively. The genes whose exons are covered by ancient deletions, and the traits associated with ancient deletions are mentioned in the plot. **B.** The significance levels (-log(p-value)) and sizes of expression level changes of nearby *HLA* genes associated with the presence of the deletion esv3608584. Each color refers to a different *HLA* gene. Each point in a given color represents a different tissue. Only those tissues whose expression level changes are statistically significant are shown here.

The effect of negative selection is stronger on deletions than SNVs

Based on previous work, we expect that deletions are more likely than SNVs to be under negative selection (Conrad et al., 2006; Kondrashov, 2017; Lin et al., 2015; Lin and Gokcumen, 2019). To investigate the magnitude of this effect, we compared the proportion of SNVs and deletion polymorphisms that are ancient. Applying the same bioinformatic pipeline to identify ancient polymorphisms in both cases, we found that 13.7% of SNVs and 9.6% of deletion polymorphisms are ancient in YRI. This result alone suggests that deletion polymorphisms are more likely than SNVs to be eliminated by negative selection, a trend that we expect to be more pronounced with increasing ages of polymorphisms. The greater intensity of negative selection acting on deletions implies that deletions are, in general, more deleterious than SNVs. It follows that younger (non-ancient) deletions currently segregating in human populations, which negative selection has not yet purged, are more likely to be deleterious (and perhaps disease-causing) than SNVs (Kondroshov, 2017).

The preceding argument makes intuitive sense since a given deletion spans more bases than does a SNV. If this intuition is correct, we expect that larger deletions should, on average, experience more intense negative selection. Since most large ancient deletions would have been purged by negative selection, we expect surviving ancient deletions to be, on average, smaller than non-ancient deletions. We test this using common deletions (frequency > 5% in YRI, CEU, and CHB combined). Ancient deletions are indeed 14% shorter than non-ancient deletions ($p=0.02$; permutation test). Nevertheless, there is an excess of long deletions among ancient deletions relative to non-ancient deletions (**Figure 7A**). In particular, the 95th and 98th size percentiles of ancient deletions are 33% ($p=0.04$; permutation test) and 128% ($p=0.005$; permutation test) larger than non-ancient deletions, respectively (see **Methods**). This excess of longer deletions is inconsistent with evolution under neutrality or negative selection. Therefore, the longest 5% of ancient deletions are excellent targets for future studies of balancing selection. In fact, 3 out of the 9 GWAS-associated common exonic deletions intersecting *SIRPB1*, *LCE3A*, *LCE3B*, and *UGT2B28* are in the 95th percentile of the size distribution of ancient deletions.

Strong overdominance is rare among ancient deletion polymorphisms

Having established that ancient deletion polymorphisms appear enriched for targets of balancing selection, we wanted to investigate whether classical overdominance is a common mechanism underlying this observation. To accomplish this, we first identified the genomic signatures that we expect to see in a region where a polymorphism has evolved under overdominance, and then we look for these signatures among ancient deletions. To identify the signatures of overdominance, we simulated sequence evolution under neutrality and overdominance (using a variety of selection coeffiecients), in turn, for variants that emerged one million years ago. We asked whether we can distinguish between neutrality and overdominance by calculating several population genetic statistics, including Tajima's D, □, □, etc., on sequences generated from the neutral and overdominance simulations (see **Methods** for full list) . We found that none of these statistics alone can distinguish overdominance from neutrality, even for strong selection coefficients (**Figure 7—figure supplement 1**)

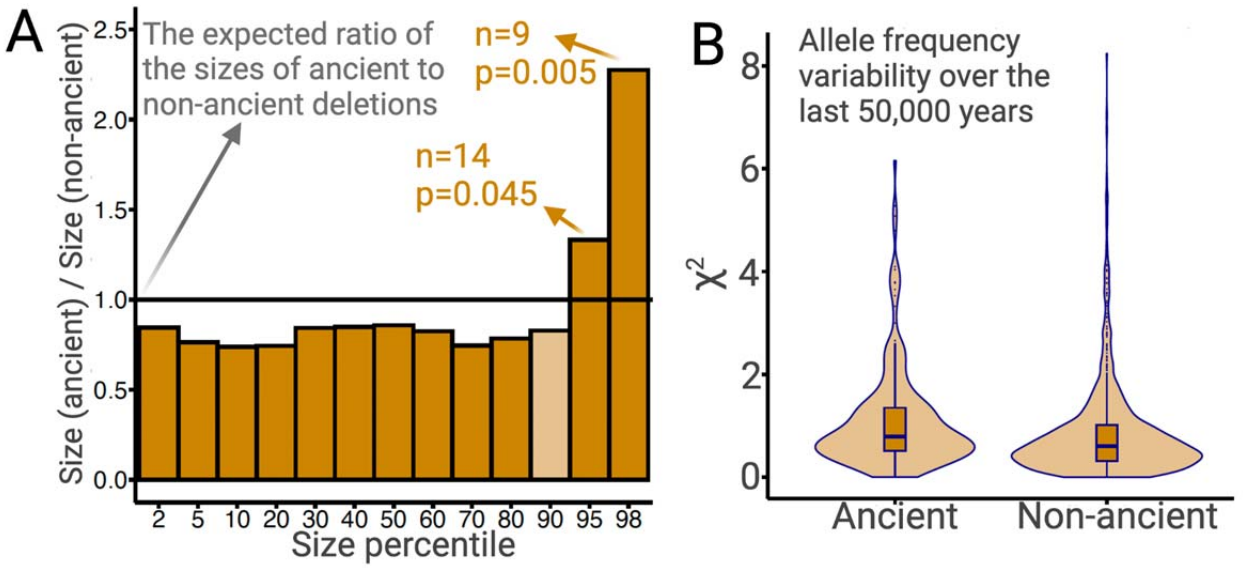

Figure 7. A. The ratios of sizes of ancient deletions to those of non-ancient deletions at different size percentiles. The black horizontal line refers to the expected ratio of 1.0. Dark orange bars refer to a statistically significant (permuatation test) deviation from the expected ratio. Light orange bars mean that the deviation from the extend ratio of 1.0 is not statistically significant. **B.** The estimated measure of allele frequency change (□2) between 50,000 and 5,000 years before present in common ancient versus common non-ancient deletions. Ancient deletions have significantly ($p=2x10^{-7}$, Wilcoxon) higher frequency variability over the last 50,000 years.

Instead, we found that a distinct feature of overdominance is that the allele frequency rapidly increases (similar to a selective sweep) until it reaches an equilibrium frequency, whereafter it remains remarkably stable across time (**Figure 7—figure supplement 1**). In contrast, under neutrality, a random change in allele frequency in every generation produced elevated noise across time in allele frequency trajectories. To ascertain if overdominance is a common mechanism of evolution for ancient deletions, we inferred the allele frequency trajectories of ancient deletions using *Relate* (Speidel et al., 2019) and quantified the variation in allele frequency between 5,000 and 50,000 years ago by squaring the standardized allele frequency difference ($□^2$) (**Methods**). We already know that ancient deletions are enriched for targets of balancing selection. If large proportion of these balancing selection targets have evolved under overdominance, we expect ancient deletions to have more stable allele frequencies across time, relative to non-ancient deletions, leading to smaller $□^2$ values on average. However, we do not observe this trend among common (frequency > 5% in YRI, CEU, and CHB combined) deletions (**Figure 7B**). Consequently, at least with our current resolution of allele frequency trajectory estimation, we found no evidence for overdominance being the prime mode of balancing selection operating on ancient polymorphisms in AMHs.

In fact, we observe that ancient deletion polymorphisms exhibit greater allele frequency variation than do non-ancient deletions (**Figure 7B;** $p=2x10^{-7}$, Wilcoxon). This suggests that a large proportion of the instances of medium-term balancing selection likely involve temporally and spatially variable selection, which lead to elevated levels of allele-frequency variation over time. This is consistent with our locus-specific analyses of ancient deletion polymorphisms. For example, we recently reported that the deletion (esv3604875) of the third exon of the human growth hormone receptor gene (*GHR*) has evolved under temporally and geographically variable adaptive constraints (Saitou et al., 2021b). In fact, this deletion is in the 93rd percentile of $□^2$ values. Ancient deletions like these are common and old, and also exhibit high population differentiation. Collectively, we argue that such adaptive maintenance of ancient, functional alleles may be due to varying selection trends across geographies and time.

CONCLUSION

This manuscript asked whether adaptive forces have maintained ancient deletion polymorphisms in humans. We provide evidence supporting the idea that AMHs exhibit an excess of ancient polymorphisms, relative to the neutral expectation. Using simulations and empirical data, we provide evidence for the notion that balancing selection is likely a considerable force shaping the extant functional deletion polymorphisms. We show that when functionality is defined conservatively, ancient deletions are disproportionately functional, compared to non-ancient deletions. In fact, 50% of such functionally relevant deletions are ancient. Additionally, ancient deletions are enriched for association related to metabolism and inflammatory response. Our results suggest that classical overdominance may not be the prime mode of balancing selection affecting the evolution of ancient deletions. Instead, geographically and temporally variable, as well as frequency-dependent selection may underlie the maintenance of ancient functional deletions. We also provide insights about the mechanisms by which a deletion could confer function: in addition to previously defined functional effects of deletions such as the loss of a gene's function and regulation of expression levels, we highlight multiple instances where the presence of ancient deletions lead to the expression of different isoforms. Overall, our study contributes to the growing body of evidence supporting the notion that balancing selection may be an important force in the evolution of genomic variation shared among human populations. These ancient variants are an important part of our legacy as a species: something we all share.

