Science

RESEARCH ART.

Check for updates

Cite as: F. Yilmaz *et al*., *Science* 10.1126/science.adn0609 (2024).

Reconstruction of the human amylase locus reveals ancient duplications seeding modern-day variation

Feyza Yılmaz[1]†, Charikleia Karageorgiou[2]†, Kwondo Kim[1]†, Petar Pajic[2], Kendra Scheer[2], Human Genome Structural Variation Consortium‡, Christine R. Beck[1,3,4], Ann-Marie Torregrossa[5,6], Charles Lee[1]*, Omer Gokcumen[2]*

[1]The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA. [2]Department of Biological Sciences, University at Buffalo, Buffalo, NY 14260, USA. [3]University of Connecticut, Institute for Systems Genomics, Storrs, CT 06269, USA. [4]The University of Connecticut Health Center, Farmington, CT 06032, USA. [5]Department of Psychology, University at Buffalo, Buffalo, NY 14260, USA. [6]University at Buffalo Center for Ingestive Behavior Research, University at Buffalo, Buffalo, NY 14260, USA.

†These authors contributed equally to this work. ‡Human Genome Structural Variation Consortium collaborators and affiliations are listed at the end of this paper.

*Corresponding author. Email: charles.lee@jax.org (C.L.); omergokc@buffalo.edu (O.G.)

Previous studies suggested that the copy number of the human salivary amylase gene, *AMY1*, correlates with starch-rich diets. However, evolutionary analyses are hampered by the absence of accurate, sequence-resolved haplotype variation maps. We identified 30 structurally distinct haplotypes at nucleotide resolution among 98 present-day humans, revealing that the coding sequences of *AMY1* copies are evolving under negative selection. Genomic analyses of these haplotypes in archaic hominins and ancient human genomes suggest that a common three-copy haplotype, dating as far back as 800 KYA, has seeded rapidly evolving rearrangements through recurrent non-allelic homologous recombination. Additionally, haplotypes with more than three *AMY1* copies have significantly increased in frequency among European farmers over the past 4,000 years, potentially as an adaptive response to increased starch digestion.

Copy number variation at the amylase locus is frequently attributed to human health and adaptation (*1*). As such, this structurally variable locus is a prime target for research on the fundamental biology of gene duplications. There are two types of amylase genes, *AMY1* and *AMY2,* which are reported to be expressed in the salivary glands and pancreas, respectively (*2*). Both genes encode for the amylase enzyme, which breaks down polymeric starch into simple sugar molecules, a crucial digestive process for starch-eating species (*3*). It has been shown that mammals that consume starch-rich diets underwent independent bursts of amylase gene duplications from the ancestral pancreatic *AMY2-like* gene (*4*). A great ape-specific duplication resulted in the formation of the salivary *AMY1* gene (*5*), which has since evolved to produce unusual copy number variations, ranging from 2 to 17 copies per diploid cell (*1*, *6*). This variation is especially prominent in human populations with high starch consumption, particularly those with a history of agriculture (*1*, *6*, *7*). These evolutionary insights indicate that copy number variation at the amylase locus may play an adaptive role in shaping the metabolic response to starchy diets, including the presence of microbes that break down amylase-resistant starch (*8*).

Given its adaptive and putative functional roles, duplications of the *AMY1* gene were linked to the advent of agriculture approximately 10,000 years ago (*1*). The lack of nucleotide-level resolution in evolutionary analysis has led to disagreements about the timing and functional significance of *AMY1* gene duplications in relation to starch-rich diets and human evolution (*7*, *9*–*12*). To address this issue, we have resolved this locus at nucleotide-level at a population scale across 98 individuals from different populations, using optical genome mapping and long-read sequencing techniques. The nucleotide-resolved haplotypes of this locus subsequently allowed us to conduct evolutionary genetic analyses on ancient human and archaic hominin genomes to investigate the timing of *AMY1* gene duplications within the context of agriculture.

Results

Structural haplotypes at the human amylase locus

The amylase locus in the human genome is a ~212.5-kbp region on chromosome 1 (GRCh38; chr1:103,554,220–103,766,732) which contains *AMY2B*, *AMY2A*, *AMY1A*, *AMY1B*, and *AMY1C* genes (Fig. 1A). This locus is largely composed of segmental duplications with > 99% sequence similarity, which complicates its accurate assembly using short-read sequencing (fig. S1). Using the sequence similarity of segmental duplications and the labeling patterns from the optical genome mapping data from the GRCh38 reference assembly in silico map, we defined six distinct amylase segments overlapping the amylase genes, depicted by colored arrows (Fig. 1A and table S1). Using optical genome mapping, which has been previously demonstrated to resolve similar complex regions (*13*–*15*), we constructed haplotype-resolved diploid assemblies for 98 individuals (n = 196 sampled alleles (i.e., haploid sample size); table S2) and characterized this locus using the copy number and orientation of the amylase segments (Fig. 1A). This approach allowed us to identify 52 distinct amylase haplotypes (fig. S2, A and B, and table S3), of which 7 were previously

reported (*7*) (fig. S3). These haplotypes were then classified based on the number of amylase gene copies, adhering to the established nomenclature; HXAYBZ: where HX: represents *AMY1*, AY: represents *AMY2A*, and BZ: represents *AMY2B* copy numbers, with superscripts "a" and "r" indicate ancestral and reference haplotypes, respectively (fig. S4) (*7*, *11*). We subsequently defined 30 high-confidence haplotypes (from 117 observed alleles in 81 individuals) that were orthogonally supported by de novo assemblies based on long-read sequencing (Fig. 1B) (*16*). This represents the first nucleotide-level reconstruction of the amylase locus at a population scale.

The length of the amylase haplotypes ranged from 111-kbp ($H1^a$.1 and $H1^a$.2) to 402-kbp (H7.1) (Fig. 1B), capturing those that are structurally identical to the GRCh38 ($H3^r$.1) and the T2T-chm13 (H7.3) (*17*) reference assemblies (Fig. 1A). Four haplotypes, $H1^a$.1 (n = 20/117), $H3^r$.1 (n = 18/117), $H3^r$.2 (n = 22/117), and $H3^r$.4 (n = 21/117), were categorized as common, each with an allele frequency exceeding 5% across all studied populations. These four common haplotypes collectively constitute approximately 70% of all amylase haplotypes (n = 81/117) in this study. Despite our limited sample size, we found that the four common haplotypes exist in all continental regions (Fig. 1C and fig. S5). In addition, *AMY1* copy number variations do not exhibit a discernible geographic specificity (p-value = 0.4312, Kruskal-Wallis rank sum test).

Out of 30 "high confidence" amylase haplotypes, we identified 19 (63%) as singletons, which are haplotypes that appear only once in our dataset. To infer the relative mutation rate of the amylase locus, we compared this number to that of tandem repeats. To avoid potential biases, we used the same 33 individuals that are present in both the tandem repeat database (EnsembleTR) and our dataset. Only 21 of the 30 distinct amylase haplotypes were present in these 33 individuals. To make sure the analysis was consistent across different genomic regions, we only considered matched tandem repeat loci that had exactly 21 detected alleles in the human population. This yielded 719 tandem repeat loci (fig. S6). We found that the proportion of singleton haplotypes was significantly higher for the amylase locus than the genome-wide average for the 719 tandem repeat loci (the observed empirical percentile = 0.017, Fig. 1D). This observation is informative for understanding the mutation rate at this locus because the allele frequency spectrum and proportion of singletons are determined by mutation rate and genetic drift (*18*). By using the same individuals and matching the number of distinct haplotypes in our comparison, we control for demographic biases and provide a relative estimation of the mutation rate for the amylase locus. Considering that short tandem repeats have mutation rates as low as 10^{-8} (similar to single nucleotide variant mutation rate) and, in some cases, can be as high as 10^{-2} mutations per locus per generation (*19*), our analysis encompasses the entire range of mutation rates found in the genome. We acknowledge that an ideal future comparison would involve other amylase-like loci, which exhibit similar mutational mechanisms and levels of structural variation, and that are resolved using comparable approaches once such databases become available. When we repeated this analysis for the complex 3q29 locus, known for its segmental duplication-rich nature and high levels of structural variation and resolved with similar approaches (*15*, *20*), we found 11 (50%) singleton haplotypes (fig. S7). Thus, the amylase locus is mutating faster than a typical structural variation hotspot and 98.3% of analyzed tandem repeats.

To understand the extent of the amylase haplotypes that are captured in our study compared to what exists in the human population, we conducted rarefaction analysis on the 98 samples. We identified all common haplotypes with a frequency of >= 5% (fig. S8) and 85% of all haplotypes overall (Fig. 1E and fig. S9).

Strong negative selection limits functional variation among amylase gene copies

To systematically assess the selection pressure on the amylase genes coding sequences, we examined the degree of protein-coding sequence variation associated with our high-confidence amylase haplotypes (30 haplotypes from 117 alleles). Gene annotation predicted 582 distinct intact protein-coding amylase gene copies in 117 alleles, and we experimentally validated these predictions using digital droplet PCR on 18 randomly selected individuals (R^2 = 0.94; fig. S10A and tables S4 and S5) (*21*). The reconstruction of the coding sequence phylogeny revealed that all human amylase gene copies can be robustly clustered into three distinct types: *AMY2B*, *AMY2A*, and *AMY1* (bootstrap value = 96%; Fig. 2A and fig. S11). We found that the *AMY2B*, *AMY2A*, and *AMY1* genes had 23, 23, and 36 fixed coding sequence variations unique to each type, resulting in 6, 11, and 19 gene-specific amino acid differences, respectively (table S6). Based on the coding sequence alignment (*22*), we estimated the synonymous and nonsynonymous substitution (dN/dS) ratios using codeml (*23*), where we found no evidence for lineage-specific selection pressure acting on any of the amylase gene types (FDR adjusted p-value from χ^2 test > 0.05; table S7). In contrast, all amylase gene types show significant signatures of negative selection (FDR adjusted p-value from χ^2 test < 0.05; Fig. 2A and table S7). These observations suggest that negative selection (dN/dS ratio < 1) has acted to retain the amino acid sequence of amylase gene copies, both within and between the three amylase gene types. It is of note that we identified two amino acid variants at positions 211 and 366 (Accession: P0DUB6) that could contribute to functional differences and may have biomedical significance (Fig. 2B, fig. S12, and table S8) (*24*).

Our coding sequence data from gene annotations cover 582 intact amylase gene copies and confirmed the explicit difference between *AMY1, AMY2A,* and *AMY2B* genes, which is crucial to distinguish the expression of these genes across tissues. Indeed,

according to existing data portals, Genotype-Tissue Expression and The Human Protein Atlas (*25*, *26*), *AMY2A* and *AMY2B* are expressed in the pancreas and, to some level, in adipose and brain tissues, while the *AMY1* gene is expressed primarily in the parotid salivary gland (fig. S13). These observations are consistent with the idea that *AMY1* first emerged in the ancestor of great apes, resulting in a gain of expression in the parotid gland tissues (*5*). Furthermore, subsequent *AMY1* gene duplications in the human lineage appear to affect the dosage of amylase in the parotid salivary glands (*1*). Among the haplotypes we identified in this study, we found 110 amylase pseudogenes and showed they share a single phylogenetic origin from an ancestral incomplete gene duplication of the *AMY2A* gene (*27*) (fig. S14). Thus, the pseudogenization of *AMY2A* is more likely due to a single mutational event rather than a loss of constraint and the repeated occurrence of novel loss-of-function variants.

Evolution of the copy number of AMY1 gene

We show that all human amylase gene copies can be robustly clustered into three distinct types: *AMY2B*, *AMY2A*, and *AMY1.* However, to specifically study the evolution of the copy number of the salivary *AMY1* gene, we needed to identify the sequences within the amylase locus that were the most phylogenetically informative. To achieve this, we systematically evaluated the variation in 117 alleles by aligning sequences and identified a consensus sequence for each amylase segment (fig. S15, A and B). We identified an interval (from 22,850 to 26,730 bp) (*22*) within the AMY1 segment (fig. S15C), where all observed AMY1 segments (n = 337) can be phylogenetically structured into three distinct clusters; AMY1A (n = 124), AMY1B (n = 99), and AMY1C (n = 114) (fig. S15D). These clusters correspond to the segments represented in the GRCh38 reference assembly. Further, the chimpanzee (panTro6) and gorilla (gorGor6) reference genomes each contain only one AMY1C-like segment. Despite the structural similarity of these haplotypes, we found that the nonhuman primate sequences cluster distinctly from those of the human $H1^a.1$ haplotypes (figs. S16 and S17). These findings suggest that the common ancestor of humans and chimpanzees possessed a single AMY1C-like segment, and the AMY1A and AMY1B segments have evolved only in the human lineage (fig. S17). Note that the bonobo reference genome (panPan3) contains two AMY1 segments: one ancestral and one resulting from an independent, bonobo-specific duplication, as determined through synteny and phylogenetic analysis (figs. S16 and S17). One of these duplications is likely non-functional due to a previously reported disrupted coding sequence (*1*), a finding corroborated by the most recent annotation (RS_2024_02/NHGRI_mPanPan1-v2.0_pri).

To further understand the evolution of the *AMY1* gene copy number, we aligned all AMY1 segments from the most common haplotypes, $H3^r.1$ and $H3^r.2$, that harbor all three AMY1 segment types. Two independent Bayesian phylogenies based on these alignments indicate that the AMY1B segment arose from AMY1C approximately 140 to 270 thousand years ago (KYA), followed by the duplication of AMY1A from the AMY1B segment approximately 120 to 240 KYA (fig. S18 and table S9). Gene conversion between the GC-rich segmental duplications complicates time estimation based on a molecular clock and is a known phenomenon at the amylase locus (*6*, *28*). Considering gene conversion between AMY1 segments, the actual duplication dates are expected to be older than the estimates above. Some studies have used single nucleotide variants in the flanking regions to infer the phylogenetic history of this locus, positing coalescence dates for human amylase locus at ~279 and ~450 KYA (*11*, *12*). However, as described previously (*7*), we found that the linkage disequilibrium between the flanking single nucleotide variants and amylase haplotypes is low (e.g., $H1^a.1$, average: R^2 = ~0.26 and median R^2 = ~0.03) (fig. S19), complicating the time estimation of *AMY1* duplications using flanking regions. Therefore, our estimates avoid these complications and support the conclusion that the initial *AMY1* gene duplications substantially predated out-of-Africa migrations, by at least 30 KYA (table S9).

AMY1 copy number variation in archaic hominin genomes

A complementary approach for estimating the relative timing of gene duplications involves analyzing the read-depth of unique k-mers in ancient human and archaic hominin genomes. We first tested a k-mer approach using the GeneToCN algorithm (*29*) on short-read sequencing data to estimate the copy numbers of the *AMY1*, *AMY2A*, and *AMY2B* genes in 116 present-day human genomes (table S5). Notably, the k-mer approach achieved an R^2 > 0.99 correlation for 32 individuals, each with both haplotypes of the amylase locus reconstructed (fig. S10B). We tested our k-mer approach in 101 samples (a subset of the 116 individuals mentioned above) with digital droplet PCR validation data (R^2 = 0.95; fig. S10A). These results suggest that the GeneToCN estimates are consistent with the digital droplet PCR estimations and are a viable option for estimating *AMY1* gene copy numbers in short-read whole genome sequencing datasets.

