RESEARCH ARTICLE SUMMARY

ARCHAIC HOMININS

The *MUC19* gene: An evolutionary history of recurrent introgression and natural selection

Fernando A. Villanea†, David Peede†, Eli J. Kaufman, Valeria Añorve-Garibay, Elizabeth T. Chevy, Viridiana Villa-Islas, Kelsey E. Witt, Roberta Zeloni, Davide Marnetto, Priya Moorjani, Flora Jay, Paul N. Valdmanis, María C. Ávila-Arcos, Emilia Huerta-Sánchez*

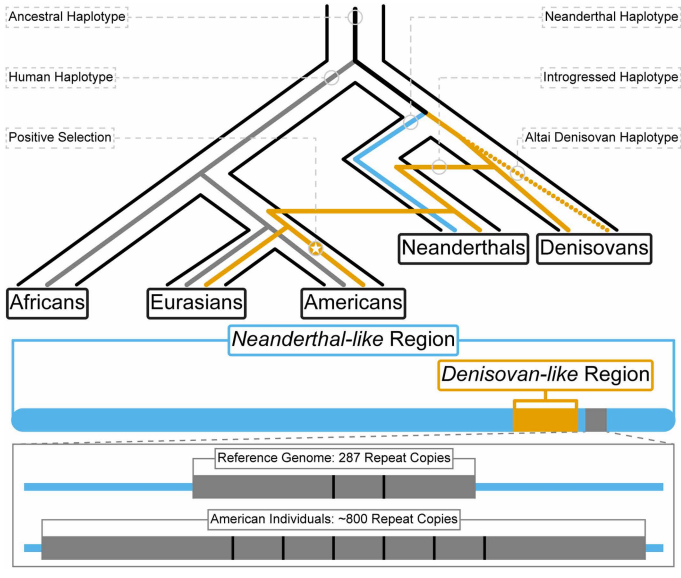
Full article and list of author affiliations: https://doi.org/10.1126/science.adl0882

INTRODUCTION: Modern human genomes contain a small number of archaic variants, the legacy of past interbreeding events with Neanderthals and Denisovans. Most of these variants are putatively neutral, but some archaic variants found in modern humans have been targets of positive natural selection and may have been pivotal for adapting to new environments as humans populated the world. American populations encountered a myriad of novel environments, providing the opportunity for natural selection to favor archaic variants in these new environmental contexts. Indigenous and admixed American populations have been understudied in this regard but present great potential for studying the underlying evolutionary processes of local adaptation.

RATIONALE: Previous studies identified the gene *MUC19*—which codes for a mucin involved in immunity—as a candidate for introgression from Denisovans as well as a candidate for positive natural selection in present-day Indigenous and admixed American populations. Therefore, we sought to confirm and further characterize signatures of both archaic introgression and positive selection at *MUC19*, with particular interest in modern and ancient American populations.

RESULTS: We identify an archaic haplotype segregating at high frequency in most admixed American populations, and among ancient genomes from 23 ancient Indigenous American individuals who predate admixture with Europeans and Africans. We conclude that the archaic haplotype has undergone positive natural selection in these populations, which is tied to their Indigenous components of ancestry. We also find that modern admixed American individuals exhibit an elevated number of variable number tandem repeats (VNTRs) at *MUC19*, which codes for the functional domain of the MUC19 protein, where it binds to oligosaccharides to form a glycoprotein, a characteristic of the mucins. Remarkably, we find an association between the number of VNTRs and the number of introgressed haplotypes; individuals harboring introgressed haplotypes tend to have a higher number of VNTRs. In addition to the differences in VNTRs, we find that the archaic *MUC19* haplotype contains nine Denisovan-specific, nonsynonymous variants found at high frequencies in American populations. Finally, we observed that the Denisovan-specific variants are contained in a 72-kb region of the *MUC19* gene, but that region is embedded in a larger 742-kb region that contains Neanderthal-specific variants. When we studied *MUC19* in three high-coverage Neanderthal individuals, we found that the Chagyrskaya and Vindija Neanderthals carry the Denisovan-like haplotype in its heterozygous form. These two Neanderthals also carry another haplotype that is shared with the Altai Neanderthals.

CONCLUSION: Our study identifies several aspects of the gene *MUC19* that highlight its importance for studying adaptive introgression: One of the haplotypes that span this gene in modern humans is of archaic origin, and modern humans inherited this haplotype from Neanderthals who likely inherited it from Denisovans. Then, as modern human populations expanded into the Americas, our results suggest that they experienced a massive coding VNTR expansion, which occurred on an archaic haplotype background in *MUC19*. The functional impact of the variation at this gene may help explain how mainland Indigenous Americans adapted to their environments, which remains underexplored. Our results point to a complex pattern of multiple introgression events, from Denisovans to Neanderthals and Neanderthals to modern humans, which may have later played a distinct role in the evolutionary history of Indigenous American populations. □



The proposed evolutionary history of *MUC19*. The Denisovan-like haplotype (in orange) was first introgressed from Denisovans into Neanderthals and then introgressed into modern humans. The introgressed haplotype later experienced positive selection in populations from the Americas. The introgressed *MUC19* haplotype is composed of a 742-kb region that contains Neanderthal-specific variants (blue). Embedded within this Neanderthal-like region is a 72-kb region containing a high density of Denisovan-specific variants (orange), and an exonic variable number tandem repeat (VNTR) region (gray). The box below the 742-kb region depicts zooming into the *MUC19* VNTR region, in which admixed American individuals carry an elevated number of tandem repeat copies.