DATA AVAILABILITY

All data generated can be found in the supplementary files. The codes that we used to generate our datasets and simulations can be found either in Methods or on our GitHub page **(https://github.com/GokcumenLab).** The consolidated file including all significant UKBioBank GWAS associations for SNVs is available at our FigShare (https://figshare.com/articles/dataset/Table_S3_for_Aqil_et_al_2022/19606192).

SUPPLEMENTARY FILES

Supplementary File 1: Deletion genotypes in archaic hominin genomes
Supplementary File 2: Haplotype-level analysis of deletion polymorphisms
Supplementary File 3: Curated GWAS meta-categories

FUNDING

O.G. acknowledges support from the National Science Foundation (Grant No. 2123284). L.S. is funded by a Sir Henry Wellcome fellowship (220457/Z/20/Z). This research was funded in part by the Wellcome Trust. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

ACKNOWLEDGEMENTS

We thank Dr. Victor Albert and Dr. Vincent Lynch for their careful reading of this manuscript. We acknowledge Petar Pajic and Charikleia Karageorgiou for their insightful discussions throughout the development of this project.

ETHICS

All data used in this study have been previously published in peer-reviewed journals.

COMPETING INTERESTS

We have no competing interests

METHODS

Proportion of ancient polymorphisms in neutral simulations versus observation:

We compared the proportion of ancient polymorphisms in randomly chosen YRI SNVs against the expected distribution of this proportion under neutrality. We obtained SNVs from the 1000G phase-3 vcf files (1000 Genomes Project Consortium et al., 2015) for the analysis described above. Using a script written in AWK (Aho et al., 1978), we subsetted the vcf files to retain only those biallelic SNVs that contained information about the ancestral/derived status of the two alleles. We used the --keep option in vcftools (Danecek et al., 2011) to retain individuals only from the Yoruba population. We then used the --mac option to retain only those SNVs for which the minor allele count was greater than 1. The allele-count filter was used to exclude singletons which could create spurious results for the linkage disequilibrium analysis described below. On the resulting vcf files, we used the SelectVariants tool with the --select-random-fraction option in GATK (Van der Auwera and O'Connor, 2020) to randomly retain 0.25% of the variants. This resulted in a set of 38,231 SNVs. Next, we investigated whether the random SNVs are in linkage disequilibrium with any other SNVs in their vicinity. In particular, we used the --hap-r2-positions and --min-r2 option in vcftools, along with the ancestral/derived status of alleles, to retain only those random SNVs wherein the derived allele was in linkage disequilibrium ($r^2 > 0.9$) with the derived allele of another SNV within a 50 kb radius of the random SNV. This resulted in 28,491 SNVs. We only retained the random SNVs with variants in LD in their vicinity to rule out cases of recurrence of the SNV between AMHs and archaic hominins, as described below. There seems to be one bias that may be introduced by eliminating polymorphisms with no SNVs in LD in AMHs: this would bias our analysis to regions with low recombination rates. In these regions, we would expect higher background selection due to the Hill-Robertson effect, leading to a deflation in the proportion of ancient polymorphisms. Since we observe a larger than expected proportion of ancient polymorphisms despite this bias, we can conclude that this bias only makes our analysis more conservative.

We then inspected the 28,491 random SNVs to see whether the derived alleles are shared with any of the four high coverage archaic genomes (Altai Neanderthal, Vindija Neanderthal, Chagyrskaya Neanderthal, and the Denisovan). We found that 4,616 SNVs (16.2%) had their derived allele shared (either homo- or heterozygously) with at least one of the four archaic hominins. Since we are only interested in polymorphisms older than 700,000 years, we want to focus only on SNVs where the derived allele is shared with archaic hominins by *common descent*. We thus excluded the SNVs where the derived allele emerged independently (recurrence) in AMHs and archaic hominins in the following way. For each of the 4,616 core SNVs with shared derived alleles in archaic hominins, we tested whether any of the derived alleles in LD with the core derived allele is also present in any of the same archaic hominins which carry the derived allele for the core SNV itself. If any of the archaic hominins contain both the derived allele of the core SNV along with at least one derived allele in LD with the derived allele of the core SNV, we classify that core SNV as "shared by common descent." If any core SNV has a derived allele shared with archaic hominins but not a derived allele in LD with the core SNV, we classify the core SNV as "recurrent." This approach yielded 3,894 SNVs (13.7%) wherein the derived allele is shared with at least one of the archaic hominins by common descent.

In order to investigate whether this percentage (13.7%) of ancient polymorphisms (700,000 years old polymorphisms) is significantly higher than the neutral expectation, we wanted to calculate the same percentage for a set of neutrally simulated SNVs. We used the program *ms* (Hudson, 2002) to neutrally simulate a set of 20,000 (independent, and therefore freely recombining) variants 2,000 times using various models. All of these models included 216

haploid genomes representing YRI (matching the YRI sample size in 1000G dataset) and 2 haploid genomes representing each of the four archaic hominins. For every model, and for each of the 2,000 runs we calculated the proportion of SNVs present in the YRI with minor allele-count > 1 that were shared with at least one of the archaic hominins. Thereby, we obtained a distribution of the proportion of ancient polymorphisms under each of the models.

We used a total of 36 models. The models varied by three parameters: N_e of Yoruba/AMH, N_e of archaic hominins, and the time of divergence between the AMH and archaic hominin lineages. For 27 of these models, N_e for humans was constant across time (ranging from 10,000 to 30,000 across models); for 9, it varied over time (Speidel et al., 2019) (**Figure 1B**). The N_e of the archaic hominin lineage was constant over time in each model ranged from 1,000 to 3,000 across models; and the time of divergence between AMH and archaic hominins ranges from 500 to 700 kya across models (Bergström et al., 2021; Mafessoni et al., 2020; Meyer et al., 2012; Prüfer et al., 2014). In the main text, we focused on the model with variable AMH-N_e (Speidel et al., 2019), using the well-accepted archaic hominin N_e of 1,000 and a divergence time of 700 kya; we referred to this as the "base model". For each model, we assumed that Denisovans diverged from Neanderthals ~400,000 years ago, the Altai Neanderthal lineage separated from the Vindija-Chagyrskaya lineage ~130,000 years ago, and the Vindija lineage separated from the Chagyrskaya lineage ~90,000 years ago (Bergström et al., 2021; Gravel et al., 2011; Mafessoni et al., 2020; Meyer et al., 2012; Prüfer et al., 2014). Moreover, the generation time was assumed to be 29 years (Fenner, 2005; Langergraber et al., 2012; Li and Durbin, 2011).

To test if the empirical excess of ancient polymorphisms is more pronounced among low derived allele frequency variants (due to recurrence), we repeated our analysis using the base model (variable AMH N_e, Archaic N_e = 1,000, divergence time = 700 kya) for simulations, dividing the empirical and simulated SNVs into derived allele frequency bins. To ensure there are enough variants in each frequency bin, we used a larger set of randomly chosen YRI SNVs. In particular, we used 300,000 SNVs that have a minor allele count > 1 and at least one variant in LD ($r^2 > 0.9$) with them. Using the same method as described above, we identified the SNVs that are shared with archaic hominins by common descent. We used the base model to simulate 1500 runs of 400,000 SNVs. Thereupon, we divided the empirical and simulated SNVs into 10 derived allele frequency bins of uniform length, and compared the simulated distribution of the proportion of ancient polymorphisms with the observed proportion in each bin **(Figure 1—figure supplement 1)**. Moreover, to gauge whether recurrence at CpG sites leads to the excess of ancient polymorphisms, we subsetted the 300,000 SNVs to retain only the 21,402 A⇄T SNVs, and calculated the proportion of ancient polymorphisms therein.

Additionally, we also performed another set of neutral simulations, this time with structure introduced in the population ancestral to the archaic hominins and AMHs. This too was done using Hudson's *ms*. This was done using a constant size model with YRI/AMH N_e = 14,474 (Gravel et al., 2011), Archaic N_e = 1,000, and divergence time = 700 kya. The effective population size for each of the subgroups in the population ancestral to both AMHs and archaic hominins was set to 10,000. We define $m_{i,j}$ as the fraction of subgroup i that is formed by the migrants of subgroup j in each generation, where $i \neq j$ and $i,j \in \{1,2,3\}$. For all i *and* j, where $i \neq j$, we set $m_{i,j} = m$. The program *ms* takes this parameter in the form of $M = 4Nm$ (where N = 10,000 is the effective population size of each subgroup). We performed simulations for 10,000 different values of M chosen uniformly on the log scale from the range $(0.01,100)$. This is akin to running simulations using 10,000 different values of m in the range $(0.25 \times 10^{-7}, 0.25 \times 10^{-2})$.

For each *m*, 1000 variants were simulated. Thus, for each *m*, we calculated the percentage of variants in Yoruba (with allele-count >1) wherein the derived allele was shared with the archaic hominins. The proportion of allele-sharing in simulations equaled or exceeded the proportion (13.7%) observed in real-life at approximately $m \leq 0.0075\%$.

Identifying deletions in archaic genomes

The identification of deletions in archaic hominins was predicated on the concept that a deletion in an archaic hominin would correspond to a low read depth in the window of deletion in the hominin's genome. We started with two main types of input files: 1) The VCF file for the 1000 Genomes Phase 3 dataset; and 2) A BAM file for each high-coverage archaic hominin.