We then aimed to estimate *AMY1* gene copy number in eight archaic hominin genomes using two approaches: (1) the validated k-mer method described above, as well as (2) an independent read-depth analysis (table S10). Given the varying genome-wide coverage in most of these genomes, we needed to conduct a downsampling analysis to empirically determine that 1X and 5X genome-wide coverage provides >85% and >95% accuracy in estimating *AMY1* copy number, respectively (fig. S20 and table S11) (*30*). Controlling for GC bias and coverage within the amylase locus for each sample, we were able to reliably estimate the *AMY1* copy number for eight archaic hominin genomes (table S10). We found increased *AMY1* copy numbers in two Eastern and one Western Neanderthals, as well as in one Denisovan genome (Fig. 3A, fig. S21, and table S10). These include the Altai Neanderthal

(2.6 copies), Denisova 2 (8 copies), GoyetQ56-1 (5.0 copies), and Mezmaiskaya 2 (4.7 copies). Previous read-depth analysis of Altai Neanderthal and Denisovan genomes found no evidence for an increase of *AMY1* gene copy number (*10*, *31*). By incorporating additional archaic hominin genomes that have not been previously analyzed for *AMY1* copy number (total: n = 8), we have now detected signatures of *AMY1* gene duplication in a total of four archaic hominins.

The *AMY1* duplication in these archaic hominins could be postulated by four scenarios. First, it is possible that because of the complexity of the amylase locus and the complications inherent in archaic hominin genomic sequencing, there may be a technical bias in our detection. However, the observation of duplications using two different approaches and in multiple genomes provides confidence in our results. Second, it is plausible that introgression into archaic hominins from humans may explain the presence of duplications in the former, especially in the light of recently revealed complex interactions and gene flow events between Neanderthals and present-day humans over time (*32*). The approach recently developed by Li and colleagues (*32*) will be valuable for more formally testing for introgression from and to Neanderthals, potentially underlying the origins of the observed *AMY1* duplications in Neanderthals. However, the current conservative filtering approach ends up filtering ~89% of the bases in the amylase locus. Thus, the specific signals of introgression, even if they exist, remain hidden. Third, it is plausible that the duplication evolved independently in the archaic hominin lineage. However, we argue that two independent duplications (one in humans and another in archaic hominins) in less than a million years are unlikely, given that initial duplications from single *AMY1* copy haplotypes are rare in nonhuman primates (*4*) (fig. S16). The fourth and, in our opinion, the most plausible scenario is that the *AMY1* gene might already have been copy number variable before the human-Neanderthal/Denisovan divergence (~800 KYA (*33*)), albeit to a limited extent as compared to what is observed in present-day humans. Overall, our results suggest a complex history of *AMY1* duplications, which will be further scrutinized as more high-coverage archaic hominin genomes become available.

The AMY1 copy number has increased in the last 4,000 years among European farmers

To explore the changes in frequency of *AMY1* copy number since the migration out of Africa ~60 KYA (*34*), we analyzed the genomes of 68 ancient human genomes (table S12). The oldest genome analyzed was the Ust'-Ishim sample from Siberia (~45,000 BP), which has six *AMY1* gene copies per diploid cell. Similarly, the oldest modern human from Europe, the Peştera Muierii sample from Romania (~34,000 BP) has eight *AMY1* gene copies per diploid cell. These copy numbers indicate that high *AMY1* gene copies (defined here as >= 6 copies per diploid cell) had already spread across Eurasia as far back as ~45,000 BP (*35*) (Fig. 3B).

We next analyzed the ancient human genomes in the context of agricultural development and found a general trend where the *AMY1* gene copy number is significantly higher among samples excavated from archaeologically agricultural contexts compared to those from hunter-gatherer contexts (p-value = 0.023; fig. S22). To further investigate this trend, we examined ancient human genomes from Europe, where we have a clear timeline of the Neolithic transition. Specifically, the Neolithic transition of Europe started with the cultural and genetic influx from Anatolian farmers around ~9,000 BP and progressed into Northwestern Europe by 5,000 BP (*36*), with small, isolated groups of hunter-gatherers persisting until at least 4,000 BP (*37*). Our analysis encompasses the geographic and temporal span of this transition. The oldest sample in our dataset from an agricultural European archaeological context is AKT16 from Anatolia, dated to 8,547 BP (*38*), while the youngest sample from a hunter-gatherer archaeological context is SRA62 from Ireland, dated to 5,215 BP (*39*). Based on the dates and archaeological context of the samples described in their respective studies (table S12), we parsed the ancient European human genomes into the following periods for visualization purposes: i) the pre-agricultural period (> 9,000 BP), where all our samples are from hunter-gatherers; ii) the agricultural transition period (9,000 to 4,000 BP), representing the long period of transition to agriculture in Europe where both hunter-gatherer and agriculturalist groups co-existed; and iii) the post-agricultural period, during which all of Europe has completely transitioned into an agriculturalist life style (< 4,000 BP) (*38*–*59*). We found that pre-agricultural genomes already harbored four to eight *AMY1* copies per diploid cell (table S12). We also observed a consistent and significant increase in *AMY1* copy numbers across these periods (p-value = 0.005; Fig. 3, C and D), and found similar trends for the *AMY2A* genes (Fig. 3C) and non-European samples (figs. S23 and S24). These findings support the notion that amylase haplotypes with higher numbers of amylase gene copies have increased in frequency over the last 4,000 years. We found no significant differences in *AMY1* copy number between samples excavated from agricultural versus hunter-gatherer archaeological contexts during the "agricultural transition" period when farmers and hunter-gatherers shared the same habitat.

Taken together, these findings are consistent with the idea that either neutral evolution or weak-adaptive forces, perhaps due to preagricultural experimentation with food processing techniques (*60*), such as flour production from wild cereals (*61*), have retained the wide range of standing *AMY1* copy number variation in preagricultural Europe. The gradual increase in starch availability as Europe transitioned to an agricultural lifestyle (*62*) may underlie selective forces acting on high copy number haplotypes, explaining the increase of *AMY1* copy number in late post-agriculture European populations. Given that the exact mechanisms through which *AMY1* copy number may confer adaptive advantage are unknown and the dietary intake even within

agricultural and hunter-gatherer groups is highly diverse (*63*), it is challenging to reach definitive conclusions that the high *AMY1* copy number had an adaptive role during the agricultural transition. Additional samples with comprehensive archaeological context will help further elucidate the potential adaptive role of amylase in relation to the specific composition of ancient diets, as well as demographic considerations such as population replacements and gene flow from different groups that shaped the European Neolithic gene pool.

Multiple mutational mechanisms underlie the amylase copy number variation

We next investigated the likely mutational origins of the present-day amylase haplotypes. To do this, we first examined the relationship among the four most common haplotypes, H3^{r}.1, H3^{r}.2, H3^{r}.4, and H1^{a}.1. Building upon the previously reported link between one-copy and three-copy haplotypes (*64*) and the fact that we observe only three types of AMY1 segments, we propose an evolutionary model linking the ancestral chimpanzee-like haplotype (H1^{a}-like) to the common three-copy haplotypes (H3^{r}.1 and H3^{r}.2) (fig. S25). According to this model, the initial duplications of *AMY1* starting from H1^{a}-like ancestral haplotype led to the emergence of H3^{r} haplotypes. Given that multiple mutational steps are required to move from one-copy common haplotypes to other common haplotypes and that we do not observe any intermediate haplotypes in the present-day human genomes, the duplication from one-copy to three-copy haplotypes is likely to have occurred only once in the human lineage. This is also supported by the absence of duplications in the amylase segments within the H1^{a}.1 haplotype, which would impede non-allelic homologous recombination (NAHR) events (*65*). In contrast, H3^{r}.1 haplotypes harbor copies of identical and unidirectional sequences (e.g., AMY2A.2 segments), providing an ideal template for recurrent NAHR events leading to the diverse haplotypes we see today.

To further investigate the mutational relationships between the H3^{r} haplotypes and other extant haplotypes, we utilized dotplots and sequence alignments to delineate the breakpoints of structural differences in amylase haplotypes (*66*, *67*). In parallel, we conducted a scan for PRDM9-binding motifs within the amylase locus to pinpoint possible recombination sites (Fig. 4A, fig. S26, and table S13). By integrating all these observations, we were able to construct a putative evolutionary path with the fewest plausible mutational steps that can explain the origin of present-day amylase haplotypes starting from NAHR-prone H3^{r} haplotypes (figs. S27 and S28; (*66*)). Our proposed evolutionary model of mutational events is consistent with our hypothesis regarding the central role of the H3^{r} haplotypes in seeding extant haplotype variation and offers three major insights into the evolution of the amylase locus in humans.

First, we found evidence for recurrent NAHR events among common haplotypes (H3^{r}.1, H3^{r}.2, and H3^{r}.4) harboring the AMY1A and AMY1B segments with breakpoints in the AMY2A.2 segment. These NAHR events, which may occur among different haplotype combinations, could concurrently result in the duplication and deletion of two *AMY1* gene copies (e.g., Fig. 4B and fig. S28) (*66*). Therefore, while other less likely scenarios are possible, NAHR-based deletion and duplications underlie copy number variation of the AMY1 segments and thus *AMY1* genes. Specifically, as this proposed mechanism always adds or deletes two copies of the *AMY1* genes, our finding explains how most human diploid genomes harbor even-numbered *AMY1* gene copies (*7*) (fig. S29). Thus, the majority of H1^{a}.1 haplotypes among present-day humans may have predominantly originated from H3^{r} haplotypes. If true, this hypothesis explains the homoplastic occurrence of H1^{a} haplotypes across the amylase phylogenetic tree (fig. S19C) and the lack of an out-of-Africa signal in H1^{a} nucleotide diversity, which would be expected if the H1^{a} haplotypes arose before out-of-Africa migrations (table S14). We argue that although the H1^{a}.1 haplotype is structurally nearly identical to the ancestral (chimpanzee) human amylase haplotype, the H1^{a}.1 haplotypes observed in extant humans have arisen recurrently from H3^{r} haplotypes.

Second, we characterized three microhomology-mediated break-induced replication events and identified the accompanying microhomologies at the breakpoint junctions (Fig. 4C) (*66*). Even though these three rearrangements constitute only five alleles (H2A2B2.1, H3^{r}.6, and H3B2.1) (~ 4%), they hold substantial biological relevance since H2A2B2.1 and H3B2.1 harbor duplications of the *AMY2* genes. Taken together, different mechanisms drive the copy number variation of the salivary *AMY1* genes and pancreatic *AMY2* genes, with the lower copy number variation of *AMY2* genes explained by the slower rate of non-recurrent microhomology-mediated break-induced replication events (*68*).

Third, recurrent NAHR-mediated inversion events at the amylase locus, similar to those described previously (*69*), underpin the mutational connections between the common H3^{r}.1, H3^{r}.2, and H3^{r}.4 haplotypes (fig. S30), as well as several other inversions among extant haplotypes (*66*). Given that inversions underlie the structural differences between the majority of the amylase haplotypes, their functional and adaptive relevance presents an intriguing avenue for future research.

Discussion

In this study, we have dissected the evolution of the amylase locus. First, we hypothesize that the initial duplications of the *AMY1* genes occurred once through multiple duplications, evolving from a one-copy ancestral haplotype to the three-copy present-day haplotypes frequently observed in our dataset. Analysis of archaic hominin genomes suggests that these initial duplications may have occurred well before the split of the human-Neanderthal/Denisovan. This observation is concordant with the recent

evidence of Neanderthal starch consumption (*70*), and perhaps the availability of cooked starch in archaic hominins made possible through the domestication of fire (*71*) (Fig. 5).

Second, we hypothesize that selection acted on abundant standing *AMY1* copy number variation at this locus rather than on de novo variants. We observed a wide range of *AMY1* gene copy number variation (3 - 9 copies) in samples that predate agriculture. We further found that late agricultural populations consistently harbored amylase haplotypes with higher *AMY1* copy numbers (Fig. 5). However, the lack of linkage disequilibrium between the flanking SNPs and the *AMY1* copy number variation prevented us from conducting haplotype-based tests of positive selection acting on *AMY1* copy number. We hypothesize that partial soft sweeps involving pre-existing amylase haplotypes may have influenced *AMY1* copy number variation in relation to historical starch consumption patterns in different populations. The effects of amylase gene duplications on taste preferences and starch metabolism may have also predisposed humans to prefer and tolerate the consumption of wild grains, as reported for the Mesolithic groups in the Balkans (*72*), facilitating the adoption of starch-rich diets and the eventual domestication of these plants. Overall, our results support the complex narrative of the transition to agriculture, which includes the replacement of Western hunter-gatherers by Anatolian farmers in Europe (*73*), potentially bringing with them amylase haplotypes that harbor higher *AMY1* gene copies. Similarly, transient interactions between hunter-gatherer and agricultural groups (*38*) could explain the similar copy numbers observed between these groups during the transition period. Given the unknowns concerning how *AMY1* copy number affects metabolic function and its adaptive value under different life histories and starch consumption levels, it is challenging to draw definitive conclusions. Further, the agricultural transition varied across different regions and periods, involving diverse types of starches, which likely influenced the putative role of *AMY1* copy number in local adaptation. Anthropologically contextualized studies in indigenous populations—such as the Andeans (*74*, *75*), could provide further insights into the relationships between dietary practices, metabolic outcomes, and amylase genetic variation.

Third, we found that NAHR and microhomology-mediated break-induced replication underlie the copy number variation of *AMY1* and *AMY2* genes, respectively, explaining the different rates of their evolution. The extremely high rate of structural variation due to NAHR led to remarkable *AMY1* copy number variation with distinct mutational propensities. For example, our evolutionary model involving common $H3^r$ haplotypes suggests recurrent NAHR events mediated by highly similar sequences (> 99%), resulting in duplications or deletions of both *AMY1A* and *AMY1B* genes. In contrast, the $H1^a$ haplotypes harboring a single *AMY1* copy and divergent *AMY2* genes are less susceptible to NAHR events. Therefore, the mutation types and rates at the amylase locus may differ depending on extant haplotype variation in a population, especially in bottlenecked populations such as Indigenous Americans (*76*). It is a distinct possibility that a bottlenecked population ends up with a very high frequency of $H1^a$ due to drift. In this case, the absence of segments with highly similar sequences within $H1^a$ haplotype would mitigate recurrent NAHR events, resulting in a slower accumulation of variation in this population. In contrast, if one of the larger amylase haplotypes were to become prevalent due to drift, the rate of variation would increase exponentially. Within this general context, one interesting question for future work is whether larger amylase haplotypes experience negative selection due to increased genomic instability.

Taken together, our study underscores how gene duplications in early human history provided the genetic foundation for dietary flexibility during agricultural innovations, contributing to modern human evolution.

Materials and methods summary

In this study, samples (n = 98) from the 1000 Genomes Project (*77*), as part of the datasets from the Human Genome Structural Variation Consortium (HGSVC) (*20*), and Human Pangenome Reference Consortium (HPRC) (*78*), and the Genome In a Bottle (*79*) were analyzed. For each sample, datasets from Bionano Genomics optical genome mapping and PacBio HiFi sequencing were utilized. A ~212.5-kbp region on chromosome 1 containing amylase genes was characterized using optical genome mapping and blastn (BLASTN 2.9.0+) alignments (*80*). The amylase segments were identified and validated across human and nonhuman primate samples. De novo assembly of optical genome maps was performed using Bionano Solve v3.5, followed by alignment and haplotype reconstruction using Bionano Access v1.7 software (*81*). PacBio HiFi long-read sequencing data were also de novo assembled using hifiasm 0.16.1-r375 (*82*). Population-specific patterns in the *AMY1* gene copy numbers were evaluated using the Kruskal-Wallis test. The singleton proportions at the amylase locus were compared to those of genome-wide tandem repeats. Rarefaction analysis was conducted to assess haplotype diversity at the amylase locus.