The 1000G phase-3 VCF file was obtained from https://www.internationalgenome.org/data. This includes 84.4 million variants from 2504 individuals across 26 populations.. The file was then converted to a BED file (with three tab-separated columns representing chromosome numbers, start positions, and end positions of deletions) using a script written in AWK. This VCF file was filtered to retain only biallelic autosomal deletions. This amounted to 32,154 deletions. We genotyped all these deletions in the four high-coverage archaic hominin genomes. (Note that in the main text, we focused only on 4,863 deletions with both allele-count > 1 in YRI, CEU, and CHB combined, and at least one SNV in LD.)

The sequence files for archaic genomes mapped to hg19 were obtained from https://www.eva.mpg.de/genetics/genome-projects.html?Fsize=0%2C%252%27A%3D0. These BAM files containing mapping information (such as the start and end coordinates of the part of the genome to which a read maps) were converted to BED files. This was done using the bamToBed command in the bedtools module (Quinlan and Hall, 2010). We then used the two types of BED files to count the number of reads for each archaic genome that mapped to a region of the genome that is polymorphically deleted in AMHs. In order to achieve this, we used the intersectBed command with the -c option within the bedtools module. This command counts the number of reads in an archaic genome that intersects with the region of the genome harboring a deletion polymorphism in AMHs.

Next, for every archaic genome, we normalized the number of reads at each window of deletion by the size of the window.

$$\textit{Normalized Read Depth} = \frac{\#\ of\ reads\ intersecting\ the\ window\ of\ deletion}{Size\ of\ the\ window\ of\ deletion}$$

For each archaic genome, we wanted to calculate the Z-scores of the normalized read depths across all windows of deletion, and classify a window as a deletion if the normalized read depth was below a certain threshold. To prevent outliers from affecting measures of central tendency and spread, and therefore the Z-score threshold, we use the modified Z-score to classify a region as deleted or non-deleted in an archaic genome. The modified Z-score uses median (as opposed to mean) and median absolute deviation (as opposed to standard deviation) to calculate the Z-score. For a given archaic genome, the modified Z-score of the normalized read depth at the i^{th} window of deletion is given by:

$$ModZ_i = \frac{r_i - Median(R)}{MedianAbsoluteDeviation(R)}$$

where r_i denotes the normalized read depth at a given window of deletion, and R denotes the random variable representing normalized read depth.

(Iglewicz and Hoaglin, 1993) have suggested that a threshold ± 3.5 is reasonable for outlier detection using modified z-scores. Nevertheless, for our purposes, we used a more conservative threshold of -5, which we deemed more-appropriate based on spot-checking. For example, if the modified Z-score (of the normalized read-depth) at a window of deletion was less than -5 in the Vindija Neanderthal, that window was classified as deleted in the Vindija Neanderthal. The distributions of these modified Z-scores across windows of deletions for the four high-coverage archaic genomes are illustrated in **figure 2—figure supplement 1.** All calculations downstream of obtaining the raw numbers of reads from archaic genomes intersecting with the windows of deletion, were performed using a script that we wrote in R. The read-depth analysis was done using all 32,154 AHM deletions (results for the status of these deletions in the four high-coverage archaic genomes are available in **Supplementary File 1**).

Identifying SNVs that are in linkage disequilibrium with deletion polymorphisms in AMHs:

We subsetted the 1000G phase-3 VCF files (there is a separate file for each chromosome) obtained from https://www.internationalgenome.org/data to retain individuals only from CEU (n=103), CHB (n=99), and YRI (n=108) populations. This filter was applied using the `--keep` option in the module VCFtools (Danecek et al., 2011). All variants that had a minor allele count of less than 2 were eliminated using the `--mac` filter in VCFtools. We used the resulting VCF files to identify SNVs in LD with each of the autosomal biallelic deletions (with minor allele-count > 1 in YRI, CEU, and CHB combined) within a 50 kb radius of the deletion. We did this using the `--hap-r2-positions` and `--min-r2 0.9` flags in VCFtools. For each autosomal biallelic deletion with allele-count > 1, this gave us a list of SNVs in LD with the deletion with $r^2 > 0.9$, if such SNVs existed. At least one such SNV in LD existed for 4,863 deletions. We called this set of deletion the “deletion dataset” and based all our downstream analysis on it.

It is important to describe why we only focused on deletions with identifiable variants (most of them SNVs) in LD with them. We can only eliminate potentially introgressed deletions by checking whether at least one of the SNVs in LD with a deletion is already known to be introgressed. Moreover, we can confirm whether a deletion shared between archaic hominins and AMHs is identical by descent (thereby eliminating recurrence) by checking whether the same SNVs accompany the deletion in AMHs and archaic hominins. This filtering would not be possible if our deletions were not flanked by variants in LD with them. A shortcoming of this approach is that it fails to capture balanced deletions that are not in LD with at least one SNV.

Eliminating instances of recurrence and introgression

We found that 575 human polymorphic deletions are also present in at least one archaic hominin genome. In order to ensure that we do further analysis only on deletions that are shared with archaic hominins by common descent, we wanted to eliminate shared deletions that were recurrent or introgressed.

We removed recurrent deletions (those emerging in AMHs and archaic hominins independently) by retaining only those shared deletions for which at least one allele in LD with the deletion in AMHs was also present in at least one archaic genome that harbored the deletion. To do this, we needed to know whether variants in LD with deletions are present or absent in the archaic genomes. We started with two types of inputs: 1) VCF files for each of the archaic genomes

(mapped to hg19) and 2) a file containing all the variants (SNVs) in LD with polymorphic deletions in AMHs. We filtered the VCF files to include only the SNVs in LD with shared deletions. This was done using the --positions flag in VCFtools. The presence or absence of every variant in LD was then determined using the vcf files for the archaic hominins. The procedure was implemented using an AWK script. 53 shared deletions were classified as "recurrent" using this approach.

In order to eliminate introgressed shared deletions, we used the results published by (Taskent et al., 2020). In their study, the authors had identified introgressed haplotypes in Eurasians using the S* statistic. They had also published a list of S*-significant SNVs that characterize introgressed haplotypes. We stamped out the shared deletions that were both absent in Yoruba and for which at least one allele in LD was among the S*-significant variants listed in the study mentioned above. We thus eliminated 92 deletions that were likely introgressed from archaic hominins into AMHs.

Age of deletions and allele frequency trajectories

We estimate the ages of the deletions in the *deletion dataset* using two methods: 1) Human Genome Dating database (https://human.genome.dating/download/index); and 2) RELATE (Speidel et al., 2019).

The Human Genome Dating database (https://human.genome.dating/download/index) hosts age estimates for over 45 million single nucleotide variants (SNVs) (Albers and McVean, 2020). This database reports multiple age estimates for each SNV. We used the median age estimate calculated using the joint clock. Since this database only includes age estimates for SNVs (and not for deletions), we could only date a deletion if the dating database contained the age estimate for at least one of the variants in LD ($r^2 > 0.9$) with the deletion. If age estimates were available for only one variant in LD, the same age estimate was assigned to the deletion. If age estimates were available for more than one variant in LD with the deletion, we used the highest age estimate, which may be inaccurate in certain cases.

Relate is a method that estimates genome-wide genealogies and can be used to infer the age of a variant (Speidel et al., 2019). We used Relate to infer the ages of the deletions in the *deletion dataset.* To this end, we used previously inferred genome-wide genealogies for samples of the SGDP dataset (Mallick et al., 2016; Speidel et al., 2021), available from https://www.dropbox.com/sh/2gjyxe3kqzh932o/AAAQcipCHnySgEB873t9EQjNa?dl=0. For each deletion, we used SNVs in LD where the derived allele was tagging the deletion at an r^2 exceeding 0.9 and calculated the mean age of such SNVs to date each deletion.

To quantify allele frequency variation, we computed the ratio of lineages carrying the derived allele by the total number of lineages remaining at 5,000 years and 50,000 years before present, but only if the number of lineages remaining at 50,000 years exceeded 10% of the present-day sample size. We then standardized the allele frequency change stratified by present-day allele frequency, by calculating the mean and standard deviation given present-day frequency. Finally, we squared this standardized allele frequency change to obtain our statistic $\square^2$, which is expected to have a Chi-squared distribution with one degree of freedom under neutrality, and smaller values for more stable trajectories. This approach was inspired by (Edge and Coop, 2019), who used a similar approach to quantify polygenic positive selection using genealogies.

Beta measure

We used a recent and robust measure of balancing selection (Siewert and Voight, 2017), std□2, to investigate whether ancient deletion are enriched for targets of balancing selection. A high std□2 for a variant is indicative of balancing selection.

In our study, we estimated std□2 for the deletions in the deletion dataset using SNVs in LD with them. We did this for the CEU, CHB, and YRI population separately. The std□2 scores for SNVs are publicly available (https://github.com/ksiewert/BetaScan) for the CEU, CHB, and YRI populations. For each deletion in each of these populations, we obtained the std□2 values for variants in LD with deletions whenever they were available. For a given deletion, when the std□2 values were available for more than one variant in LD with the deletion, we used two approaches to estimate the std□2 for the deletions. In the first approach, we used the highest std□2 among the LD variants as the estimate for the std□2 for the deletion. We call this BETAMAX. In the second approach, we focused on the std□2 values for the SNVs that were in LD with the deletion with the highest r^2 value. If multiple SNVs were in LD with the deletion with the highest r^2 value, we used the std□2 value of the SNV that was closest to the deletion among these SNVs as the estimate for std□2 for the deletion. We call this BETAPRIME. We performed this process for YRI, CEU, and CHB populations separately to arrive at std□2 estimates for deletions in our deletion dataset in each of these three populations. Using both BETAMAX and BETAPRIME gave us similar trends across populations. These values are available in **Supplementary File 2**.

Ascribing phenotypic relevance to deletion

We used two criteria to ascribe phenotypic relevance to deletions: 1) intersection of the deletion with at least one exon; and 2) association of a SNV in LD with the deletion with a GWAS trait.

In order to identify deletions that intersect with exons, we started with the genome annotation file download from https://hgdownload-test.gi.ucsc.edu/goldenPath/hg19/bigZips/genes/hg19.refGene.gtf.gz. Using an AWK script, this GTF file was then converted to a BED file containing five columns: 1) annotation's chromosome number; 2) annotation's start position; 3) annotation's end position; 4) gene name; and 5) type of feature. Only the rows wherein the type of feature was "exon" and the chromosome number was between 1 and 22 were retained. All repeated entries (rows) were eliminated. The resulting file contained only columns 1) to column 4). We then used a BED file containing information about the AMH deletions in our deletion dataset and the BED file mentioned above to identify deletions spanning exons. This was done using the intersectBed option with -wa and -wb flags in the BEDtools module. On the resulting file, we used the groupby tool with the "-o freqdesc" flag in the BEDtools module in order to obtain a file containing the names of the genes (and the number of exons within each intersecting gene) that overlap the deletions. 243 (5%) of the 4,863 deletions in the deletion dataset were exonic.

The second method to ascribe phenotypic relevance to deletions was to use results from previously published Genome Wide Association Studies (GWAS). We used a publicly available catalog of GWAS results based on the UK BioBank data (http://www.nealelab.is/uk-biobank/). In particular, we used data for 4,113 traits. For each trait, we used data that produced results using both sexes. For continuous traits, we used the raw version of the data, as opposed to the inverse rank normalized version. For each trait, only those SNVs were retained that were associated with the phenotype with a p-value less than 10^{-8}. We are making available a consolidated table with all statistically significant associations from this dataset (https://figshare.com/articles/dataset/Table_S3_for_Aqil_et_al_2022/19606192). We hope this

will make it easier for the scientific community to use GWAS results than the currently available datasets which store associations with each trait in a different table. Then, for each of the 4,863 deletions in our deletion dataset, we checked if any of the SNVs in LD were among the SNVs that were significantly ($p < 10^{-8}$) associated with a phenotype. We then obtained the phenotype that was associated with one of the SNVs in LD with the lowest p-value, and ascribed it to the deletion. Thus, 433 (8.9%) of the 4,863 deletions in the deletion dataset had phenotypic associations.

Enrichment analysis for ancient deletions

We performed enrichment analyses for phenotypic relevance and length among ancient deletions using variants with a pooled frequency > 5% in YRI, CEU, and CHB combined. First, we investigate whether a higher proportion of ancient deletions, relative to non-ancient deletions, have phenotypic relevance. To this end, we defined phenotypic relevance in three ways: 1) GWAS association, 2) exonic overlap, and 3) both GWAS association and exonic overlap. For each definition, we first calculated the observed proportions of phenotypic deletions among both ancient and non-ancient categories in turn. Then we shuffled the “ancient” and “non-ancient” labels among the deletions in 10,000 permutations, calculating the proportion of phenotypic deletions among both ancient and non-ancient labels for each permutation. Using the number of permutations in which the difference in proportions of phenotypic deletions was more extreme than the difference in observed proportions, we obtained an empirical p-value for phenotypic enrichment among ancient deletions.

We also investigated whether certain phenotypic categories are overrepresented in ancient deletions relative to non-ancient deletions. For this, we used the UKBioBank traits associated with the deletions. In total, 1,675 traits were associated with the deletions in the deletion dataset. We manually placed each of these traits into one of 18 categories such that any deletion could be associated with one or more phenotypic categories (**Supplementary File 3**). Only deletions with a pooled frequency > 5% in YRI, CEU, and CHB combined were retained for analysis. For each phenotypic category, we obtained the proportion of deletions associated with that category among ancient and non-ancient deletions. We then shuffled the “ancient” and “non-ancient” labels in 10,000 permutations. Just as above, we used the number of permutations in which the difference in proportions of deletions associated with a phenotypic category was more extreme than the difference in observed proportions, and we obtained an empirical p-value.

Now, we turn to length enrichment. We calculated the 2nd, 5th, 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th, 95th, and 98th length percentiles for deletions in ancient and non-ancient categories. We calculated the differences in corresponding percentiles in ancient and non-ancient deletions. Again, we shuffled the “ancient” and “non-ancient” labels into 10,000 random permutations, calculating the differences in corresponding percentiles in ancient versus non-ancient deletions for each permutation. This gave us empirical p-values for differences in the length of ancient and non-ancient deletions at various percentiles.

Simulations to identify signatures of overdominance

We set out to identify signatures associated with a locus that has evolved under overdominance.

Methodologically, we approached the problem of separating overdominance from neutrality on two fronts: (i) The trajectory of the allele-frequency of the mutation conditioning on the age of the mutation, and (ii) the patterns of neutral polymorphisms around the so-called focal mutation which has evolved either under overdominance or neutrally (with the same age of the mutation). Given that a mutation under overdominance (heterozygote advantage) may be at intermediate frequency, we also (iii) studied simple coalescent simulations conditioning on the existence of at least one SNV within a certain frequency range (e.g. 50%-60%). The age of a mutation is crucial for the study of balancing selection. We considered two values for the age of a mutation: (i) 10,000 generation and (ii) 40,000 generations old. In the first case, the age of the mutation corresponds to 10,000 x 29 years = 290,000 years old mutation. In the second case, the mutation is 1,160,000 years old. Conditioning on the age of the mutation, we generated allele frequency trajectories of the mutation, *i.e.,* the frequency of the mutation at each time point from its onset until the present-day. For the overdominance scenario, we used the software *trajdemognpops,* implemented using tools from *ms* and *mssel* (kindly provided by R.R. Hudson), to generate trajectories of a mutation under overdominance. The dominance coefficient is characterized by a large value (here, $h = 10$) in order to assign a benefit for the heterozygote. Thus, the fitness for genotypes at a biallelic locus is by:

Aa: (1.+sh), where s is the selection coefficient and h the dominance coefficient. Here, s = 0.005 and h = 10, *i.e.,* the heterozygote Aa has a fitness value 1.05. If we had considered codominance (as opposed to overdominance), we would have set h = 0.5.
AA: 1 + s, thus, the fitness for the AA is 1.005
aa: 1.

Given the trajectory of a mutation, a population is split in two kinds of genotypes. The haplotypes carrying the mutation (or the derived allele) and the haplotypes that carry the ancestral allele. Each *neutral* (i.e. passenger mutation) can change “population” (genetic background) by recombination. Therefore, the coalescent in this case is described as a structured coalescent of two populations (a population with the derived allele and another population with the ancestral allele) that communicate between themselves via recombination. The size of each population is determined at each time point by the trajectory of the derived allele.

In order to understand the effect of the age of the allele (and also to test the mssel code for correctness), we performed coalescent simulations (using Hudson’s ms) conditioning on the presence of at least one SNV at frequency within a given range. Since we are interested in the mutations that are approximately at 50% frequency in the population, we conditioned on the presence of derived mutations in the sample in 22-28 (out of 50) haplotypes. This set of simulations are called *pseCoal.* For each simulated dataset, we calculated the relevant population genetics statistics available through Comus (Papadantonakis et al., 2016) including number of segregating sites, θ, □, Tajima’s D, ZnS, Fay and Wu’s H, dvk, and dvh. Then, we conducted a summary of all these statistics using PCA. The goal is to understand whether the different scenarios can get separated by using polymorphic patterns.

Conceptual and methodological concerns:

Human polymorphisms wherein derived allele is shared with archaic hominins are older than 700,000 years

AMHs and archaic hominins are estimated to have diverged around ~700,000 years ago. Therefore, if a human polymorphism has been maintained for more than ~700,000 years, it was also present in the common ancestral population of AMH and archaic lineages. It follows that if a polymorphism (the presence of both ancestral and derived alleles in a population) is present in AMHs and archaic hominins, then (barring recurrence and introgression), by parsimony, the polymorphism was also present in their common ancestral population (**Figure 1A**). Thus, a polymorphism that is shared by common descent between AMHs and archaic hominins, has been maintained for over ~700,000 years. Moreover, because the ancestral allele is fixed in chimpanzees by definition, AMH polymorphisms wherein the derived allele is fixed in archaic hominins were also present (in the polymorphic state) in the common ancestral population of archaic hominins and AMHs. In essence, AMH polymorphisms for which archaic hominins carry the derived allele (fixed or polymorphic) have been maintained for more than 700,000 years.

Why use SNVs (as opposed to deletion polymorphisms) for comparison of the real-life proportion of ancient polymorphisms against neutrally simulated SNVs.

It is worth explaining why we used SNVs, instead of deletions – the class of variants that we are interested in – for comparing observed versus simulated (under neutral conditions) proportions of ancient polymorphisms. Deletions are not suitable for such a comparison because, in general, they are targets of strong negative selection (Conrad et al., 2006; Kondrashov, 2017; Lin et al., 2015; Lin and Gokcumen, 2019). Thus, negative selection would have purged a large proportion of deletions that emerged in the common ancestral population of AMHs and archaic hominins. It follows that a smaller proportion of AMH polymorphic deletions than expected under neutral conditions will be shared by common descent with archaic hominins. Even if balancing selection were inflating the proportion of deletions that are shared with archaic hominins, it would not be observable due to the opposite deflationary effect of negative selection. Since negative selection is not as strong a force in the evolution of SNVs as it is for deletions, this problem would not be as pronounced if we used SNVs instead of deletions for testing this premise. Hence, our choice of SNVs for this analysis. In the main text,, we have expanded on the idea that deletions are targets of stronger negative selection.