To predict amylase gene coordinates and copy number, we used Exonerate v2.4.0 “protein2genome” tool (*83*) with default settings and a maximum intron length of 20-kbp, employing amylase protein sequences from UNIPROT (*84*). Overlapping hits were clustered, and the best hit was selected. Hits translating into full-length polypeptides (511 aa) were kept. Any conflicts with predicted amylase gene copy numbers were manually curated. De novo genome annotations were obtained using the T2T-chm13 gff annotation by liftoff v1.6.3 (*85*) and compared with the homology-based annotations. Digital droplet polymerase chain reaction (ddPCR) with custom primers was used to validate *AMY1* copy numbers in human DNA samples, with HindIII digestion prior

to ddPCR. Alignments of amylase segments and coding sequences were constructed using MAFFT v7.310 (*86*) and PAGAN v1.53 (*87*), incorporating chimpanzee sequences as an outgroup. Phylogenetic trees were reconstructed using IQ-TREE v2.2.0 (*88*) and visualized with FigTree v1.4.4. dN/dS ratios were estimated using PAML v4.10.6 codeml (*23*), testing for gene type-specific selection pressures. Functional analysis of amylase protein sequences was performed with CLUSTALO v1.2.3 (*89*) and AlphaMissense (*90*). The expression data of amylase genes were analyzed using TPM data from the Genotype-Tissue Expression (GTEx) portal (*26*), focusing on three tissues with the highest expression values. Since major salivary glands were not included in GTEx, supplementary TPM data from Saitou *et al*. (*91*) were used, considering AMY1A, AMY1B, and AMY1C transcripts collectively as representative of AMY1 expression. Sequence origins and breakpoints in AMY1 segments were identified using BLAT v37x1 (*92*), nucmer v3.1, and mummerplot v3.5 (*93*, *94*). *AMY1* duplication events were dated using BEASTv 2.7.5 (*95*), without tree topology optimization for H3^{r}.1 and H3^{r}.2 haplotypes independently, and results were visualized with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

We used GeneToCN (*29*) to determine the amylase copy number in 68 ancient human and 8 archaic hominin genomes. Raw sequencing data were sourced from the European Nucleotide Archive. GeneToCN, utilizing 25-mers, calculated the copy numbers of *AMY1*, *AMY2A*, and *AMY2B* genes. This analysis was validated with ddPCR for present-day humans. *AMY1*, *AMY2A*, and *AMY2B* copy numbers were compared between hunter-gatherer and agricultural lifestyles using the Wilcoxon Rank Sum Test. The copy number trends over time were also analyzed. To assess coverage bias, eight ancient genomes were downsampled, and their *AMY1* copy numbers were recalculated and compared to full genome estimates. This analysis considered potential batch effects and sample diversity.

We obtained raw sequencing reads of archaic hominin genomes (n = 38) from public datasets, removed adaptors, and merged overlapping paired-end reads using leeHom v1.2.17 (*96*). For the Chagyrskaya Neanderthal genome, preprocessed reads were used. Clean reads were mapped to the GRCh38 reference genome using bwa v0.7.17 aln (*97*). GC bias was assessed and corrected with deepTools v3.5.1 (*98*), resulting in the exclusion of 29 genomes. Average read-depth at both genome-wide level and the amylase locus was calculated with mosdepth v0.3.5 (*99*), excluding one more sample for low coverage (< 1X). GC-corrected reads were realigned to T2T-chm13 and GRCh38 assemblies with bwa v0.7.17 mem (*97*) and further aligned to a reference containing a single *AMY1* gene copy. Amylase gene copy numbers were calculated from the average read-depth of each amylase gene normalized by the genome-wide average.

To identify potential PRDM9 binding sites in the H3^{r}.1 haplotype, we used FIMO v5.5.4 from the MEME package (*100*) with the degenerate 13-bp motif "CCNCCNTNNCCNC" and a non-redundant DNA database background frequency matrix. Hits with a FIMO score above 10 and a p-value below 0.00011 were considered. To investigate structural variation mechanisms, we aligned haplotypes using nucmer v3.1 (*93*, *94*) and visualized alignments with mummerplot and SVbyEye (https://github.com/daewoooo/SVbyEye/tree/master). Potential breakpoints were identified and 20-kbp sequences upstream and downstream were aligned using MAFFT v7.522 (*86*). We evaluated replication and DNA recombination-based processes and inferred NAHR when breakpoints were within paralogous segments and lacked replication-based sequence motifs.

REFERENCES AND NOTES

1. G. H. Perry, N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, N. P. Carter, C. Lee, A. C. Stone, Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**, 1256–1260 (2007). doi:10.1038/ng2123 Medline
2. T. Nishide, M. Emi, Y. Nakamura, K. Matsubara, G. Kosaki, S. Himeno, K. Matsubara, Sequences of cDNAs for human salivary and pancreatic α-amylases. *Gene* **28**, 263–270 (1984). doi:10.1016/0378-1119(84)90265-8 Medline
3. C. Peyrot des Gachons, P. A. S. Breslin, Salivary Amylase: Digestion and Metabolic Syndrome. *Curr. Diab. Rep.* **16**, 102 (2016). doi:10.1007/s11892-016-0794-7 Medline
4. P. Pajic, P. Pavlidis, K. Dean, L. Neznanova, R.-A. Romano, D. Garneau, E. Daugherity, A. Globig, S. Ruhl, O. Gokcumen, Independent amylase gene copy number bursts correlate with dietary preferences in mammals. *eLife* **8**, e44628 (2019). doi:10.7554/eLife.44628 Medline
5. M. H. Meisler, C. N. Ting, The remarkable evolutionary history of the human amylase genes. *Crit. Rev. Oral Biol. Med.* **4**, 503–509 (1993). doi:10.1177/10454411930040033501 Medline
6. P. C. Groot, W. H. Mager, R. R. Frants, Interpretation of polymorphic DNA patterns in the human alpha-amylase multigene family. *Genomics* **10**, 779–785 (1991). doi:10.1016/0888-7543(91)90463-O Medline
7. C. L. Usher, R. E. Handsaker, T. Esko, M. A. Tuke, M. N. Weedon, A. R. Hastie, H. Cao, J. E. Moon, S. Kashin, C. Fuchsberger, A. Metspalu, C. N. Pato, M. T. Pato, M. I. McCarthy, M. Boehnke, D. M. Altshuler, T. M. Frayling, J. N. Hirschhorn, S. A. McCarroll, Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. *Nat. Genet.* **47**, 921–925 (2015). doi:10.1038/ng.3340 Medline
8. A. C. Poole, J. K. Goodrich, N. D. Youngblut, G. G. Luque, A. Ruaud, J. L. Sutter, J. L. Waters, Q. Shi, M. El-Hadidi, L. M. Johnson, H. Y. Bar, D. H. Huson, J. G. Booth, R. E. Ley, Human Salivary Amylase Gene Copy Number Impacts Oral and Gut Microbiomes. *Cell Host Microbe* **25**, 553–564.e7 (2019). doi:10.1016/j.chom.2019.03.001 Medline
9. S. Mathieson, I. Mathieson, FADS1 and the Timing of Human Adaptation to Agriculture. *Mol. Biol. Evol.* **35**, 2957–2970 (2018). doi:10.1093/molbev/msy180 Medline
10. G. H. Perry, L. Kistler, M. A. Kelaita, A. J. Sams, Insights into hominin phenotypic and dietary evolution from ancient DNA sequence data. *J. Hum. Evol.* **79**, 55–63 (2015). doi:10.1016/j.jhevol.2014.10.018 Medline
11. D. Bolognini, A. Halgren, R. N. Lou, A. Raveane, J. L. Rocha, A. Guarracino, N. Soranzo, J. Chin, E. Garrison, P. H. Sudmant, Global diversity, recurrent evolution, and recent selection on amylase structural haplotypes in humans. bioRxiv 2024.02.07.579378 [Preprint] (2024); doi:10.1101/2024.02.07.579378.
12. C. E. Inchley, C. D. A. Larbey, N. A. A. Shwan, L. Pagani, L. Saag, T. Antão, G. Jacobs, G. Hudjashov, E. Metspalu, M. Mitt, C. A. Eichstaedt, B. Malyarchuk, M. Derenko, J. Wee, S. Abdullah, F.-X. Ricaut, M. Mormina, R. Mägi, R. Villems, M. Metspalu, M. K. Jones, J. A. L. Armour, T. Kivisild, Selective sweep on human

amylase genes postdates the split with Neanderthals. *Sci. Rep.* **6**, 37198 (2016). doi:10.1038/srep37198 Medline

13. W. Demaerel, Y. Mostovoy, F. Yilmaz, L. Vervoort, S. Pastor, M. S. Hestand, A. Swillen, E. Vergaelen, E. A. Geiger, C. R. Coughlin, S. K. Chow, D. McDonald-McGinn, B. Morrow, P.-Y. Kwok, M. Xiao, B. S. Emanuel, T. H. Shaikh, J. R. Vermeesch, The 22q11 low copy repeats are characterized by unprecedented size and structural variability. *Genome Res.* **29**, 1389–1401 (2019). doi:10.1101/gr.248682.119 Medline
14. Y. Mostovoy, F. Yilmaz, S. K. Chow, C. Chu, C. Lin, E. A. Geiger, N. J. L. Meeks, K. C. Chatfield, C. R. Coughlin II, U. Surti, P.-Y. Kwok, T. H. Shaikh, Genomic regions associated with microdeletion/microduplication syndromes exhibit extreme diversity of structural variation. *Genetics* **217**, iyaa038 (2021). doi:10.1093/genetics/iyaa038 Medline
15. F. Yilmaz, U. Gurusamy, T. J. Mosley, P. Hallast, K. Kim, Y. Mostovoy, R. H. Purcell, T. H. Shaikh, M. E. Zwick, P.-Y. Kwok, C. Lee, J. G. Mulle, High level of complexity and global diversity of the 3q29 locus revealed by optical mapping and long-read sequencing. *Genome Med.* **15**, 35 (2023). doi:10.1186/s13073-023-01184-5 Medline
16. F. Creators Yilmaz, Amylase Segments, Haplotypes, Liftoff Gene Annotations and OrthogonalValidation-MoleculeSupport; https://zenodo.org/doi/10.5281/zenodo.13247145.
17. S. Nurk, S. Koren, A. Rhie, M. Rautiainen, A. V. Bzikadze, A. Mikheenko, M. R. Vollger, N. Altemose, L. Uralsky, A. Gershman, S. Aganezov, S. J. Hoyt, M. Diekhans, G. A. Logsdon, M. Alonge, S. E. Antonarakis, M. Borchers, G. G. Bouffard, S. Y. Brooks, G. V. Caldas, N.-C. Chen, H. Cheng, C.-S. Chin, W. Chow, L. G. de Lima, P. C. Dishuck, R. Durbin, T. Dvorkina, I. T. Fiddes, G. Formenti, R. S. Fulton, A. Fungtammasan, E. Garrison, P. G. S. Grady, T. A. Graves-Lindsay, I. M. Hall, N. F. Hansen, G. A. Hartley, M. Haukness, K. Howe, M. W. Hunkapiller, C. Jain, M. Jain, E. D. Jarvis, P. Kerpedjiev, M. Kirsche, M. Kolmogorov, J. Korlach, M. Kremitzki, H. Li, V. V. Maduro, T. Marschall, A. M. McCartney, J. McDaniel, D. E. Miller, J. C. Mullikin, E. W. Myers, N. D. Olson, B. Paten, P. Peluso, P. A. Pevzner, D. Porubsky, T. Potapova, E. I. Rogaev, J. A. Rosenfeld, S. L. Salzberg, V. A. Schneider, F. J. Sedlazeck, K. Shafin, C. J. Shew, A. Shumate, Y. Sims, A. F. A. Smit, D. C. Soto, I. Sović, J. M. Storer, A. Streets, B. A. Sullivan, F. Thibaud-Nissen, J. Torrance, J. Wagner, B. P. Walenz, A. Wenger, J. M. D. Wood, C. Xiao, S. M. Yan, A. C. Young, S. Zarate, U. Surti, R. C. McCoy, M. Y. Dennis, I. A. Alexandrov, J. L. Gerton, R. J. O'Neill, W. Timp, J. M. Zook, M. C. Schatz, E. E. Eichler, K. H. Miga, A. M. Phillippy, The complete sequence of a human genome. *Science* **376**, 44–53 (2022). doi:10.1126/science.abj6987 Medline
18. A. Harpak, A. Bhaskar, J. K. Pritchard, Mutation Rate Variation is a Primary Determinant of the Distribution of Allele Frequencies in Humans. *PLOS Genet.* **12**, e1006489 (2016). doi:10.1371/journal.pgen.1006489 Medline
19. M. Verbiest, M. Maksimov, Y. Jin, M. Anisimova, M. Gymrek, T. Bilgin Sonay, Mutation and selection processes regulating short tandem repeats give rise to genetic and phenotypic diversity across species. *J. Evol. Biol.* **36**, 321–336 (2023). doi:10.1111/jeb.14106 Medline
20. P. Ebert, P. A. Audano, Q. Zhu, B. Rodriguez-Martin, D. Porubsky, M. J. Bonder, A. Sulovari, J. Ebler, W. Zhou, R. Serra Mari, F. Yilmaz, X. Zhao, P. Hsieh, J. Lee, S. Kumar, J. Lin, T. Rausch, Y. Chen, J. Ren, M. Santamarina, W. Höps, H. Ashraf, N. T. Chuang, X. Yang, K. M. Munson, A. P. Lewis, S. Fairley, L. J. Tallon, W. E. Clarke, A. O. Basile, M. Byrska-Bishop, A. Corvelo, U. S. Evani, T.-Y. Lu, M. J. P. Chaisson, J. Chen, C. Li, H. Brand, A. M. Wenger, M. Ghareghani, W. T. Harvey, B. Raeder, P. Hasenfeld, A. A. Regier, H. J. Abel, I. M. Hall, P. Flicek, O. Stegle, M. B. Gerstein, J. M. C. Tubio, Z. Mu, Y. I. Li, X. Shi, A. R. Hastie, K. Ye, Z. Chong, A. D. Sanders, M. C. Zody, M. E. Talkowski, R. E. Mills, S. E. Devine, C. Lee, J. O. Korbel, T. Marschall, E. E. Eichler, Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* **372**, eabf7117 (2021). doi:10.1126/science.abf7117 Medline
21. P. Pajic, Digital Droplet PCR Protocol for Copy Number Variation Estimation, Zenodo (2024). https://doi.org/10.5281/ZENODO.13155899.
22. K. Kim, AmylaseSegmentCDSAlignments,DatingAMY1,DNdSAnalysis,OriginBreakPointAMY1,Protein2genome_GeneAnnotation,TandemRepeatAnalysis, Zenodo (2024); https://doi.org/10.5281/zenodo.13169856.
23. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007). doi:10.1093/molbev/msm088 Medline
24. P. Pajic, Functional analysis data (AlphaMissense) for Amylase proteins, Zenodo (2024); https://doi.org/10.5281/ZENODO.13156521.
25. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015). doi:10.1126/science.1260419 Medline
26. F. Aguet, S. Anand, K. G. Ardlie, S. Gabriel, G. A. Getz, A. Graubert, K. Hadley, R. E. Handsaker, K. H. Huang, S. Kashin, X. Li, D. G. MacArthur, S. R. Meier, J. L. Nedzel, D. T. Nguyen, A. V. Segrè, E. Todres, B. Balliu, A. N. Barbeira, A. Battle, R. Bonazzola, A. Brown, C. D. Brown, S. E. Castel, D. F. Conrad, D. J. Cotter, N. Cox, S. Das, O. M. de Goede, E. T. Dermitzakis, J. Einson, B. E. Engelhardt, E. Eskin, T. Y. Eulalio, N. M. Ferraro, E. D. Flynn, L. Fresard, E. R. Gamazon, D. Garrido-Martín, N. R. Gay, M. J. Gloudemans, R. Guigó, A. R. Hame, Y. He, P. J. Hoffman, F. Hormozdiari, L. Hou, H. K. Im, B. Jo, S. Kasela, M. Kellis, S. Kim-Hellmuth, A. Kwong, T. Lappalainen, X. Li, Y. Liang, S. Mangul, P. Mohammadi, S. B. Montgomery, M. Muñoz-Aguirre, D. C. Nachun, A. B. Nobel, M. Oliva, Y. S. Park, Y. Park, P. Parsana, A. S. Rao, F. Reverter, J. M. Rouhana, C. Sabatti, A. Saha, M. Stephens, B. E. Stranger, B. J. Strober, N. A. Teran, A. Viñuela, G. Wang, X. Wen, F. Wright, V. Wucher, Y. Zou, P. G. Ferreira, G. Li, M. Melé, E. Yeger-Lotem, M. E. Barcus, D. Bradbury, T. Krubit, J. A. McLean, L. Qi, K. Robinson, N. V. Roche, A. M. Smith, L. Sobin, D. E. Tabor, A. Undale, J. Bridge, L. E. Brigham, B. A. Foster, B. M. Gillard, R. Hasz, M. Hunter, C. Johns, M. Johnson, E. Karasik, G. Kopen, W. F. Leinweber, A. McDonald, M. T. Moser, K. Myer, K. D. Ramsey, B. Roe, S. Shad, J. A. Thomas, G. Walters, M. Washington, J. Wheeler, S. D. Jewell, D. C. Rohrer, D. R. Valley, D. A. Davis, D. C. Mash, P. A. Branton, L. K. Barker, H. M. Gardiner, M. Mosavel, L. A. Siminoff, P. Flicek, M. Haeussler, T. Juettemann, W. J. Kent, C. M. Lee, C. C. Powell, K. R. Rosenbloom, M. Ruffier, D. Sheppard, K. Taylor, S. J. Trevanion, D. R. Zerbino, N. S. Abell, J. Akey, L. Chen, K. Demanelis, J. A. Doherty, A. P. Feinberg, K. D. Hansen, P. F. Hickey, F. Jasmine, L. Jiang, R. Kaul, M. G. Kibriya, J. B. Li, Q. Li, S. Lin, S. E. Linder, B. L. Pierce, L. F. Rizzardi, A. D. Skol, K. S. Smith, M. Snyder, J. Stamatoyannopoulos, H. Tang, M. Wang, L. J. Carithers, P. Guan, S. E. Koester, A. R. Little, H. M. Moore, C. R. Nierras, A. K. Rao, J. B. Vaught, S. Volpi; GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020). doi:10.1126/science.aaz1776 Medline
27. D. Carpenter, S. Dhar, L. M. Mitchell, B. Fu, J. Tyson, N. A. A. Shwan, F. Yang, M. G. Thomas, J. A. L. Armour, Obesity, starch digestion and amylase: Association between copy number variants at human salivary (AMY1) and pancreatic (AMY2) amylase genes. *Hum. Mol. Genet.* **24**, 3472–3480 (2015). doi:10.1093/hmg/ddv098 Medline
28. M. R. Vollger, P. C. Dishuck, W. T. Harvey, W. S. DeWitt, X. Guitart, M. E. Goldberg, A. N. Rozanski, J. Lucas, M. Asri, K. M. Munson, A. P. Lewis, K. Hoekzema, G. A. Logsdon, D. Porubsky, B. Paten, K. Harris, P. Hsieh, E. E. Eichler, M. J. P. Chaisson, P.-C. Chang, X. H. Chang, H. Cheng, J. Chu, S. Cody, V. Colonna, D. E. Cook, R. M. Cook-Deegan, O. E. Cornejo, M. Diekhans, D. Doerr, P. Ebert, J. Ebler, J. M. Eizenga, S. Fairley, O. Fedrigo, A. L. Felsenfeld, X. Feng, C. Fischer, P. Flicek, G. Formenti, A. Frankish, R. S. Fulton, Y. Gao, S. Garg, E. Garrison, N. A. Garrison, C. G. Giron, R. E. Green, C. Groza, A. Guarracino, L. Haggerty, I. M. Hall, M. Haukness, D. Haussler, S. Heumos, G. Hickey, T. Hourlier, K. Howe, M. Jain, E. D. Jarvis, H. P. Ji, E. E. Kenny, B. A. Koenig, A. Kolesnikov, J. O. Korbel, J. Kordosky, S. Koren, H. J. Lee, H. Li, W.-W. Liao, S. Lu, T.-Y. Lu, J. K. Lucas, H. Magalhães, S. Marco-Sola, P. Marijon, C. Markello, T. Marschall, F. J. Martin, A. McCartney, J. McDaniel, K. H. Miga, M. W. Mitchell, J. Monlong, J. Mountcastle, M. N. Mwaniki, M. Nattestad, A. M. Novak, S. Nurk, H. E. Olsen, N. D. Olson, B. Paten, T. Pesout, A. M. Phillippy, A. B. Popejoy, P. Prins, D. Puiu, M. Rautiainen, A. A. Regier, A. Rhie, S. Sacco, A. D. Sanders, V.

A. Schneider, B. I. Schultz, K. Shafin, J. A. Sibbesen, J. Sirén, M. W. Smith, H. J. Sofia, A. N. Abou Tayoun, F. Thibaud-Nissen, C. Tomlinson, F. F. Tricomi, F. Villani, M. R. Vollger, J. Wagner, B. Walenz, T. Wang, J. M. D. Wood, A. V. Zimin, J. M. Zook, K. M. Munson, A. P. Lewis, K. Hoekzema, G. A. Logsdon, D. Porubsky, B. Paten, K. Harris, P. H. Hsieh, E. E. Eichler; Human Pangenome Reference Consortium, Increased mutation and gene conversion within human segmental duplications. *Nature* **617**, 325–334 (2023). doi:10.1038/s41586-023-05895-y Medline

29. F.-D. Pajuste, M. Remm, GeneToCN: An alignment-free method for gene copy number estimation directly from next-generation sequencing reads. *Sci. Rep.* **13**, 17765 (2023). doi:10.1038/s41598-023-44636-z Medline
30. K. Scheer, Zenodo Downsampling Infomation, Zenodo (2024); https://doi.org/10.5281/ZENODO.13251208.
31. K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014). doi:10.1038/nature12886 Medline
32. L. Li, T. J. Comi, R. F. Bierman, J. M. Akey, Recurrent gene flow between Neanderthals and modern humans over the past 200,000 years. *Science* **385**, eadi1768 (2024). doi:10.1126/science.adi1768 Medline
33. A. Gómez-Robles, Dental evolutionary rates and its implications for the Neanderthal-modern human divergence. *Sci. Adv.* **5**, eaaw1268 (2019). doi:10.1126/sciadv.aaw1268 Medline
34. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G. Poletti, F. Balloux, G. van Driem, P. de Knijff, I. G. Romero, A. R. Jha, D. M. Behar, C. M. Bravi, C. Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M. S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C. Jeong, E. B. Starikovskaya, E. Metspalu, J. Parik, R. Villems, B. M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J. T. S. Wee, R. Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D. Comas, M. F. Hammer, T. Kivisild, W. Klitz, C. A. Winkler, D. Labuda, M. Bamshad, L. B. Jorde, S. A. Tishkoff, W. S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K. Thangaraj, S. Pääbo, J. Kelso, N. Patterson, D. Reich, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi:10.1038/nature18964 Medline
35. L. Vallini, C. Zampieri, M. J. Shoaee, E. Bortolini, G. Marciani, S. Aneli, T. Pievani, S. Benazzi, A. Barausse, M. Mezzavilla, M. D. Petraglia, L. Pagani, The Persian plateau served as hub for Homo sapiens after the main out of Africa dispersal. *Nat. Commun.* **15**, 1882 (2024). doi:10.1038/s41467-024-46161-7 Medline
36. M. Furholt, Mobility and Social Change: Understanding the European Neolithic Period after the Archaeogenetic Revolution. *J. Archaeol. Res.* **29**, 481–535 (2021). doi:10.1007/s10814-020-09153-x
37. L. Saag, L. Varul, C. L. Scheib, J. Stenderup, M. E. Allentoft, L. Saag, L. Pagani, M. Reidla, K. Tambets, E. Metspalu, A. Kriiska, E. Willerslev, T. Kivisild, M. Metspalu, Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe. *Curr. Biol.* **27**, 2185–2193.e6 (2017). doi:10.1016/j.cub.2017.06.022 Medline
38. N. Marchi, L. Winkelbach, I. Schulz, M. Brami, Z. Hofmanová, J. Blöcher, C. S. Reyna-Blanco, Y. Diekmann, A. Thiéry, A. Kapopoulou, V. Link, V. Piuz, S. Kreutzer, S. M. Figarska, E. Ganiatsou, A. Pukaj, T. J. Struck, R. N. Gutenkunst, N. Karul, F. Gerritsen, J. Pechtl, J. Peters, A. Zeeb-Lanz, E. Lenneis, M. Teschler-Nicola, S. Triantaphyllou, S. Stefanović, C. Papageorgopoulou, D. Wegmann, J. Burger, L. Excoffier, The genomic origins of the world's first farmers. *Cell* **185**, 1842–1859.e18 (2022). doi:10.1016/j.cell.2022.04.008 Medline
39. L. M. Cassidy, R. Ó. Maoldúin, T. Kador, A. Lynch, C. Jones, P. C. Woodman, E. Murphy, G. Ramsey, M. Dowd, A. Noonan, C. Campbell, E. R. Jones, V. Mattiangeli, D. G. Bradley, A dynastic elite in monumental Neolithic society. *Nature* **582**, 384–388 (2020). doi:10.1038/s41586-020-2378-6 Medline
40. J.-P. Bocquet-Appel, S. Naji, M. Vander Linden, J. Kozlowski, Understanding the rates of expansion of the farming system in Europe. *J. Archaeol. Sci.* **39**, 531–546 (2012). doi:10.1016/j.jas.2011.10.010
41. K. Wang, K. Prüfer, B. Krause-Kyora, A. Childebayeva, V. J. Schuenemann, V. Coia, F. Maixner, A. Zink, S. Schiffels, J. Krause, High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry. *Cell Genomics* **3**, 100377 (2023). doi:10.1016/j.xgen.2023.100377 Medline
42. A. Seguin-Orlando, R. Donat, C. Der Sarkissian, J. Southon, C. Thèves, C. Manen, Y. Tchérémissinoff, E. Crubézy, B. Shapiro, J.-F. Deleuze, L. Dalén, J. Guilaine, L. Orlando, Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France. *Curr. Biol.* **31**, 1072–1083.e10 (2021). doi:10.1016/j.cub.2020.12.015 Medline
43. E. Svensson, T. Günther, A. Hoischen, M. Hervella, A. R. Munters, M. Ioana, F. Ridiche, H. Edlund, R. C. van Deuren, A. Soficaru, C. de-la-Rua, M. G. Netea, M. Jakobsson, Genome of Peştera Muierii skull shows high diversity and low mutational load in pre-glacial Europe. *Curr. Biol.* **31**, 2973–2983.e9 (2021). doi:10.1016/j.cub.2021.04.045 Medline
44. B. Ariano, V. Mattiangeli, E. M. Breslin, E. W. Parkinson, T. R. McLaughlin, J. E. Thompson, R. K. Power, J. T. Stock, B. Mercieca-Spiteri, S. Stoddart, C. Malone, S. Gopalakrishnan, L. M. Cassidy, D. G. Bradley, Ancient Maltese genomes and the genetic geography of Neolithic Europe. *Curr. Biol.* **32**, 2668–2680.e6 (2022). doi:10.1016/j.cub.2022.04.069 Medline
45. M. Rasmussen, Y. Li, S. Lindgreen, J. S. Pedersen, A. Albrechtsen, I. Moltke, M. Metspalu, E. Metspalu, T. Kivisild, R. Gupta, M. Bertalan, K. Nielsen, M. T. P. Gilbert, Y. Wang, M. Raghavan, P. F. Campos, H. M. Kamp, A. S. Wilson, A. Gledhill, S. Tridico, M. Bunce, E. D. Lorenzen, J. Binladen, X. Guo, J. Zhao, X. Zhang, H. Zhang, Z. Li, M. Chen, L. Orlando, K. Kristiansen, M. Bak, N. Tommerup, C. Bendixen, T. L. Pierre, B. Grønnow, M. Meldgaard, C. Andreasen, S. A. Fedorova, L. P. Osipova, T. F. G. Higham, C. B. Ramsey, T. V. O. Hansen, F. C. Nielsen, M. H. Crawford, S. Brunak, T. Sicheritz-Pontén, R. Villems, R. Nielsen, A. Krogh, J. Wang, E. Willerslev, Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010). doi:10.1038/nature08835 Medline
46. I. Lazaridis, N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu, K. I. Bos, S. Nordenfelt, H. Li, C. de Filippo, K. Prüfer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fornander, N. Rohland, D. Delsate, M. Francken, J.-M. Guinet, J. Wahl, G. Ayodo, H. A. Babiker, G. Bailliet, E. Balanovska, O. Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. M. Bravi, F. Brisighelli, G. B. J. Busby, F. Cali, M. Churnosov, D. E. C. Cole, D. Corach, L. Damba, G. van Driem, S. Dryomov, J.-M. Dugoujon, S. A. Fedorova, I. Gallego Romero, M. Gubina, M. Hammer, B. M. Henn, T. Hervig, U. Hodoglugil, A. R. Jha, S. Karachanak-Yankova, R. Khusainova, E. Khusnutdinova, R. Kittles, T. Kivisild, W. Klitz, V. Kučinskas, A. Kushniarevich, L. Laredj, S. Litvinov, T. Loukidis, R. W. Mahley, B. Melegh, E. Metspalu, J. Molina, J. Mountain, K. Näkkäläjärvi, D. Nesheva, T. Nyambo, L. Osipova, J. Parik, F. Platonov, O. Posukh, V. Romano, F. Rothhammer, I. Rudan, R. Ruizbakiev, H. Sahakyan, A. Sajantila, A. Salas, E. B. Starikovskaya, A. Tarekegn, D. Toncheva, S. Turdikulova, I. Uktveryte, O. Utevska, R. Vasquez, M. Villena, M. Voevoda, C. A. Winkler, L. Yepiskoposyan, P. Zalloua, T. Zemunik, A. Cooper, C. Capelli, M. G. Thomas, A. Ruiz-Linares, S. A. Tishkoff, L. Singh, K. Thangaraj, R. Villems, D. Comas, R. Sukernik, M. Metspalu, M. Meyer, E. E. Eichler, J. Burger, M. Slatkin, S. Pääbo, J. Kelso, D. Reich, J. Krause, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014). doi:10.1038/nature13673 Medline
47. S. Schiffels, W. Haak, P. Paajanen, B. Llamas, E. Popescu, L. Loe, R. Clarke, A. Lyons, R. Mortimer, D. Sayer, C. Tyler-Smith, A. Cooper, R. Durbin, Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nat. Commun.* **7**, 10408 (2016). doi:10.1038/ncomms10408 Medline
48. C. Gamba, E. R. Jones, M. D. Teasdale, R. L. McLaughlin, G. Gonzalez-Fortes, V.