The vast majority of human deletions are derived relative to chimpanzees

In order to identify deletion polymorphisms that have been maintained in the human lineage for over 700,000 years, we focused on deletions that were present (either polymorphically or fixed) in the four high coverage archaic genomes. This technique would work only for deletions that were derived in humans, relative to chimps. However, variants that have been called as deletions (wherein deletion is the alternative allele) in the 1000 Genomes project may, in fact, be human-specific insertions, such that the reference allele (non-deletion) is derived. To investigate how common this situation is among the 4,863 deletions in our dataset, we lifted over the the coordinates of the deletions from hg19 on to the Chimpanzee reference panTro3 using the LiftOver tool in UCSC Genome Browser (Kent et al., 2002). If the liftover for a deletion fails on account of the window being completely or partially deleted in the Chimpanzee reference, it is indicative of the region being a human-specific insertion. The liftover failed for this reason for only 184 (3.8%) of the 4,863 deletions. Therefore, the vast majority of deletions are, in fact, derived relative to chimps.

1012

SUPPLEMENTARY FIGURES

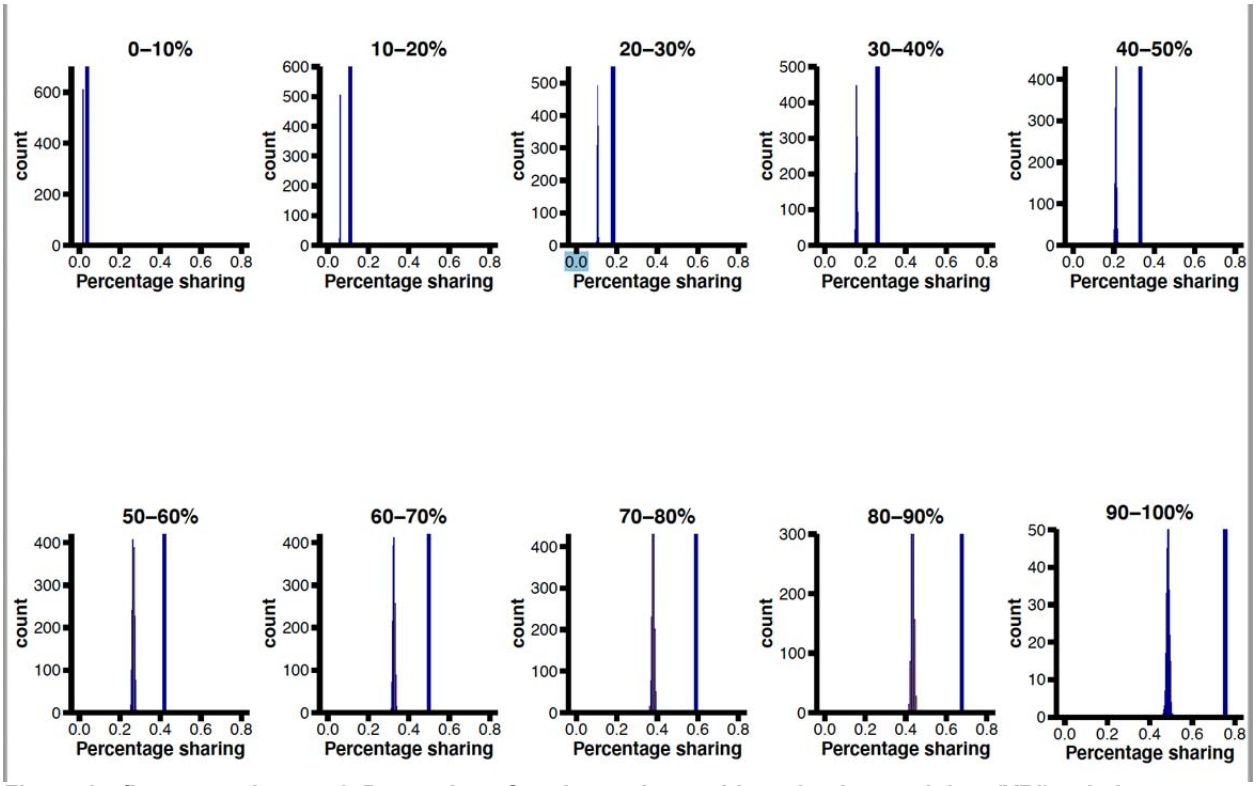

Figure 1—figure supplement 1. Proportion of ancient polymorphisms in observed data (YRI), relative to neutral expecation ("base" model parameters) in various derived allele frequency bins. The vertical blue line indicates the observed sharing, while the distributions are simulated expectations. The excess of ancient polymorphisms in observed data becomes more pronounced at higher derived allele frequencies.

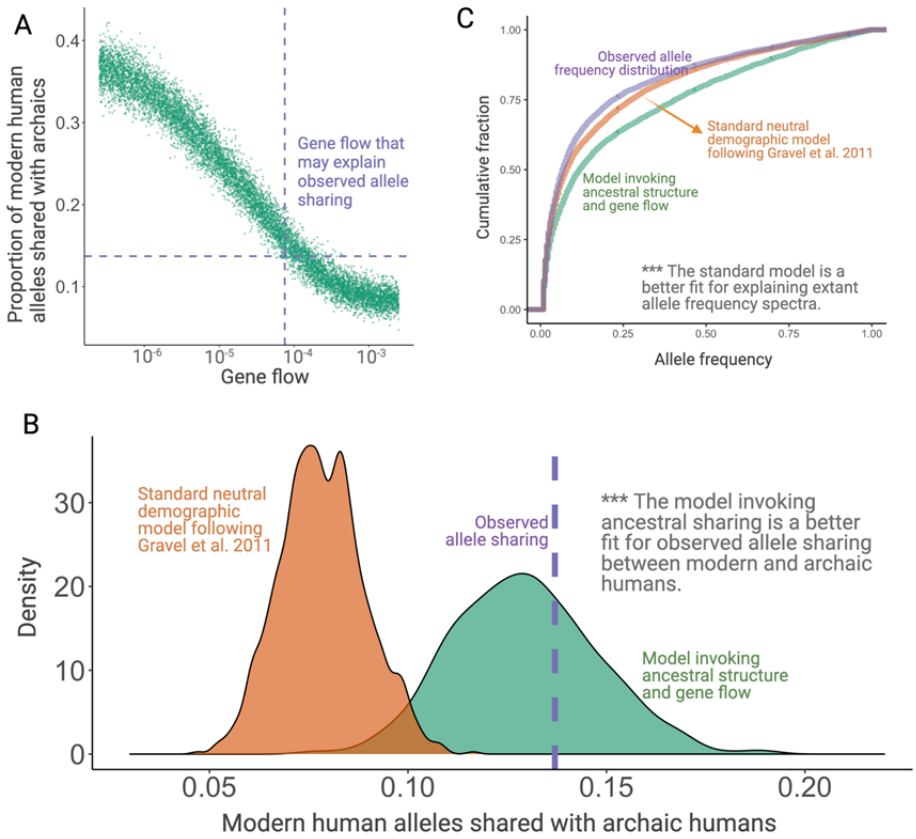

Figure 1—figure supplement 2. A) Results from simulations invoking structure in the population that was ancestral to both AMHs and archaic hominins. In this model, we have three latent subgroups in the ancestral populations. The x-axis refers to the fraction of each subgroup that is formed by the migrants of each of the other subgroups in each generation. **B.** Proportion of ancient polymorphisms in YRI. The purple line is the observed proportion of ancient polymorphisms in Yoruba (YRI). The green and orange density plots indicate the distribution of the proportion of ancient polymorphisms in neutral simulations with and without ancestral structure, respectively. We used Gravel et al. parameters for these simulations. **C.** Comparison of the allele-frequency spectra of simulated SNVs with observed SNVs. The purple, orange, and green lines represent allele frequency spectra in the YRI population using actual SNVs, neutral simulations without ancestral structure, and neutral simulations invoking ancestral structure, respectively.