Mattiangeli, L. Domboróczki, I. Kővári, I. Pap, A. Anders, A. Whittle, J. Dani, P. Raczky, T. F. G. Higham, M. Hofreiter, D. G. Bradley, R. Pinhasi, Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014). doi:10.1038/ncomms6257 Medline

49. E. R. Jones, G. Gonzalez-Fortes, S. Connell, V. Siska, A. Eriksson, R. Martiniano, R. L. McLaughlin, M. Gallego Llorente, L. M. Cassidy, C. Gamba, T. Meshveliani, O. Bar-Yosef, W. Müller, A. Belfer-Cohen, Z. Matskevich, N. Jakeli, T. F. G. Higham, M. Currat, D. Lordkipanidze, M. Hofreiter, A. Manica, R. Pinhasi, D. G. Bradley, Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015). doi:10.1038/ncomms9912 Medline
50. M. E. Allentoft, M. Sikora, A. Refoyo-Martínez, E. K. Irving-Pease, A. Fischer, W. Barrie, A. Ingason, J. Stenderup, K.-G. Sjögren, A. Pearson, B. Sousa da Mota, B. Schulz Paulsson, A. Halgren, R. Macleod, M. L. S. Jørkov, F. Demeter, L. Sørensen, P. O. Nielsen, R. A. Henriksen, T. Vimala, H. McColl, A. Margaryan, M. Ilardo, A. Vaughn, M. Fischer Mortensen, A. B. Nielsen, M. Ulfeldt Hede, N. N. Johannsen, P. Rasmussen, L. Vinner, G. Renaud, A. Stern, T. Z. T. Jensen, G. Scorrano, H. Schroeder, P. Lysdahl, A. D. Ramsøe, A. Skorobogatov, A. J. Schork, A. Rosengren, A. Ruter, A. Outram, A. A. Timoshenko, A. Buzhilova, A. Coppa, A. Zubova, A. M. Silva, A. J. Hansen, A. Gromov, A. Logvin, A. B. Gotfredsen, B. Henning Nielsen, B. González-Rabanal, C. Lalueza-Fox, C. J. McKenzie, C. Gaunitz, C. Blasco, C. Liesau, C. Martinez-Labarga, D. V. Pozdnyakov, D. Cuenca-Solana, D. O. Lordkipanidze, D. En'shin, D. C. Salazar-García, T. D. Price, D. Borić, E. Kostyleva, E. V. Veselovskaya, E. R. Usmanova, E. Cappellini, E. Brinch Petersen, E. Kannegaard, F. Radina, F. Eylem Yediay, H. Duday, I. Gutiérrez-Zugasti, I. Merts, I. Potekhina, I. Shevnina, I. Altinkaya, J. Guilaine, J. Hansen, J. E. Aura Tortosa, J. Zilhão, J. Vega, K. Buck Pedersen, K. Tunia, L. Zhao, L. N. Mylnikova, L. Larsson, L. Metz, L. Yepiskoposyan, L. Pedersen, L. Sarti, L. Orlando, L. Slimak, L. Klassen, M. Blank, M. González-Morales, M. Silvestrini, M. Vretemark, M. S. Nesterova, M. Rykun, M. F. Rolfo, M. Szmyt, M. Przybyła, M. Calattini, M. Sablin, M. Dobisíková, M. Meldgaard, M. Johansen, N. Berezina, N. Card, N. A. Saveliev, O. Poshekhonova, O. Rickards, O. V. Lozovskaya, O. Gábor, O. C. Uldum, P. Aurino, P. Kosintsev, P. Courtaud, P. Ríos, P. Mortensen, P. Lotz, P. Persson, P. Bangsgaard, P. de Barros Damgaard, P. Vang Petersen, P. P. Martinez, P. Włodarczak, R. V. Smolyaninov, R. Maring, R. Menduiña, R. Badalyan, R. Iversen, R. Turin, S. Vasilyev, S. Wåhlin, S. Borutskaya, S. Skochina, S. A. Sørensen, S. H. Andersen, T. Jørgensen, Y. B. Serikov, V. I. Molodin, V. Smrcka, V. Merts, V. Appadurai, V. Moiseyev, Y. Magnusson, K. H. Kjær, N. Lynnerup, D. J. Lawson, P. H. Sudmant, S. Rasmussen, T. S. Korneliussen, R. Durbin, R. Nielsen, O. Delaneau, T. Werge, F. Racimo, K. Kristiansen, E. Willerslev, Population genomics of post-glacial western Eurasia. *Nature* **625**, 301–311 (2024). doi:10.1038/s41586-023-06865-0 Medline
51. P. Maisano Delser, E. R. Jones, A. Hovhannisyan, L. Cassidy, R. Pinhasi, A. Manica, A curated dataset of modern and ancient high-coverage shotgun human genomes. *Sci. Data* **8**, 202 (2021). doi:10.1038/s41597-021-00980-1 Medline
52. M. Srigyan, H. Bolívar, I. Ureña, J. Santana, A. Petersen, E. Iriarte, E. Kırdök, N. Bergfeldt, A. Mora, M. Jakobsson, K. Abdo, F. Braemer, C. Smith, J. J. Ibañez, A. Götherström, T. Günther, C. Valdiosera, Bioarchaeological evidence of one of the earliest Islamic burials in the Levant. *Commun. Biol.* **5**, 554 (2022). doi:10.1038/s42003-022-03508-4 Medline
53. L. M. Cassidy, R. Martiniano, E. M. Murphy, M. D. Teasdale, J. Mallory, B. Hartwell, D. G. Bradley, Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 368–373 (2016). doi:10.1073/pnas.1518445113 Medline
54. F. Sánchez-Quinto, H. Malmström, M. Fraser, L. Girdland-Flink, E. M. Svensson, L. G. Simões, R. George, N. Hollfelder, G. Burenhult, G. Noble, K. Britton, S. Talamo, N. Curtis, H. Brzobohata, R. Sumberova, A. Götherström, J. Storå, M. Jakobsson, Megalithic tombs in western and northern Neolithic Europe were linked to a kindred society. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9469–9474 (2019). doi:10.1073/pnas.1818037116 Medline
55. L. G. Simões, R. Peyroteo-Stjerna, G. Marchand, C. Bernhardsson, A. Vialet, D. Chetty, E. Alaçamlı, H. Edlund, D. Bouquin, C. Dina, N. Garmond, T. Günther, M. Jakobsson, Genomic ancestry and social dynamics of the last hunter-gatherers of Atlantic France. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2310545121 (2024). doi:10.1073/pnas.2310545121 Medline
56. C. Valdiosera, T. Günther, J. C. Vera-Rodríguez, I. Ureña, E. Iriarte, R. Rodríguez-Varela, L. G. Simões, R. M. Martínez-Sánchez, E. M. Svensson, H. Malmström, L. Rodríguez, J.-M. Bermúdez de Castro, E. Carbonell, A. Alday, J. A. Hernández Vera, A. Götherström, J.-M. Carretero, J. L. Arsuaga, C. I. Smith, M. Jakobsson, Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 3428–3433 (2018). doi:10.1073/pnas.1717762115 Medline
57. S. S. Ebenesersdóttir, M. Sandoval-Velasco, E. D. Gunnarsdóttir, A. Jagadeesan, V. B. Guðmundsdóttir, E. L. Thordardóttir, M. S. Einarsdóttir, K. H. S. Moore, Á. Sigurðsson, D. N. Magnúsdóttir, H. Jónsson, S. Snorradóttir, E. Hovig, P. Møller, I. Kockum, T. Olsson, L. Alfredsson, T. F. Hansen, T. Werge, G. L. Cavalleri, E. Gilbert, C. Lalueza-Fox, J. W. Walser 3rd, S. Kristjánsdóttir, S. Gopalakrishnan, L. Árnadóttir, Ó. Þ. Magnússon, M. T. P. Gilbert, K. Stefánsson, A. Helgason, Ancient genomes from Iceland reveal the making of a human population. *Science* **360**, 1028–1032 (2018). doi:10.1126/science.aar2625 Medline
58. M. Lipson, A. Szécsényi-Nagy, S. Mallick, A. Pósa, B. Stégmár, V. Keerl, N. Rohland, K. Stewardson, M. Ferry, M. Michel, J. Oppenheimer, N. Broomandkhoshbacht, E. Harney, S. Nordenfelt, B. Llamas, B. Gusztáv Mende, K. Köhler, K. Oross, M. Bondár, T. Marton, A. Osztás, J. Jakucs, T. Paluch, F. Horváth, P. Csengeri, J. Koós, K. Sebők, A. Anders, P. Raczky, J. Regenye, J. P. Barna, S. Fábián, G. Serlegi, Z. Toldi, E. Gyöngyvér Nagy, J. Dani, E. Molnár, G. Pálfi, L. Márk, B. Melegh, Z. Bánfai, L. Domboróczki, J. Fernández-Eraso, J. Antonio Mujika-Alustiza, C. Alonso Fernández, J. Jiménez Echevarría, R. Bollongino, J. Orschiedt, K. Schierhold, H. Meller, A. Cooper, J. Burger, E. Bánffy, K. W. Alt, C. Lalueza-Fox, W. Haak, D. Reich, Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017). doi:10.1038/nature24476 Medline
59. N. E. Altınışık, D. D. Kazancı, A. Aydoğan, H. C. Gemici, Ö. D. Erdal, S. Sarıaltun, K. B. Vural, D. Koptekin, K. Gürün, E. Sağlıcan, D. Fernandes, G. Çakan, M. M. Koruyucu, V. K. Lagerholm, C. Karamurat, M. Özkan, G. M. Kılınç, A. Sevkar, E. Sürer, A. Götherström, Ç. Atakuman, Y. S. Erdal, F. Özer, A. Erim Özdoğan, M. Somel, A genomic snapshot of demographic and cultural dynamism in Upper Mesopotamia during the Neolithic Transition. *Sci. Adv.* **8**, eabo3609 (2022). doi:10.1126/sciadv.abo3609 Medline
60. A. Revedin, B. Aranguren, R. Becattini, L. Longo, E. Marconi, M. M. Lippi, N. Skakun, A. Sinitsyn, E. Spiridonova, J. Svoboda, Thirty thousand-year-old evidence of plant food processing. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18815–18819 (2010). doi:10.1073/pnas.1006993107 Medline
61. D. R. Piperno, E. Weiss, I. Holst, D. Nadel, Processing of wild cereal grains in the Upper Palaeolithic revealed by starch grain analysis. *Nature* **430**, 670–673 (2004). doi:10.1038/nature02734 Medline
62. J. Jovanović, R. C. Power, C. de Becdelièvre, G. Goude, S. Stefanović, Microbotanical evidence for the spread of cereal use during the Mesolithic-Neolithic transition in the Southeastern Europe (Danube Gorges): Data from dental calculus analysis. *J. Archaeol. Sci.* **125**, 105288 (2021). doi:10.1016/j.jas.2020.105288
63. L. Betti, R. M. Beyer, E. R. Jones, A. Eriksson, F. Tassi, V. Siska, M. Leonardi, P. Maisano Delser, L. K. Bentley, P. R. Nigst, J. T. Stock, R. Pinhasi, A. Manica, Climate shaped how Neolithic farmers and European hunter-gatherers interacted after a major slowdown from 6,100 BCE to 4,500 BCE. *Nat. Hum. Behav.* **4**, 1004–1010 (2020). doi:10.1038/s41562-020-0897-7 Medline
64. P. C. Groot, W. H. Mager, N. V. Henriquez, J. C. Pronk, F. Arwert, R. J. Planta, A. W. Eriksson, R. R. Frants, Evolution of the human alpha-amylase multigene family through unequal, homologous, and inter- and intrachromosomal crossovers. *Genomics* **8**, 97–105 (1990). doi:10.1016/0888-7543(90)90230-R Medline
65. C. Karageorgiou, Amylase locus genomic analysis, Zenodo (2024); https://doi.org/10.5281/ZENODO.13170898.
66. See the supplementary text.
67. C. Karageorgiou, Amylase locus genomic analysis, Zenodo (2024); https://doi.org/10.5281/ZENODO.13138560.
68. F. Zhang, M. Khajavi, A. M. Connolly, C. F. Towne, S. D. Batish, J. R. Lupski, The

DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat. Genet.* **41**, 849–853 (2009). doi:10.1038/ng.399 Medline

69. D. Porubsky, W. Höps, H. Ashraf, P. Hsieh, B. Rodriguez-Martin, F. Yilmaz, J. Ebler, P. Hallast, F. A. Maria Maggiolini, W. T. Harvey, B. Henning, P. A. Audano, D. S. Gordon, P. Ebert, P. Hasenfeld, E. Benito, Q. Zhu, C. Lee, F. Antonacci, M. Steinrücken, C. R. Beck, A. D. Sanders, T. Marschall, E. E. Eichler, J. O. Korbel; Human Genome Structural Variation Consortium (HGSVC), Recurrent inversion polymorphisms in humans associate with genetic instability and genomic disorders. *Cell* **185**, 1986–2005.e26 (2022). doi:10.1016/j.cell.2022.04.017 Medline
70. J. A. Fellows Yates, I. M. Velsko, F. Aron, C. Posth, C. A. Hofman, R. M. Austin, C. E. Parker, A. E. Mann, K. Nägele, K. W. Arthur, J. W. Arthur, C. C. Bauer, I. Crevecoeur, C. Cupillard, M. C. Curtis, L. Dalén, M. Díaz-Zorita Bonilla, J. C. Díez Fernández-Lomana, D. G. Drucker, E. Escribano Escrivá, M. Francken, V. E. Gibbon, M. R. González Morales, A. Grande Mateu, K. Harvati, A. G. Henry, L. Humphrey, M. Menéndez, D. Mihailović, M. Peresani, S. Rodríguez Moroder, M. Roksandic, H. Rougier, S. Sázelová, J. T. Stock, L. G. Straus, J. Svoboda, B. Teßmann, M. J. Walker, R. C. Power, C. M. Lewis, K. Sankaranarayanan, K. Guschanski, R. W. Wrangham, F. E. Dewhirst, D. C. Salazar-García, J. Krause, A. Herbig, C. Warinner, The evolution and changing ecology of the African hominid oral microbiome. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2021655118 (2021). doi:10.1073/pnas.2021655118 Medline
71. C. Larbey, S. M. Mentzer, B. Ligouis, S. Wurz, M. K. Jones, Cooked starchy food in hearths ca. 120 kya and 65 kya (MIS 5e and MIS 4) from Klasies River Cave, South Africa. *J. Hum. Evol.* **131**, 210–227 (2019). doi:10.1016/j.jhevol.2019.03.015 Medline
72. E. Cristiani, A. Radini, A. Zupancich, A. Gismondi, A. D'Agostino, C. Ottoni, M. Carra, S. Vukojičić, M. Constantinescu, D. Antonović, T. D. Price, D. Borić, Wild cereal grain consumption among Early Holocene foragers of the Balkans predates the arrival of agriculture. *eLife* **10**, e72976 (2021). doi:10.7554/eLife.72976 Medline
73. G. M. Kılınç, A. Omrak, F. Özer, T. Günther, A. M. Büyükkarakaya, E. Bıçakçı, D. Baird, H. M. Dönertaş, A. Ghalichi, R. Yaka, D. Koptekin, S. C. Açan, P. Parvizi, M. Krzewińska, E. A. Daskalaki, E. Yüncü, N. D. Dağtaş, A. Fairbairn, J. Pearson, G. Mustafaoğlu, Y. S. Erdal, Y. G. Çakan, İ. Togan, M. Somel, J. Storå, M. Jakobsson, A. Götherström, The Demographic Development of the First Farmers in Anatolia. *Curr. Biol.* **26**, 2659–2666 (2016). doi:10.1016/j.cub.2016.07.057 Medline
74. J. Lindo, R. Haas, C. Hofman, M. Apata, M. Moraga, R. A. Verdugo, J. T. Watson, C. Viviano Llave, D. Witonsky, C. Beall, C. Warinner, J. Novembre, M. Aldenderfer, A. Di Rienzo, The genetic prehistory of the Andean highlands 7000 years BP though European contact. *Sci. Adv.* **4**, eaau4921 (2018). doi:10.1126/sciadv.aau4921 Medline
75. K. Jorgensen, O. A. Garcia, M. Kiyamu, T. D. Brutsaert, A. W. Bigham, Genetic adaptations to potato starch digestion in the Peruvian Andes. *Am. J. Biol. Anthropol.* **180**, 162–172 (2023). doi:10.1002/ajpa.24656
76. J. V. Moreno-Mayar, B. A. Potter, L. Vinner, M. Steinrücken, S. Rasmussen, J. Terhorst, J. A. Kamm, A. Albrechtsen, A.-S. Malaspinas, M. Sikora, J. D. Reuther, J. D. Irish, R. S. Malhi, L. Orlando, Y. S. Song, R. Nielsen, D. J. Meltzer, E. Willerslev, Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* **553**, 203–207 (2018). doi:10.1038/nature25173 Medline
77. G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles, G. A. McVean; 1000 Genomes Project Consortium, A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010). doi:10.1038/nature09534 Medline
78. W.-W. Liao, M. Asri, J. Ebler, D. Doerr, M. Haukness, G. Hickey, S. Lu, J. K. Lucas, J. Monlong, H. J. Abel, S. Buonaiuto, X. H. Chang, H. Cheng, J. Chu, V. Colonna, J. M. Eizenga, X. Feng, C. Fischer, R. S. Fulton, S. Garg, C. Groza, A. Guarracino, W. T. Harvey, S. Heumos, K. Howe, M. Jain, T.-Y. Lu, C. Markello, F. J. Martin, M. W. Mitchell, K. M. Munson, M. N. Mwaniki, A. M. Novak, H. E. Olsen, T. Pesout, D. Porubsky, P. Prins, J. A. Sibbesen, J. Sirén, C. Tomlinson, F. Villani, M. R. Vollger, L. L. Antonacci-Fulton, G. Baid, C. A. Baker, A. Belyaeva, K. Billis, A. Carroll, P.-C. Chang, S. Cody, D. E. Cook, R. M. Cook-Deegan, O. E. Cornejo, M. Diekhans, P. Ebert, S. Fairley, O. Fedrigo, A. L. Felsenfeld, G. Formenti, A. Frankish, Y. Gao, N. A. Garrison, C. G. Giron, R. E. Green, L. Haggerty, K. Hoekzema, T. Hourlier, H. P. Ji, E. E. Kenny, B. A. Koenig, A. Kolesnikov, J. O. Korbel, J. Kordosky, S. Koren, H. Lee, A. P. Lewis, H. Magalhães, S. Marco-Sola, P. Marijon, A. McCartney, J. McDaniel, J. Mountcastle, M. Nattestad, S. Nurk, N. D. Olson, A. B. Popejoy, D. Puiu, M. Rautiainen, A. A. Regier, A. Rhie, S. Sacco, A. D. Sanders, V. A. Schneider, B. I. Schultz, K. Shafin, M. W. Smith, H. J. Sofia, A. N. Abou Tayoun, F. Thibaud-Nissen, F. F. Tricomi, J. Wagner, B. Walenz, J. M. D. Wood, A. V. Zimin, G. Bourque, M. J. P. Chaisson, P. Flicek, A. M. Phillippy, J. M. Zook, E. E. Eichler, D. Haussler, T. Wang, E. D. Jarvis, K. H. Miga, E. Garrison, T. Marschall, I. M. Hall, H. Li, B. Paten, A draft human pangenome reference. *Nature* **617**, 312–324 (2023). doi:10.1038/s41586-023-05896-x Medline
79. J. M. Zook, D. Catoe, J. McDaniel, L. Vang, N. Spies, A. Sidow, Z. Weng, Y. Liu, C. E. Mason, N. Alexander, E. Henaff, A. B. R. McIntyre, D. Chandramohan, F. Chen, E. Jaeger, A. Moshrefi, K. Pham, W. Stedman, T. Liang, M. Saghbini, Z. Dzakula, A. Hastie, H. Cao, G. Deikus, E. Schadt, R. Sebra, A. Bashir, R. M. Truty, C. C. Chang, N. Gulbahce, K. Zhao, S. Ghosh, F. Hyland, Y. Fu, M. Chaisson, C. Xiao, J. Trow, S. T. Sherry, A. W. Zaranek, M. Ball, J. Bobe, P. Estep, G. M. Church, P. Marks, S. Kyriazopoulou-Panagiotopoulou, G. X. Y. Zheng, M. Schnall-Levin, H. S. Ordonez, P. A. Mudivarti, K. Giorda, Y. Sheng, K. B. Rypdal, M. Salit, Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci. Data* **3**, 160025 (2016). doi:10.1038/sdata.2016.25 Medline
80. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990). doi:10.1016/S0022-2836(05)80360-2 Medline
81. H. Cao, A. R. Hastie, D. Cao, E. T. Lam, Y. Sun, H. Huang, X. Liu, L. Lin, W. Andrews, S. Chan, S. Huang, X. Tong, M. Requa, T. Anantharaman, A. Krogh, H. Yang, H. Cao, X. Xu, Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. *Gigascience* **3**, 34 (2014). doi:10.1186/2047-217X-3-34 Medline
82. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021). doi:10.1038/s41592-020-01056-5 Medline
83. G. S. C. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005). doi:10.1186/1471-2105-6-31 Medline
84. UniProt Consortium, UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019). Medline
85. A. Shumate, S. L. Salzberg, Liftoff: Accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2020). doi:10.1093/bioinformatics/btaa1016 Medline
86. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013). doi:10.1093/molbev/mst010 Medline
87. A. Löytynoja, A. J. Vilella, N. Goldman, Accurate extension of multiple sequence alignments using a phylogeny-aware graph algorithm. *Bioinformatics* **28**, 1684–1691 (2012). doi:10.1093/bioinformatics/bts198 Medline
88. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020). doi:10.1093/molbev/msaa015 Medline
89. F. Sievers, D. G. Higgins, Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145 (2018). doi:10.1002/pro.3290 Medline
90. J. Cheng, G. Novati, J. Pan, C. Bycroft, A. Žemgulytė, T. Applebaum, A. Pritzel, L. H. Wong, M. Zielinski, T. Sargeant, R. G. Schneider, A. W. Senior, J. Jumper, D. Hassabis, P. Kohli, Ž. Avsec, Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**, eadg7492 (2023). doi:10.1126/science.adg7492 Medline
91. M. Saitou, E. A. Gaylord, E. Xu, A. J. May, L. Neznanova, S. Nathan, A. Grawe, J. Chang, W. Ryan, S. Ruhl, S. M. Knox, O. Gokcumen, Functional Specialization

of Human Salivary Glands and Origins of Proteins Intrinsic to Human Saliva. *Cell Rep.* **33**, 108402 (2020). doi:10.1016/j.celrep.2020.108402 Medline
92. W. J. Kent, BLAT—The BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002). Medline
93. G. Marçais, A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg, A. Zimin, MUMmer4: A fast and versatile genome alignment system. *PLOS Comput. Biol.* **14**, e1005944 (2018). doi:10.1371/journal.pcbi.1005944 Medline
94. S. Kurtz, A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S. L. Salzberg, Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004). doi:10.1186/gb-2004-5-2-r12 Medline
95. R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, A. J. Drummond, BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014). doi:10.1371/journal.pcbi.1003537 Medline
96. G. Renaud, U. Stenzel, J. Kelso, leeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014). doi:10.1093/nar/gku699 Medline
97. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi:10.1093/bioinformatics/btp324 Medline
98. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W1605 (2016). doi:10.1093/nar/gkw257 Medline
99. B. S. Pedersen, A. R. Quinlan, Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018). doi:10.1093/bioinformatics/btx699 Medline
100. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015). doi:10.1093/nar/gkv416 Medline
101. Y. Shimoyama, pyGenomeViz: A genome visualization python package for comparative genomics (2022); https://github.com/moshi4/pyGenomeViz.
102. C.-S. Chin, S. Behera, A. Khalak, F. J. Sedlazeck, P. H. Sudmant, J. Wagner, J. M. Zook, Multiscale analysis of pangenomes enables improved representation of genomic diversity for repetitive and clinically relevant genes. *Nat. Methods* **20**, 1213–1221 (2023). doi:10.1038/s41592-023-01914-y Medline
103. K. H. M. A. A. C. T. C. Hsieh, K. H. Ma, A. Chao, iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451–1456 (2016). doi:10.1111/2041-210X.12613
104. D. V. Hinkley, Inference about the change-point in a sequence of random variables. *Biometrika* **57**, 1–17 (1970). doi:10.1093/biomet/57.1.1
105. F. J. Martin, M. R. Amode, A. Aneja, O. Austine-Orimoloye, A. G. Azov, I. Barnes, A. Becker, R. Bennett, A. Berry, J. Bhai, S. K. Bhurji, A. Bignell, S. Boddu, P. R. Branco Lins, L. Brooks, S. B. Ramaraju, M. Charkhchi, A. Cockburn, L. Da Rin Fiorretto, C. Davidson, K. Dodiya, S. Donaldson, B. El Houdaigui, T. El Naboulsi, R. Fatima, C. G. Giron, T. Genez, G. S. Ghattaoraya, J. G. Martinez, C. Guijarro, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, M. Kay, V. Kaykala, T. Le, D. Lemos, D. Marques-Coelho, J. C. Marugán, G. A. Merino, L. P. Mirabueno, A. Mushtaq, S. N. Hossain, D. N. Ogeh, M. P. Sakthivel, A. Parker, M. Perry, I. Piližota, I. Prosovetskaia, J. G. Pérez-Silva, A. I. A. Salam, N. Saraiva-Agostinho, H. Schuilenburg, D. Sheppard, S. Sinha, B. Sipos, W. Stark, E. Steed, R. Sukumaran, D. Sumathipala, M.-M. Suner, L. Surapaneni, K. Sutinen, M. Szpak, F. F. Tricomi, D. Urbina-Gómez, A. Veidenberg, T. A. Walsh, B. Walts, E. Wass, N. Willhoft, J. Allen, J. Alvarez-Jarreta, M. Chakiachvili, B. Flint, S. Giorgetti, L. Haggerty, G. R. Ilsley, J. E. Loveland, B. Moore, J. M. Mudge, J. Tate, D. Thybert, S. J. Trevanion, A. Winterbottom, A. Frankish, S. E. Hunt, M. Ruffier, F. Cunningham, S. Dyer, R. D. Finn, K. L. Howe, P. W. Harrison, A. D. Yates, P. Flicek, Ensembl 2023. *Nucleic Acids Res.* **51** (D1), D933–D941 (2023). doi:10.1093/nar/gkac958 Medline
106. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017). doi:10.1038/nmeth.4285 Medline
107. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018). doi:10.1093/molbev/msx281 Medline
108. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995). doi:10.1111/j.2517-6161.1995.tb02031.x
109. D. Reich, R. E. Green, M. Kircher, J. Krause, N. Patterson, E. Y. Durand, B. Viola, A. W. Briggs, U. Stenzel, P. L. F. Johnson, T. Maricic, J. M. Good, T. Marques-Bonet, C. Alkan, Q. Fu, S. Mallick, H. Li, M. Meyer, E. E. Eichler, M. Stoneking, M. Richards, S. Talamo, M. V. Shunkov, A. P. Derevianko, J.-J. Hublin, J. Kelso, M. Slatkin, S. Pääbo, Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010). doi:10.1038/nature09710 Medline
110. Q. Fu, H. Li, P. Moorjani, F. Jay, S. M. Slepchenko, A. A. Bondarev, P. L. F. Johnson, A. Aximu-Petri, K. Prüfer, C. de Filippo, M. Meyer, N. Zwyns, D. C. Salazar-García, Y. V. Kuzmin, S. G. Keates, P. A. Kosintsev, D. I. Razhev, M. P. Richards, N. V. Peristov, M. Lachmann, K. Douka, T. F. G. Higham, M. Slatkin, J.-J. Hublin, D. Reich, J. Kelso, T. B. Viola, S. Pääbo, Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014). doi:10.1038/nature13810 Medline
111. M. G. Llorente, E. R. Jones, A. Eriksson, V. Siska, K. W. Arthur, J. W. Arthur, M. C. Curtis, J. T. Stock, M. Coltorti, P. Pieruccini, S. Stretton, F. Brock, T. Higham, Y. Park, M. Hofreiter, D. G. Bradley, J. Bhak, R. Pinhasi, A. Manica, Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350**, 820–822 (2015). doi:10.1126/science.aad2879 Medline
112. C. Jeong, A. T. Ozga, D. B. Witonsky, H. Malmström, H. Edlund, C. A. Hofman, R. W. Hagan, M. Jakobsson, C. M. Lewis, M. S. Aldenderfer, A. Di Rienzo, C. Warinner, Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7485–7490 (2016). doi:10.1073/pnas.1520844113 Medline
113. C. M. Schlebusch, H. Malmström, T. Günther, P. Sjödin, A. Coutinho, H. Edlund, A. R. Munters, M. Vicente, M. Steyn, H. Soodyall, M. Lombard, M. Jakobsson, Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* **358**, 652–655 (2017). doi:10.1126/science.aao6266 Medline
114. C. E. G. Amorim, S. Vai, C. Posth, A. Modi, I. Koncz, S. Hakenbeck, M. C. La Rocca, B. Mende, D. Bobo, W. Pohl, L. P. Baricco, E. Bedini, P. Francalacci, C. Giostra, T. Vida, D. Winger, U. von Freeden, S. Ghirotto, M. Lari, G. Barbujani, J. Krause, D. Caramelli, P. J. Geary, K. R. Veeramah, Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nat. Commun.* **9**, 3547 (2018). doi:10.1038/s41467-018-06024-4 Medline
115. C. de la Fuente, M. C. Ávila-Arcos, J. Galimany, M. L. Carpenter, J. R. Homburger, A. Blanco, P. Contreras, D. Cruz Dávalos, O. Reyes, M. San Roman, A. Moreno-Estrada, P. F. Campos, C. Eng, S. Huntsman, E. G. Burchard, A.-S. Malaspinas, C. D. Bustamante, E. Willerslev, E. Llop, R. A. Verdugo, M. Moraga, Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4006–E4012 (2018). doi:10.1073/pnas.1715688115 Medline
116. C. Ning, T. Li, K. Wang, F. Zhang, T. Li, X. Wu, S. Gao, Q. Zhang, H. Zhang, M. J. Hudson, G. Dong, S. Wu, Y. Fang, C. Liu, C. Feng, W. Li, T. Han, R. Li, J. Wei, Y. Zhu, Y. Zhou, C.-C. Wang, S. Fan, Z. Xiong, Z. Sun, M. Ye, L. Sun, X. Wu, F. Liang, Y. Cao, X. Wei, H. Zhu, H. Zhou, J. Krause, M. Robbeets, C. Jeong, Y. Cui, Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nat. Commun.* **11**, 2700 (2020). doi:10.1038/s41467-020-16557-2 Medline
117. Y. Bussi, R. Kapon, Z. Reich, Large-scale k-mer-based analysis of the informational properties of genomes, comparative genomics and taxonomy. *PLOS ONE* **16**, e0258693 (2021). doi:10.1371/journal.pone.0258693 Medline
118. K. Scheer, Zenodo Downsampling Infomation, Zenodo (2024); https://zenodo.org/doi/10.5281/zenodo.13251208.
119. N.-C. Chen, B. Solomon, T. Mun, S. Iyer, B. Langmead, Reference flow: Reducing reference bias using multiple population genomes. *Genome Biol.* **22**, 8 (2021). doi:10.1186/s13059-020-02229-3 Medline
120. S. Myers, R. Bowden, A. Tumian, R. E. Bontrop, C. Freeman, T. S. MacFie, G. McVean, P. Donnelly, Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* **327**, 876–879 (2010). doi:10.1126/science.1182363 Medline

121. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018). doi:10.1093/bioinformatics/bty191 Medline
122. G. Tesler, GRIMM: Genome rearrangements web server. *Bioinformatics* **18**, 492–493 (2002). doi:10.1093/bioinformatics/18.3.492 Medline
123. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002). doi:10.1093/nar/gkf436 Medline
124. C. M. B. Carvalho, J. R. Lupski, Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* **17**, 224–238 (2016). doi:10.1038/nrg.2015.25 Medline
125. P. Lindenbaum, JVarkit: java-based utilities for Bioinformatics. figshare (2015); https://doi.org/10.6084/M9.FIGSHARE.1425030.
126. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021). doi:10.1093/gigascience/giab008 Medline
127. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin; 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011). doi:10.1093/bioinformatics/btr330 Medline
128. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi:10.1086/519795 Medline
129. J.-H. Shin, S. Blay, B. McNeney, J. Graham, LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. *J. Stat. Softw.* **16** (Code Snippet 3), 1–9 (2006). doi:10.18637/jss.v016.c03
130. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, Ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017). doi:10.1111/2041-210X.12628
131. P. K. Albers, G. McVean, Dating genomic variants and shared ancestry in population-scale sequencing data. *PLOS Biol.* **18**, e3000586 (2020). doi:10.1371/journal.pbio.3000586 Medline

ACKNOWLEDGMENTS

We thank Sabriya Syed, Valerie Reyes-Ortiz, Nicholas Moskwa, Pille Hallast, Carmen Robinett, Stefan Ruhl, Victor Albert, Vincent Lynch, Derek Taylor, Leo Speidel, and John Novembre for technical help, discussions, suggestions, and feedback on the manuscript; Joshua Akey and Liming Li for help analyze Neanderthal introgression data; the Human Genome Structural Variation Consortium and the Human Pangenome Reference Consortium for making their data publicly available; The Jackson Laboratory Scientific Services, including the Genome Technologies Service for expert assistance with the work described herein, and Research IT for computational infrastructure and support. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on December 1, 2023. We are grateful to the people who generously contributed their samples to the 1000 Genomes Project. **Funding:** C.L., and F.Y. were supported by the National Institute of Health (NIH) U24HG007497; K.K. was supported by The Jackson Laboratory Postdoctoral Scholar Award; O.G. was supported by the National Science Foundation (NSF) Awards # 2049947 and #2123284; C.R.B was supported by NIH GM133600. **Author contributions:** O.G. and C.L. conceived the study. F.Y. performed the analysis and interpretation of the Human Genome Structural Variation Consortium (HGSVC), and Human Pangenome Reference Consortium (HPRC) Bionano Genomics optical mapping, PacBio HiFi sequencing and phased assemblies, performed haplotype-resolved assemblies of HGSVC PacBio HiFi samples using hifiasm and amylase haplotype detection using HGSVC, and HPRC datasets. C.K. performed the mutational mechanisms analyses, breakpoint characterization, linkage disequilibrium, PCA and phylogenetic analyses. C.K. contributed to evolutionary and functional analysis. K.K. and F.Y. performed gene annotation, K.K. performed singleton analysis, selection/phylogenetic analyses of amylase coding sequences/segments, and interpretation/time estimation of initial *AMY1* gene duplication events. F.Y. and K.K. performed archaic hominin genome processing. P.P. performed ddPCR validation experiments, analysis of functional site differences in amylase amino acid sequences, and data visualization. K.S. calculated amylase gene copy numbers and performed downsampling in ancient human genomes. F.Y., C.K., K.K., P.P., C.R.B., A-M.T., C.L. and O.G. drafted and critically revised the article. Final approval of the version to be published was given by F.Y., C.K., K.K., P.P., K.S., C.R.B., A-M. T, C.L. and O.G. All authors read and approved the final manuscript. **Competing interests:** C.L. is a scientific advisory board member of Nabsys and Genome Insight. **Data and materials availability:** Data files used by this project are available by the following project IDs: Bionano Genomics optical genome mapping: PRJEB41077, PRJEB58376, PRJEB77842, PRJNA315896, PRJNA701308; PacBio HiFi: PRJEB58376, PRJEB75190, PRJNA701308; hifiasm assemblies: PRJEB78558; HPRC phased assemblies: PRJNA701308; ancient human genomes: PRJEB11364, PRJEB11995, PRJEB22660, PRJEB24629, PRJEB26760, PRJEB31045, PRJEB33172, PRJEB36297, PRJEB36854, PRJEB38008, PRJEB41240, PRJEB4604, PRJEB50857, PRJEB56570, PRJEB6272, PRJEB64656, PRJEB6622, PRJEB71770, PRJEB9783, PRJNA229448, PRJNA240906, PRJNA295861, PRJNA299403, PRJNA433631, PRJNA46213, PRJNA670050, PRJNA778930; archaic hominins: PRJEB10597, PRJEB10828, PRJEB1265, PRJEB2065, PRJEB20653, PRJEB21157, PRJEB21195, PRJEB21870, PRJEB21875, PRJEB21881, PRJEB21882, PRJEB21883, PRJEB24663, PRJEB29475, PRJEB3092, PRJEB31410, PRJEB55327. Amylase segments, haplotypes, orthogonal validation of haplotypes and molecule support, and liftoff gene annotation files (*16*); amylase segments CDS alignments, dating of *AMY1*, dN/dS analysis, origin of *AMY1* break point, protein2genome gene annotations, tandem repeat analysis files (*22*); ddPCR protocol (*21*) and functional analysis files (*24*); principal component analysis, breakpoint junction characterization, genomic rearrangement analysis, alignments, nucleotide diversity estimation, linkage disequilibrium estimation, phylogenetic trees, PRDM9 binding sites (*67*); structural variation mechanisms files (*65*); and ancient genomes downsampling analysis file information (*30*) are available at Zenodo.

Human Genome Structural Variation Consortium

Peter A. Audano[7], Olanrewaju Austine-Orimoloye[8], Christine R. Beck[7,9], Evan E. Eichler[10], Pille Hallast[7], William T. Harvey[10], Alex R. Hastie[11], Kendra Hoekzema[10], Sarah Hunt[8], Jan O. Korbel[12], Jennifer Kordosky[10], Charles Lee[7], Alexandra P. Lewis[10], Tobias Marschall[13], Katherine M. Munson[10], Andy Pang[11], Feyza Yilmaz[7]

[7]The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA. [8]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [9]The University of Connecticut Health Center, Farmington, CT, USA. [10]Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA. [11]Bionano Genomics, San Diego, CA, USA. [12]European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany. [13]Institute for Medical Biometry and Bioinformatics, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adn0609
Materials and Methods
Supplementary Text
Figs. S1 to S40
Tables S1 to S15

References (*101*–*131*)
MDAR Reproducibility Checklist

Submitted 27 November 2023; resubmitted 27 May 2024
Accepted 24 September 2024
Published online 17 October 2024

10.1126/science.adn0609

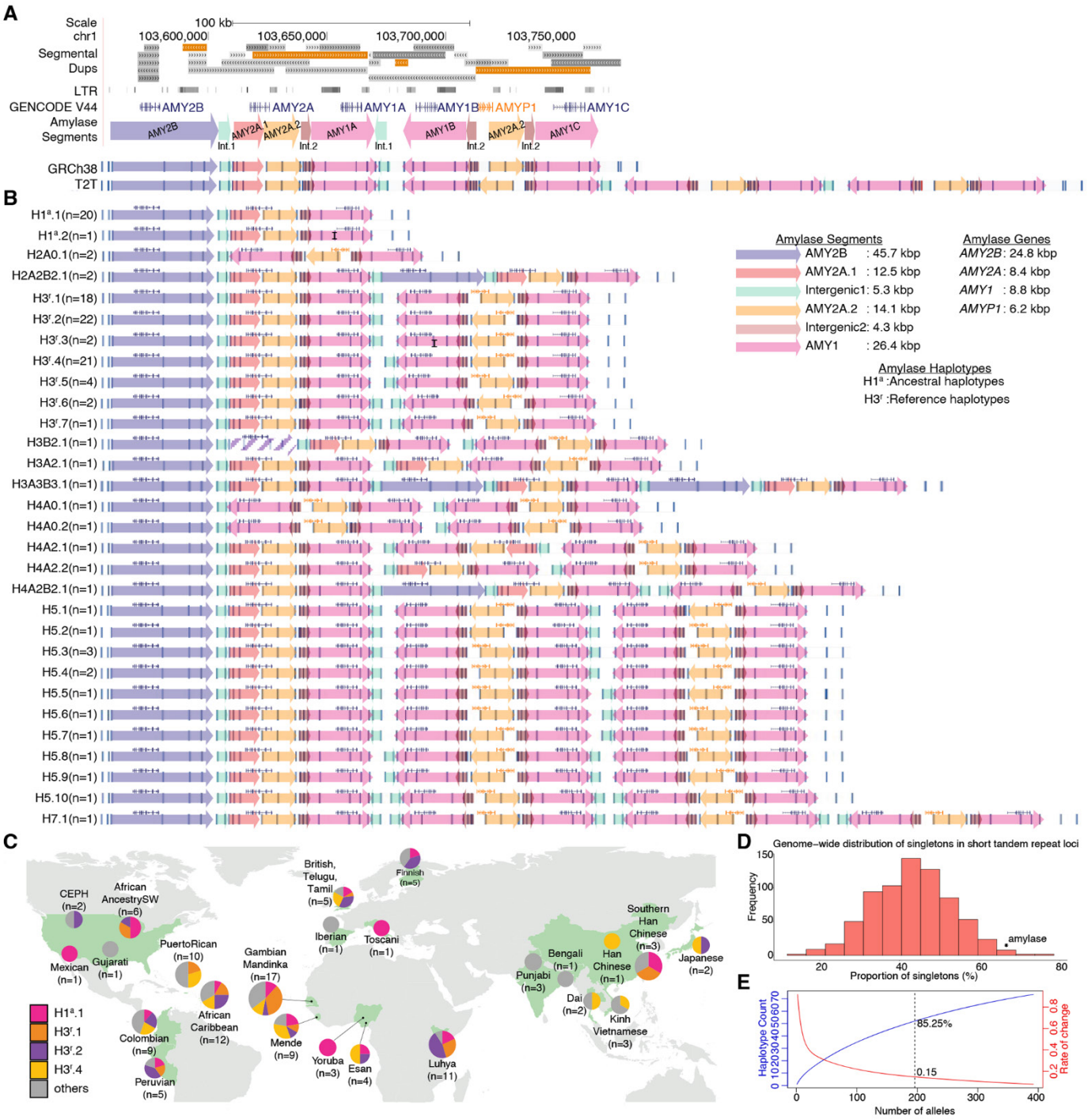

Fig. 1. Amylase structural haplotypes identified from present-day humans in this study. (**A**) Segmental duplications (light to dark gray: 90-98% similarity, light to dark orange: > 99% similarity), GENCODE V44 gene annotations, and long terminal repeats (LTRs) are represented as tracks. The lower panel shows amylase segments (colored arrows), and haplotype structures of GRCh38 and T2T-chm13 reference assemblies, represented as in silico maps with white backgrounds and vertical blue lines displaying optical mapping labels. The AMY2B segment overlaps the *AMY2B* gene, AMY2A.1 and AMY2A.2 segments overlap the *AMY2A* gene, and the AMY1 segment overlaps the *AMY1* gene. (**B**) The high-confidence amylase structural haplotypes resolved in our dataset (n = 30). The vertical black line in the second AMY1 segment of H3^{r}.3 represents the polymorphic label present in three alleles. The diagonal stripes in the second AMY2B segment of H3B2.1 indicate that it is a partial copy of the first AMY2B segment. Haplotype IDs describe: HX: X denotes the number of *AMY1* copies; AX: X denotes the number of *AMY2A* copies; BX: X denotes the number of *AMY2B* copies. The superscript "a'' denotes the ancestral amylase haplotype structure, and the superscript "r" denotes the reference amylase haplotype structure. The number in parentheses indicates the number of alleles. (**C**) The distribution of common amylase haplotypes across 26 population samples. (**D**) The proportion of singletons for tandem repeat loci (EnsembleTR) across the genome. For adequate comparison, we used the same individuals (n = 33) for whom we were able to reconstruct amylase haplotypes in our dataset. Additionally, we filtered the tandem repeat loci (719 loci, unit length 1-6 bp) that we analyzed to match the number of distinct alleles (n = 21) observed in the amylase locus. The asterisk (*) represents the proportion of singletons among all distinct haplotypes (~67%, 14 out of 21) detected at the amylase locus. (**E**) Rarefaction and extrapolation sampling curve based on 52 amylase haplotypes, displaying how the number of distinct haplotypes (blue line) is projected to saturate with the increase in the number of alleles. The rate of change (red line, 0.15) indicates the number of novel haplotypes discovered per unit increase in the number of analyzed alleles. The dashed line shows the proportion of estimated number of samples (85.25%) captured in our study.