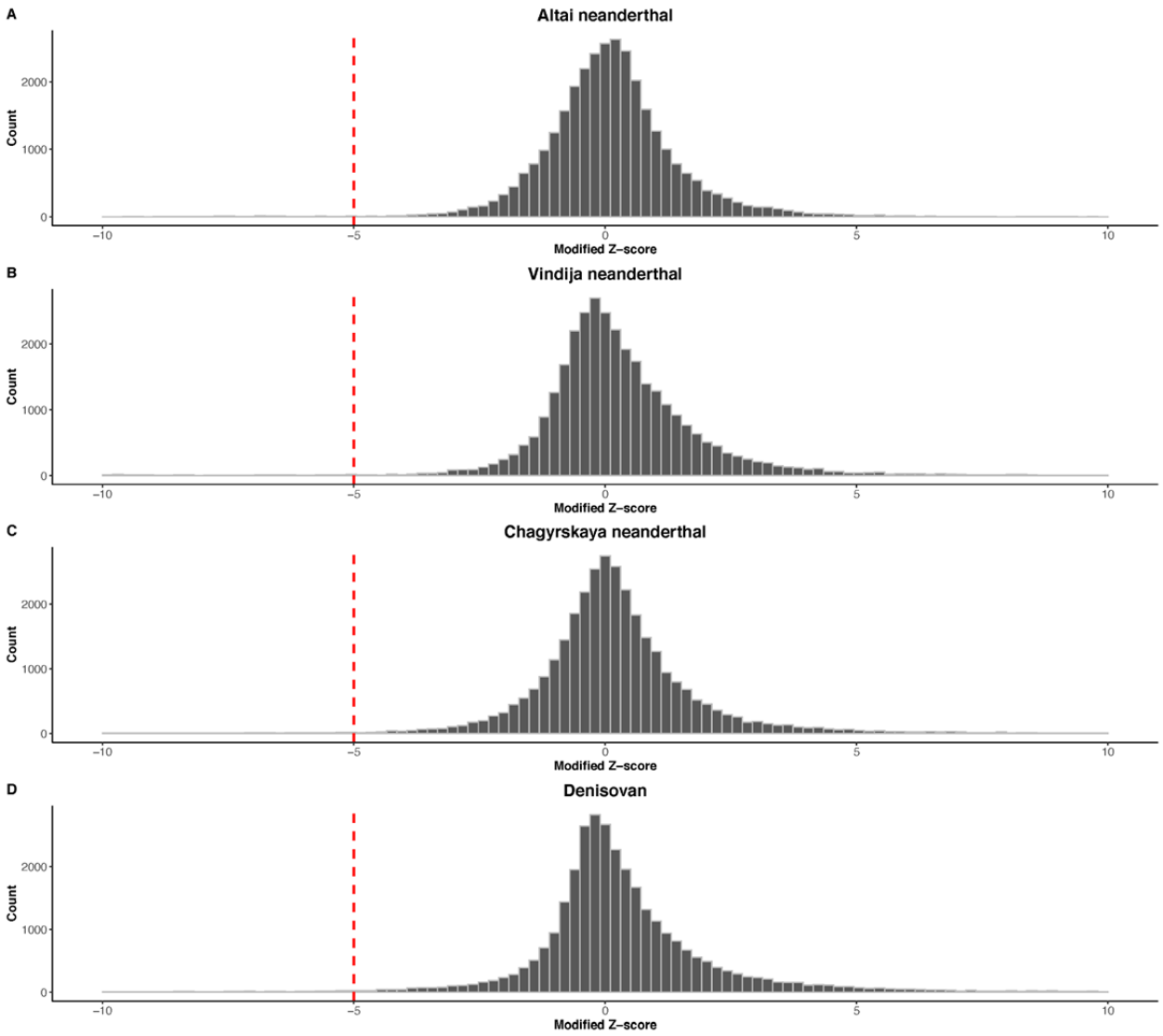

Figure 2—figure supplement 1. Read-depth-based pipeline to identify deletions in archaic hominin genomes: Distribution of the modified Z-score of the read-depth across the 32,154 biallelic AMH deletions in the archaic genomes. **A.** Altai neanderthal. **B.** Vindija neanderthal. **C.** Chagyrskaya neanderthal. **D.** Denisovan.

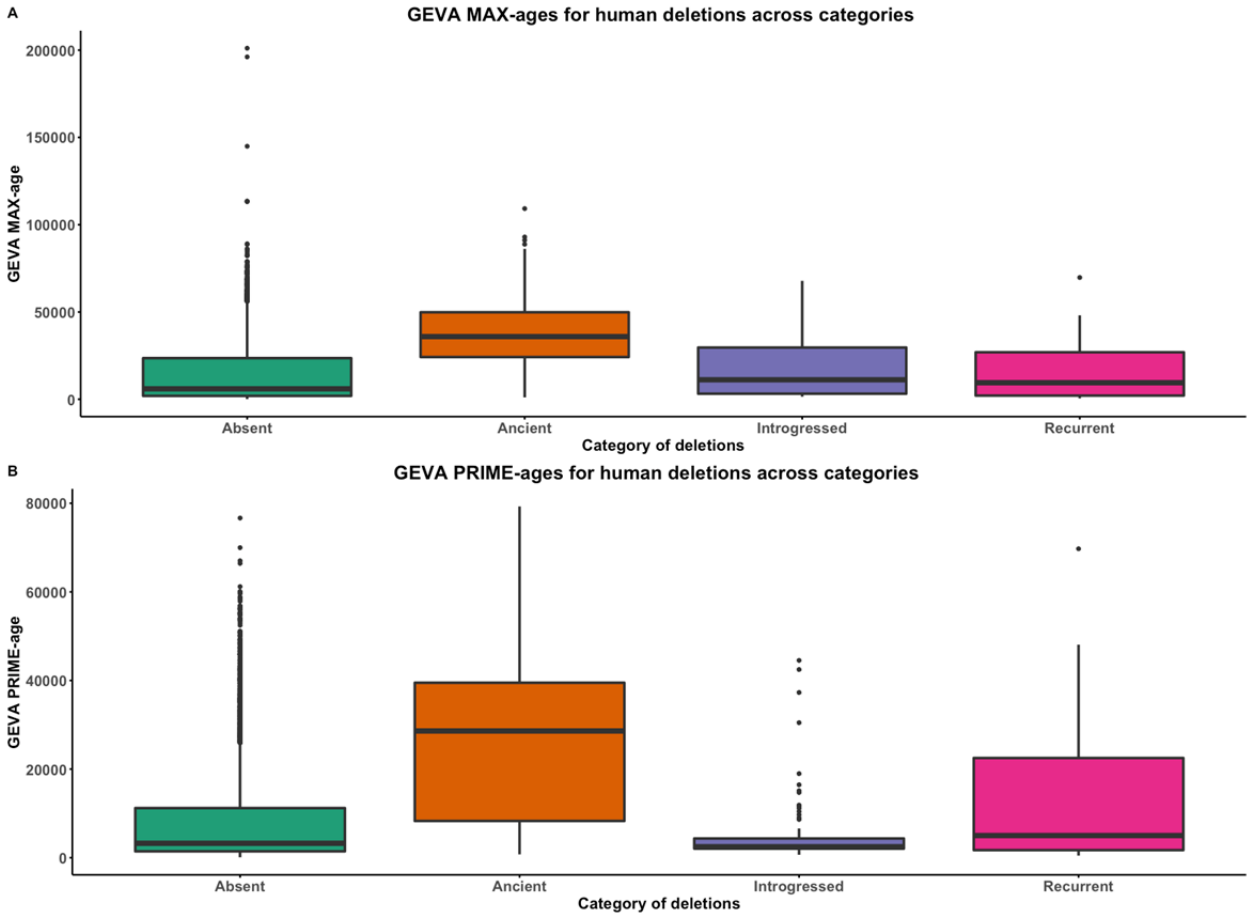

Figure 3—figure supplement 1. GEVA ages of deletions across categories. Absent denotes polymorphic deletions in AMHs that are not present in any of the four high coverage archaic genomes. Introgressed refers to the shared deletions that were introgressed from archaic hominins into AMHs. Recurrent refers to the shared deletions that emerged independently in the AMH and archaic hominin lineages. Ancient refers to the AMH deletions that are shared with archaic hominins by common descent. **A.** GEVA PRIME-ages. **B.** GEVA MAX-ages. With both GEVA PRIME and GEVA MAX measures, we observe that ancient deletions are significantly older than absent, recurrent, and introgressed deletions. This implies that our pipeline to identify ancient deletions is sound.

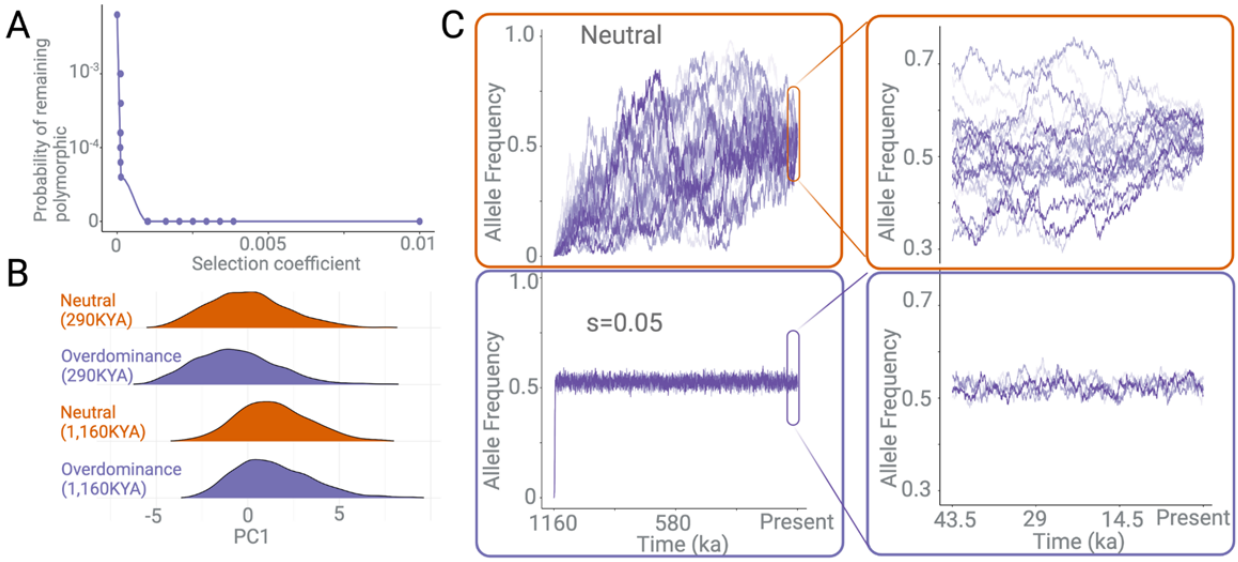

Figure 7—figure supplement 1. A. The probability of a polymorphism persisting in the population for 1,000,000 years under different negative selection pressures. **B.** Density plots of the first principle component of multiple summary statistics based on variants simulated under neutral versus overdominance (s=0.05) scenarios. This is shown for two categories of variants: 1) those that emerged 290 kya, and 2) those that emerged 1,160 kya. There is no discernible difference between overdominance and neutrality within the time frame of these simulations. **C.** The allele frequency trajectories of variants over 1,000,000 years, under neutrally (top), versus under overdominance (bottom). The x-axis represents the time since the emergence of a variant in years, assuming a 29 year generation time. The right panel is a zoomed-in version of the same allele frequency trajectories in the last ~50 thousand years.

REFERENCES

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. 2015. A global reference for human genetic variation. *Nature* **526**:68–74.

Abdul-Rahman F, Tranchina D, Gresham D. 2021. Fluctuating Environments Maintain Genetic Diversity through Neutral Fitness Effects and Balancing Selection. *Mol Biol Evol* **38**:4362–4375.

Agarwal V, Kommaddi RP, Valli K, Ryder D, Hyde TM, Kleinman JE, Strobel HW, Ravindranath V. 2008. Drug metabolism in human brain: high levels of cytochrome P4503A43 in brain and metabolism of anti-anxiety drug alprazolam to its active metabolite. *PLoS One* **3**:e2337.

Aho AV, Kernighan BW, Weinberger PJ. 1978. Awk, a Pattern Scanning and Processing Language. Bell Laboratories.

Albers PK, McVean G. 2020. Dating genomic variants and shared ancestry in population-scale sequencing data. *PLoS Biol* **18**:e3000586.

Alharbi AF, Sheng N, Nicol K, Strömberg N, Hollox EJ. 2022. Balancing selection at the human salivary agglutinin gene (DMBT1) driven by host-microbe interactions. *iScience* **25**:104189.

Allison AC. 1954a. Protection afforded by sickle-cell trait against subtertian malareal infection. *Br Med J* **1**:290–294.

Allison AC. 1954b. The distribution of the sickle-cell trait in East Africa and elsewhere, and its apparent relationship to the incidence of subtertian malaria. *Trans R Soc Trop Med Hyg* **48**:312–318.

Andrés AM, Dennis MY, Kretzschmar WW, Cannons JL, Lee-Lin S-Q, Hurle B, NISC Comparative Sequencing Program, Schwartzberg PL, Williamson SH, Bustamante CD, Nielsen R, Clark AG, Green ED. 2010. Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. *PLoS Genet* **6**:e1001157.

Barghi N, Hermisson J, Schlötterer C. 2020. Polygenic adaptation: a unifying framework to understand positive selection. *Nat Rev Genet* **21**:769–781.

Benton ML, Abraham A, LaBella AL, Abbot P, Rokas A, Capra JA. 2021. The influence of evolutionary history on human health and disease. *Nat Rev Genet* **22**:269–283.

Berg JJ, Coop G. 2014. A population genetic signal of polygenic adaptation. *PLoS Genet* **10**:e1004412.

Bergström A, Stringer C, Hajdinjak M, Scerri EML, Skoglund P. 2021. Origins of modern human ancestry. *Nature* **590**:229–237.

Bitarello BD, de Filippo C, Teixeira JC, Schmidt JM, Kleinert P, Meyer D, Andrés AM. 2018. Signatures of Long-Term Balancing Selection in Human Genomes. *Genome Biol Evol* **10**:939–955.

Bromberg Y, Kahn PC, Rost B. 2013. Neutral and weakly nonneutral sequence variants may define individuality. *Proc Natl Acad Sci U S A* **110**:14255–14260.

Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. 2006. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* **38**:75–81.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AWC, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Wellcome Trust Case Control Consortium, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME. 2010. Origins and functional impact of copy number variation in the human genome. *Nature* **464**:704–712.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, Others. 2011. 1000 Genomes Project Analysis Group. The variant call format and vcftools. *Bioinformatics* **27**:2156–2158.

DeGiorgio M, Lohmueller KE, Nielsen R. 2014. A model-based approach for identifying

signatures of ancient balancing selection in genetic data. *PLoS Genet* **10**:e1004561.
de Guzman Strong C, Conlan S, Deming CB, Cheng J, Sears KE, Segre JA. 2010. A milieu of regulatory elements in the epidermal differentiation complex syntenic block: implications for atopic dermatitis and psoriasis. *Hum Mol Genet* **19**:1453–1460.
Dudley JT, Kim Y, Liu L, Markov GJ, Gerold K, Chen R, Butte AJ, Kumar S. 2012. Human genomic disease variants: a neutral evolutionary explanation. *Genome Res* **22**:1383–1394.
Edge MD, Coop G. 2019. Reconstructing the History of Polygenic Scores Using Coalescent Trees. *Genetics* **211**:235–262.
Eyre-Walker A. 2010. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc Natl Acad Sci U S A* **107**:1752–1756.
Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol* **128**:415–423.
Fijarczyk A, Babik W. 2015. Detecting balancing selection in genomes: limits and prospects. *Mol Ecol* **24**:3529–3545.
Fisher RA. 1922. XXI.—On the Dominance Ratio. *Proceedings of the Royal Society of Edinburgh* **42**:321–341.
Gibson G. 2018. Population genetics and GWAS: A primer. *PLoS Biol*.
Gokcumen O, Zhu Q, Mulder LCF, Iskow RC, Austermann C, Scharer CD, Raj T, Boss JM, Sunyaev S, Price A, Stranger B, Simon V, Lee C. 2013. Balancing Selection on a Regulatory Region Exhibiting Ancient Variation That Predates Human–Neandertal Divergence. *PLoS Genet* **9**:e1003404.
Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, 1000 Genomes Project, Bustamante CD. 2011. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci U S A* **108**:11983–11988.
Hamid I, Korunes KL, Beleza S, Goldberg A. 2021. Rapid adaptation to malaria facilitated by admixture in the human population of Cabo Verde. *Elife* **10**. doi:10.7554/eLife.63177
Hedrick PW. 2011. Population genetics of malaria resistance in humans. *Heredity* **107**:283–304.
Hedrick PW. 2006. Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics. *Annu Rev Ecol Evol Syst* **37**:67–93.
Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*. doi:10.1093/bioinformatics/18.2.337
Iglewicz B, Hoaglin DC. 1993. How to detect and handle outliers. Asq Press.
Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**:996–1006.
Kharitonenkov A, Chen Z, Sures I, Wang H, Schilling J, Ullrich A. 1997. A family of proteins that inhibit signalling through tyrosine kinase receptors. *Nature* **386**:181–186.
Klunk J, Vilgalys TP, Demeure CE, Cheng X, Shiratori M, Madej J, Beau R, Elli D, Patino MI, Redfern R, DeWitte SN, Gamble JA, Boldsen JL, Carmichael A, Varlik N, Eaton K, Grenier J-C, Golding GB, Devault A, Rouillard J-M, Yotova V, Sindeaux R, Ye CJ, Bikaran M, Dumaine A, Brinkworth JF, Missiakas D, Rouleau GA, Steinrücken M, Pizarro-Cerdá J, Poinar HN, Barreiro LB. 2022. Evolution of immune genes is associated with the Black Death. *Nature*. doi:10.1038/s41586-022-05349-x
Kondrashov AS. 2017. Crumbling Genome: The Impact of Deleterious Mutations on Humans. John Wiley & Sons.
Lachance J, Tishkoff SA. 2013. Population Genomics of Human Adaptation. *Annu Rev Ecol Evol Syst* **44**:123–143.
Langergraber KE, Prüfer K, Rowney C, Boesch C, Crockford C, Fawcett K, Inoue E, Inoue-Muruyama M, Mitani JC, Muller MN, Robbins MM, Schubert G, Stoinski TS, Viola B, Watts D, Wittig RM, Wrangham RW, Zuberbühler K, Pääbo S, Vigilant L. 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human