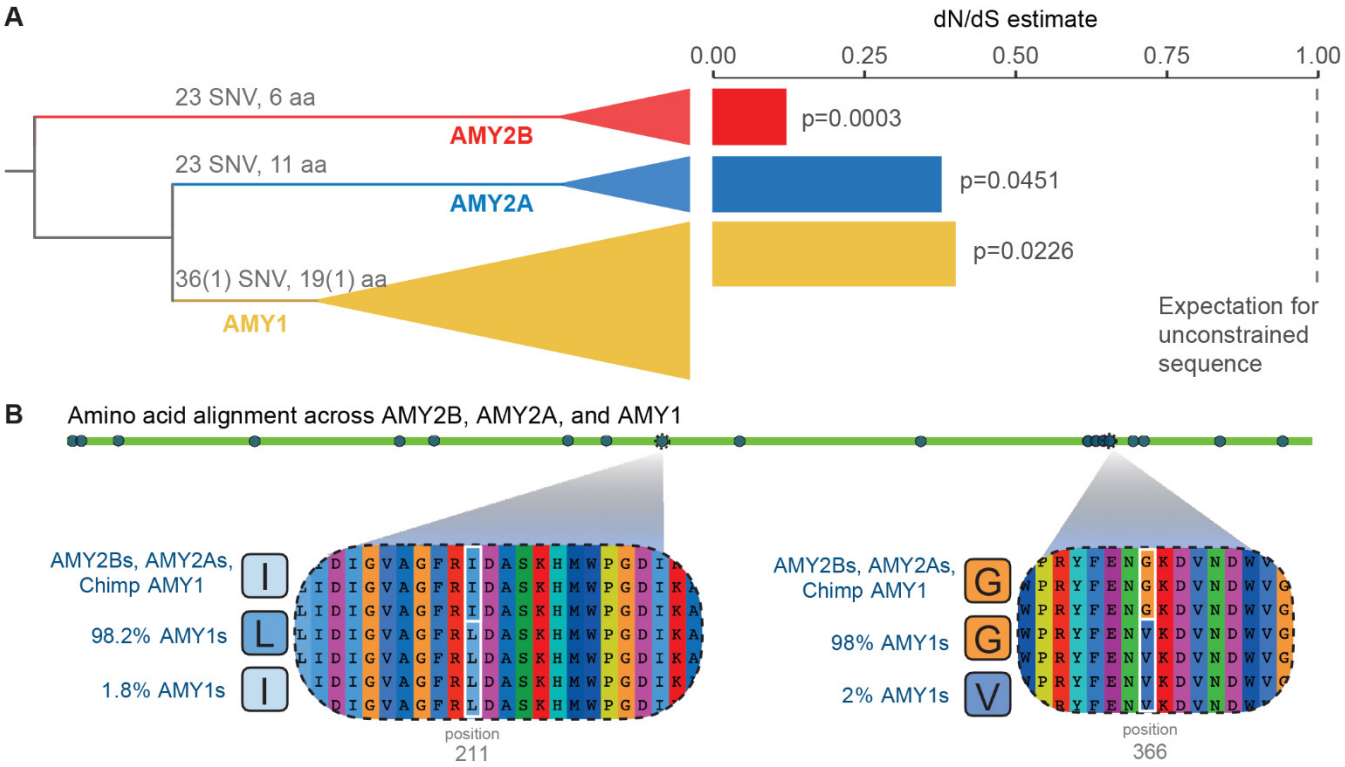

Fig. 2. The variants in amylase coding sequences and negative selection on three amylase gene types. (**A**) The maximum likelihood phylogenetic tree of amylase coding sequences (left) and dN/dS estimate for each amylase gene type (right). The phylogenetic tree is rooted with a coding sequence from the sheep genome (Oar_rambouillet_v1.0). The number of nucleotide and resultant amino acid changes that are paralog-specific are indicated. The numbers in parenthesis indicate nucleotide and amino acid changes that are variable within the *AMY1* branch. The numbers to the right of each bar represent the FDR-adjusted p-value for the likelihood ratio between $H0_2$ (dN/dS ratio is fixed to one on the foreground branches) and H1 (two dN/dS ratios are allowed on the foreground and background branches, respectively) for each gene type. (**B**) The positions of amino acid variants within and between *AMY2B, AMY2A* and *AMY1* protein sequences. The 211 and 366 positions are highlighted because they overlap with a conserved section of the amylase protein sequence and have a predicted functional impact (AlphaMissense (*90*). The coordinates are based on the residues of the amylase enzyme from UniProtKB (Accession: P0DUB6).

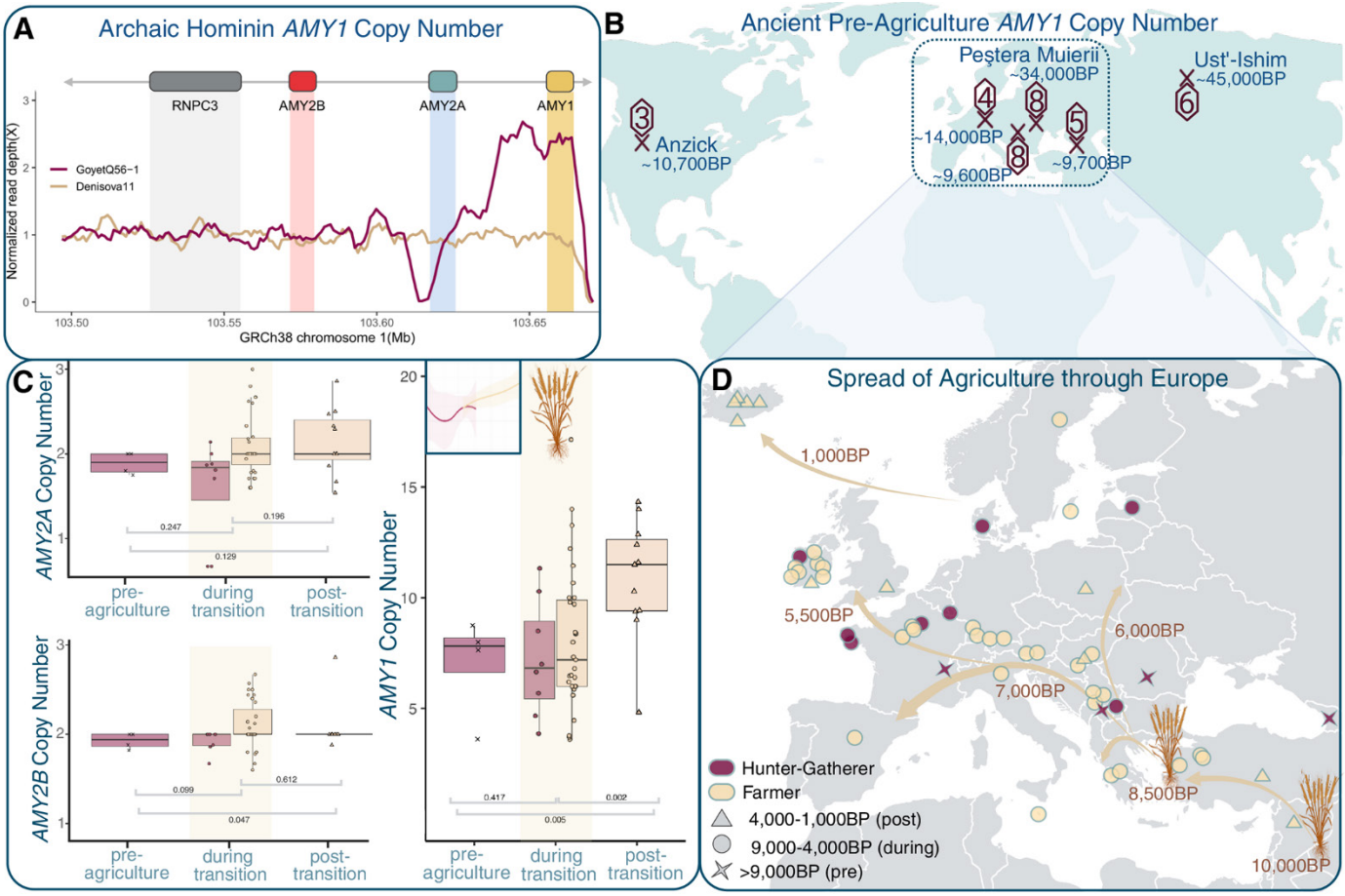

Fig. 3. Amylase gene duplication and the history of agriculture. (**A**) The read-depth of amylase locus spanning *RNPC3*, *AMY2B*, *AMY2A*, and *AMY1* genes (chr1:103494306-103668306) for GoyetQ56-1 (maroon line), a Neanderthal excavated in present-day Belgium, showing signatures of *AMY1* duplication, and for Denisova 11 (beige line), a hybrid hominin (Neanderthal and Denisovan) excavated from present-day Russia, showing signatures consistent with an ancestral single-copy AMY1 haplotype. The maroon and beige lines indicate average read-depths with 5-kbp window and 1-kbp step for each sample. The average read-depth of each 5-kbp window was normalized by the average read-depth of the *RNPC3* gene. Only uniquely mapped reads were used for this visualization. (**B**) A world map displaying the locations of ancient human samples. Sample locations are indicated with an 'X', with corresponding hexagons showing the estimated *AMY1* copy number. Carbon dating (number of years Before Present; BP) estimated for each sample is indicated in blue. (**C**) Amylase copy number estimations from Europeans who were farmers (beige) and hunter-gatherers (maroon). Samples are binned on the x-axis according to three time periods; pre-agriculture (before the transition to agriculture, >9,000 BP), during the transition (~ 9,000 - 4,000 BP), and post-transition (complete transition to agriculture, ~ 4,000 BP - 1000 BP). The shape of data points corresponds to sample dating estimates (Xs for more than 9,000 BP, circles for 9,000 - 4,000 BP, and triangles for 4,000 BP - 1000 BP). The inset of the right panel shows nonparametric regression lines for *AMY1* copy number across time, for hunter-gatherers (maroon) and farmers (beige), with confidence intervals in lighter corresponding color, respectively. (**D**) Zoomed-in map showing the spread of agriculture into Europe from Asia. Major agricultural footholds are indicated by wheat pictograms. Tan arrows show general trends of human agricultural migration throughout Europe, with predicted time periods (*58*, *59*). Ancient human samples that were analyzed for amylase copy number are annotated with shapes and colors for time period and lifestyle, respectively. Figure created using Biorender.

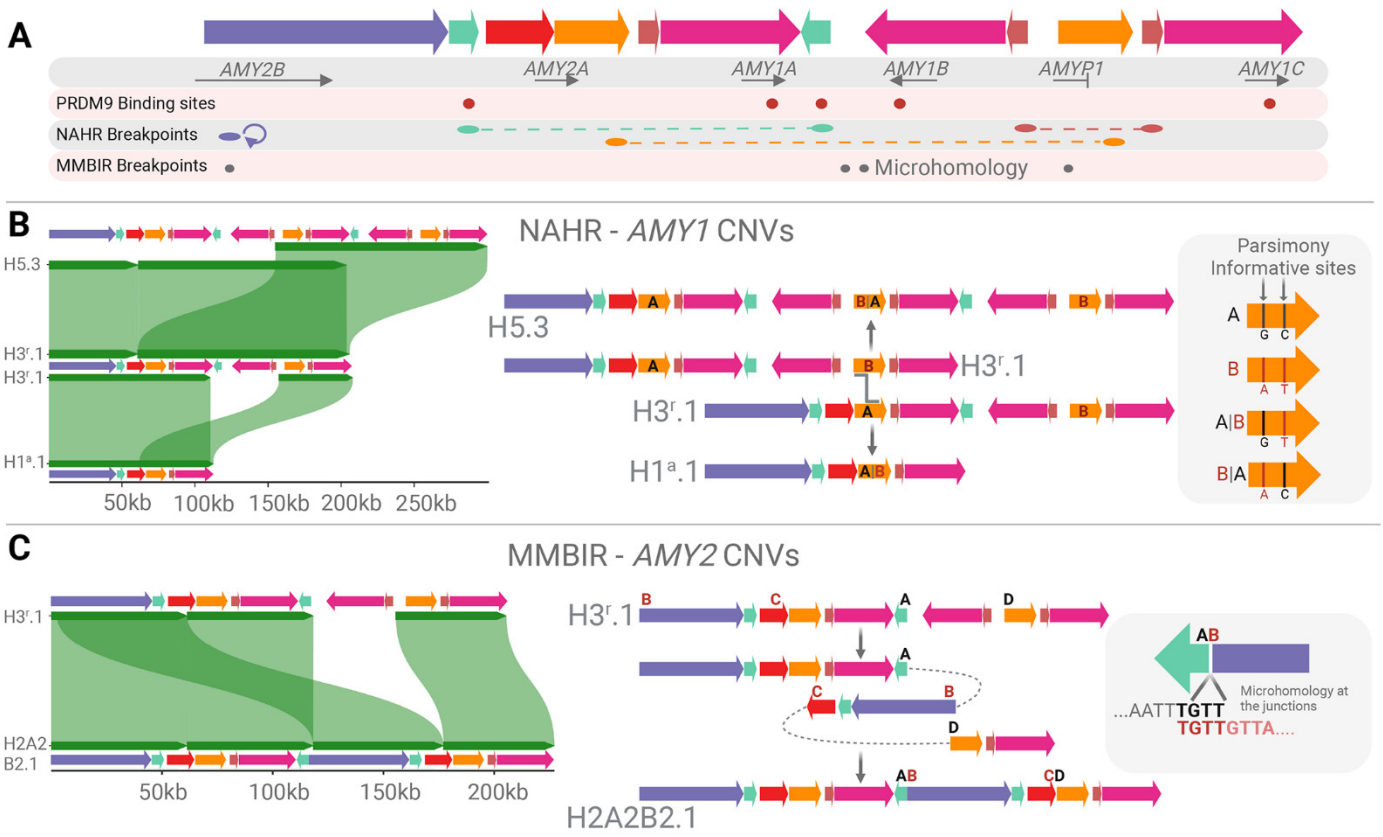

Fig. 4. The evolutionary and mutational connections among common haplotypes. (**A**) The structural variation breakpoints and recombination hotspots in the amylase locus. The colored arrows represent amylase segments. The PRDM9 binding sites are represented with red dots. The non-allelic homologous recombination (NAHR) breakpoints are represented with purple, green and orange dots, and dashed lines. The microhomology-mediated break-induced replication breakpoints are represented with gray dots. (**B**) NAHR-mediated duplication and deletion of the *AMY1A-AMY1B* cluster. The AMY2.2 (orange) segment serves as the recombination substrate for the crossover, resulting in the duplication or deletion of the *AMY1A-AMY1B* cluster as illustrated in the middle panel. Chimeric AMY2.2 segments have been identified, utilizing parsimony informative sites within the AMY2.2 segment (right panel). (**C**) Microhomology-mediated break-induced replication based copy-number gain resulting in the formation of H2A2B2.1. The middle panel shows the mutational mechanism. Four nucleotides of microhomology internal to the breakends were identified at the breakpoint junction (right panel). Figure created using Biorender.

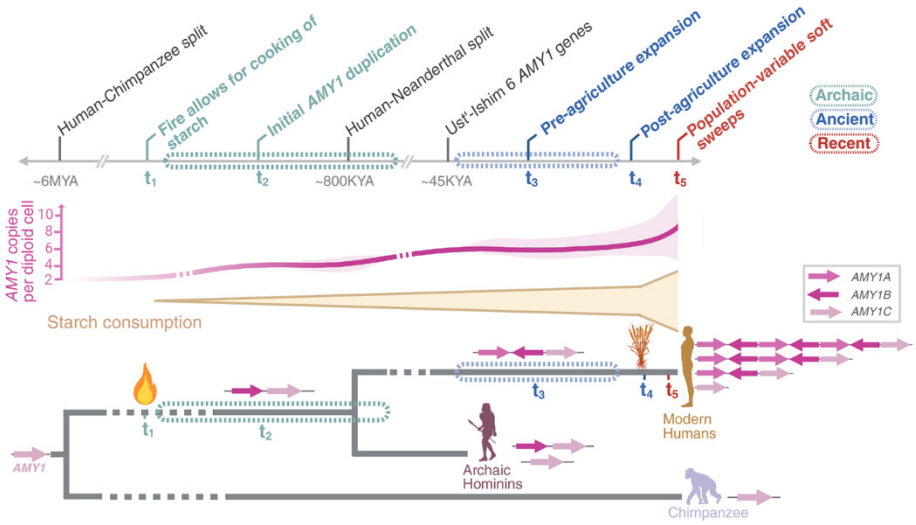

Fig. 5. An evolutionary model of the human amylase genes and resulting hypotheses. Top: A timeline of human amylase locus evolution based on the results of this study, with relevant events indicated on top. Middle: A schematic view showing the increase in *AMY1* copy number variation and the mean number of *AMY1* copies in present-day human populations throughout history as starch consumption increases. Bottom: A phylogenetic representation of the hypothesized amylase duplication timeline. Figure created using Biorender.