evolution. *Proc Natl Acad Sci U S A* **109**:15716–15721.
Lederberg J. 1999. JBS Haldane (1949) on infectious disease and evolution. *Genetics* **153**:1–3.
Leffler EM, Gao Z, Pfeifer S, Ségurel L, Auton A, Venn O, Bowden R, Bontrop R, Wall JD, Sella G, Donnelly P, McVean G, Przeworski M. 2013. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science* **339**:1578–1582.
Lettre G. 2014. Rare and low-frequency variants in human common diseases and other complex traits. *J Med Genet* **51**:705–714.
Levene H. 1953. Genetic Equilibrium When More Than One Ecological Niche is Available. *Am Nat* **87**:331–333.
Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* **475**:493–496.
Lin Y-L, Gokcumen O. 2019. Fine-Scale Characterization of Genomic Structural Variation in the Human Genome Reveals Adaptive and Biomedically Relevant Hotspots. *Genome Biol Evol* **11**:1136–1151.
Lin Y-L, Pavlidis P, Karakoc E, Ajay J, Gokcumen O. 2015. The evolution and functional impact of human deletion variants shared with archaic hominin genomes. *Mol Biol Evol* **32**:1008–1019.
Loos RJF. 2020. 15 years of genome-wide association studies and no signs of slowing down. *Nat Commun* **11**:5900.
Mafessoni F, Grote S, de Filippo C, Slon V, Kolobova KA, Viola B, Markin SV, Chintalapati M, Peyrégne S, Skov L, Skoglund P, Krivoshapkin AI, Derevianko AP, Meyer M, Kelso J, Peter B, Prüfer K, Pääbo S. 2020. A high-coverage Neandertal genome from Chagyrskaya Cave. *Proc Natl Acad Sci U S A* **117**:15132–15136.
Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, Skoglund P, Lazaridis I, Sankararaman S, Fu Q, Rohland N, Renaud G, Erlich Y, Willems T, Gallo C, Spence JP, Song YS, Poletti G, Balloux F, Van Driem G, De Knijff P, Romero IG, Jha AR, Behar DM, Bravi CM, Capelli C, Hervig T, Moreno-Estrada A, Posukh OL, Balanovska E, Balanovsky O, Karachanak-Yankova S, Sahakyan H, Toncheva D, Yepiskoposyan L, Tyler-Smith C, Xue Y, Abdullah MS, Ruiz-Linares A, Beall CM, Di Rienzo A, Jeong C, Starikovskaya EB, Metspalu E, Parik J, Villems R, Henn BM, Hodoglugil U, Mahley R, Sajantila A, Stamatoyannopoulos G, Wee JTS, Khusainova R, Khusnutdinova E, Litvinov S, Ayodo G, Comas D, Hammer MF, Kivisild T, Klitz W, Winkler CA, Labuda D, Bamshad M, Jorde LB, Tishkoff SA, Watkins WS, Metspalu M, Dryomov S, Sukernik R, Singh L, Thangaraj K, Paäbo S, Kelso J, Patterson N, Reich D. 2016. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**:201–206.
Mathieson S, Mathieson I. 2018. FADS1 and the Timing of Human Adaptation to Agriculture. *Mol Biol Evol* **35**:2957–2970.
McArthur E, Rinker D, Capra JA. 2020. Quantifying the contribution of Neanderthal introgression to the heritability of complex traits. *bioRxiv*. doi:10.1101/2020.06.08.140087
Mendoza-Revilla J, Chacón-Duque JC, Fuentes-Guajardo M, Ormond L, Wang K, Hurtado M, Villegas V, Granja V, Acuña-Alonzo V, Jaramillo C, Arias W, Lozano RB, Gómez-Valdés J, Villamil-Ramírez H, de Cerqueira CCS, Badillo Rivera KM, Nieves-Colón MA, Gignoux CR, Wojcik GL, Moreno-Estrada A, Hunemeier T, Ramallo V, Schuler-Faccini L, Gonzalez-José R, Bortolini M-C, Canizales-Quinteros S, Gallo C, Poletti G, Bedoya G, Rothhammer F, Balding D, Fumagalli M, Adhikari K, Ruiz-Linares A, Hellenthal G. 2021. Disentangling signatures of selection before and after European colonization in Latin Americans. *bioRxiv*. doi:10.1101/2021.11.15.467418
Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Tandon A, Siebauer M, Green RE, Bryc K, Briggs AW, Stenzel U, Dabney J, Shendure J, Kitzman J, Hammer MF,

Shunkov MV, Derevianko AP, Patterson N, Andrés AM, Eichler EE, Slatkin M, Reich D, Kelso J, Pääbo S. 2012. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* **338**:222–226.
Muller HJ. 1918. Genetic Variability, Twin Hybrids and Constant Hybrids, in a Case of Balanced Lethal Factors. *Genetics* **3**:422–499.
Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Pääbo S, Pritchard JK, Rubin EM. 2006. Sequencing and analysis of Neanderthal genomic DNA. *Science* **314**:1113–1118.
Pajic P, Lin Y-L, Xu D, Gokcumen O. 2016. The psoriasis-associated deletion of late cornified envelope genes LCE3B and LCE3C has been maintained under balancing selection since Human Denisovan divergence. *BMC Evol Biol* **16**:265.
Papadantonakis S, Poirazi P, Pavlidis P. 2016. CoMuS: Simulating coalescent histories and polymorphic data from multiple species. *Mol Ecol Resour*. doi:10.1111/1755-0998.12544
Posth C, Wißing C, Kitagawa K, Pagani L, van Holstein L, Racimo F, Wehrberger K, Conard NJ, Kind CJ, Bocherens H, Krause J. 2017. Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. *Nat Commun* **8**:16046.
Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol* **20**:R208–15.
Prüfer K, de Filippo C, Grote S, Mafessoni F, Korlević P, Hajdinjak M, Vernot B, Skov L, Hsieh P, Peyrégne S, Reher D, Hopfe C, Nagel S, Maricic T, Fu Q, Theunert C, Rogers R, Skoglund P, Chintalapati M, Dannemann M, Nelson BJ, Key FM, Rudan P, Kućan Ž, Gušić I, Golovanova LV, Doronichev VB, Patterson N, Reich D, Eichler EE, Slatkin M, Schierup MH, Andrés AM, Kelso J, Meyer M, Pääbo S. 2017. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* **358**:655–658.
Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, Li H, Mallick S, Dannemann M, Fu Q, Kircher M, Kuhlwilm M, Lachmann M, Meyer M, Ongyerth M, Siebauer M, Theunert C, Tandon A, Moorjani P, Pickrell J, Mullikin JC, Vohr SH, Green RE, Hellmann I, Johnson PLF, Blanche H, Cann H, Kitzman JO, Shendure J, Eichler EE, Lein ES, Bakken TE, Golovanova LV, Doronichev VB, Shunkov MV, Derevianko AP, Viola B, Slatkin M, Reich D, Kelso J, Pääbo S. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**:43–49.
Qiu Q-W, Wu D-D, Yu L-H, Yan T-Z, Zhang W, Li Z-T, Liu Y-H, Zhang Y-P, Xu X-M. 2013. Evidence of recent natural selection on the Southeast Asian deletion (--SEA) causing α-thalassemia in South China. *BMC Evol Biol* **13**:63.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842.
Saitou M, Gokcumen O. 2020. An Evolutionary Perspective on the Impact of Genomic Copy Number Variation on Human Health. *J Mol Evol* **88**:104–119.
Saitou M, Masuda N, Gokcumen O. 2021a. Similarity-based analysis of allele frequency distribution among multiple populations identifies adaptive genomic structural variants. *Mol Biol Evol*. doi:10.1093/molbev/msab313
Saitou M, Resendez S, Pradhan AJ, Wu F, Lie NC, Hall NJ, Zhu Q, Reinholdt L, Satta Y, Speidel L, Nakagome S, Hanchard NA, Churchill G, Lee C, Atilla-Gokcumen GE, Mu X, Gokcumen O. 2021b. Sex-specific phenotypic effects and evolutionary history of an ancient polymorphic deletion of the human growth hormone receptor. *Sci Adv* **7**:eabi4476.
Sella G, Barton NH. 2019. Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annu Rev Genomics Hum Genet* **20**:461–493.
Siewert KM, Voight BF. 2020. BetaScan2: Standardized Statistics to Detect Balancing Selection Utilizing Substitution Data. *Genome Biol Evol* **12**:3873–3877.
Siewert KM, Voight BF. 2017. Detecting Long-Term Balancing Selection Using Allele Frequency

Correlation. *Mol Biol Evol* **34**:2996–3005.
Smith Maynard J, Smith JM, Smith JM. 1998. Evolutionary Genetics. Oxford University Press.
Soni V, Vos M, Eyre-Walker A. 2022. A new test suggests hundreds of amino acid polymorphisms in humans are subject to balancing selection. *PLoS Biol* **20**:e3001645.
Speidel L, Cassidy L, Davies RW, Hellenthal G, Skoglund P, Myers SR. 2021. Inferring Population Histories for Ancient Genomes Using Genome-Wide Genealogies. *Mol Biol Evol* **38**:3497–3511.
Speidel L, Forest M, Shi S, Myers SR. 2019. A method for genome-wide genealogy estimation for thousands of samples. *Nat Genet* **51**:1321–1329.
Starr I, Seiffert-Sinha K, Sinha AA, Gokcumen O. 2021. Evolutionary context of psoriatic immune skin response. *Evolution, Medicine, and Public Health*. doi:10.1093/emph/eoab042
Takahashi Y, Kawata M. 2013. A comprehensive test for negative frequency-dependent selection. *Popul Ecol* **55**:499–509.
Taskent O, Lin YL, Patramanis I, Pavlidis P, Gokcumen O. 2020. Analysis of Haplotypic Variation and Deletion Polymorphisms Point to Multiple Archaic Introgression Events, Including from Altai Neanderthal Lineage. *Genetics* **215**:497–509.
Taskent RO, Alioglu ND, Fer E, Melike Donertas H, Somel M, Gokcumen O. 2017. Variation and Functional Impact of Neanderthal Ancestry in Western Asia. *Genome Biol Evol* **9**:3516–3524.
Van der Auwera GA, O'Connor BD. 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. "O'Reilly Media, Inc."
Vernot B, Akey JM. 2014. Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**:1017–1021.
Wallace B. 1970. Genetic load: its biological and conceptual aspects. *Genetic load: its biological and conceptual aspects*.
Wittmann MJ, Bergland AO, Feldman MW, Schmidt PS, Petrov DA. 2017. Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. *Proc Natl Acad Sci U S A* **114**:E9932–E9941.
Xue Y, Sun D, Daly A, Yang F, Zhou X, Zhao M, Huang N, Zerjal T, Lee C, Carter NP, Hurles ME, Tyler-Smith C. 2008. Adaptive evolution of UGT2B17 copy-number variation. *Am J Hum Genet* **83**:337–346.
Zeng J, de Vlaming R, Wu Y, Robinson MR, Lloyd-Jones LR, Yengo L, Yap CX, Xue A, Sidorenko J, McRae AF, Powell JE, Montgomery GW, Metspalu A, Esko T, Gibson G, Wray NR, Visscher PM, Yang J. 2018. Signatures of negative selection in the genetic architecture of human complex traits. *Nat Genet* **50**:746–753.