*Corresponding author. Email: emilia_huerta-sanchez@brown.edu †These authors contributed equally to this work. Cite this article as F. Villanea *et al.*, *Science* **389**, eadl0882 (2025). DOI:10.1126/science.adl0882

ARCHAIC HOMININS

The *MUC19* gene: An evolutionary history of recurrent introgression and natural selection

Fernando A. Villanea[1]†, David Peede[2,3,4]†, Eli J. Kaufman[5], Valeria Añorve-Garibay[3], Elizabeth T. Chevy[3], Viridiana Villa-Islas[6,7], Kelsey E. Witt[8], Roberta Zeloni[9,10], Davide Marnetto[10], Priya Moorjani[11,12], Flora Jay[13], Paul N. Valdmanis[5], María C. Ávila-Arcos[6], Emilia Huerta-Sánchez[2,3,14,15]*

We study the gene *MUC19*, for which some modern humans carry a Denisovan-like haplotype. *MUC19* is a mucin, a glycoprotein that forms gels with various biological functions. We find diagnostic variants for the Denisovan-like *MUC19* haplotype at high frequencies in admixed American individuals and at highest frequency in 23 ancient Indigenous American individuals, all pre-dating population admixture with Europeans and Africans. We find that the Denisovan-like *MUC19* haplotype is under positive selection and carries a higher copy number of a 30–base-pair variable number tandem repeat, and that copy numbers of this repeat are exceedingly high in admixed American populations. Finally, we find that some Neanderthals carry the Denisovan-like *MUC19* haplotype, and that it was likely introgressed into modern human populations through Neanderthal introgression rather than Denisovan introgression.

Most modern humans of non-African ancestry carry both Neanderthal and Denisovan genomic variants (*1*–*3*). Although most of these variants are putatively neutral, some archaic variants found in modern humans have been targets of positive natural selection (*4*–*9*). Interbreeding with Neanderthals and Denisovans may have thereby facilitated adaptation to the myriad novel environments that modern humans encountered as they populated the globe (*10*). Indeed, several studies have identified signatures of adaptive introgression in Eurasian and Oceanian populations (*11*–*20*). Indigenous American populations, however, present great potential for studying the underlying evolutionary processes of local adaptation (*21*). In the 25,000 years since the first individuals populated the American continent, these populations would have encountered manifold novel environments, far different from the Beringian steppe, to which their ancestral population was adapted (*22*).

Previous studies identified *MUC19*—a gene involved in immunity—as a candidate for adaptive introgression among populations from the 1000 Genomes Project (1KG). These studies found the region surrounding *MUC19* to harbor a high density of Denisovan alleles in Mexicans (MXL) (*23*, *24*). *MUC19* was also reported to be under positive selection in North American Indigenous populations using population branch statistic (*PBS*) and integrated haplotype score (*iHS*) methods for detecting positive selection (*25*).

In this study, we confirm and further characterize signatures of both introgression and positive selection at *MUC19* in MXL. We find an archaic haplotype segregating at high frequency in most populations on the American continent, which is also present in two of the late high-coverage Neanderthal genomes, Chagyrskaya and Vindija. MXL individuals harbor Denisovan-specific coding mutations in *MUC19* at high frequencies and exhibit elevated copy number of a tandem repeat region within *MUC19* compared with other worldwide populations. Our results point to a complex pattern of multiple introgression events, from Denisovans to Neanderthals and Neanderthals to modern humans, which may have played a distinct role in the evolutionary history of Indigenous American populations.

Results

Signatures of adaptive introgression at MUC19 in admixed populations from the Americas

We compiled introgressed tracts that overlap the NCBI RefSeq coordinates for *MUC19* (hg19, Chr12:40787196-40964559) by at least one base pair (bp). Fig. 1A shows the density of introgressed tracts for all non-African populations in the region, using introgression maps inferred with hmmix (*26*). All non-African populations harbor introgressed tracts overlapping this region, but at much lower frequencies than the admixed American populations [(AMR) tract frequency: ~0.183, non-AMR tract frequency: ~0.087; proportions *Z*-test, *P*-value: 5.011e-14; Fisher's exact test, *P*-value: 2.144e-12; table S1). MXL, a population with a large component of Indigenous American genetic ancestry (~48%) (*27*), exhibit the highest frequency of the introgressed tracts (0.305; table S2). Given this, we examined a 742-kb window containing the longest introgressed tract found in MXL (hg19, Chr12:40272001-41014000; fig. S1). This region contains 135 Denisovan-specific single nucleotide polymorphisms (SNPs), classified as such because they are rare or absent in African populations (<1%), present in MXL (>1%), and shared only with the Altai Denisovan. All 135 of these SNPs are sequestered within a core 72-kb region (hg19, Chr12:40759001-40831000; shaded gray region in Fig. 1A) that has the highest introgressed tract density among individuals in the 1KG [see (*28*)], making both the 742-kb and 72-kb regions outliers for Denisovan-specific SNP density in MXL (742-kb region *P*-value: <3.164e-4; 72-kb region *P*-value: <3.389e-5; fig. S2 and tables S3 and S4). By contrast, there are 80 Neanderthal-specific SNPs in MXL found within the larger 742-kb region (*P*-value: 0.159; fig. S3 and table S5), with only four located in the 72-kb region (*P*-value: 0.263; fig. S3 and table S6).

To test whether natural selection is acting on this region, we computed three statistics. One developed to detect adaptive introgression ($U_{A,B,C}(w, x, y)$: *A*, African superpopulation; *B*, non-African populations; *C*, Altai Denisovan; (w, x, y) are allele frequency thresholds in *A*, *B* and *C*, respectively (*24*), and two for positive selection (*PBS* and *iHS*). For each gene, we computed $U_{AFR,B,Denisovan}(w = 1\%, x = 30\%, y = 100\%)$, which measures the number of Denisovan alleles found in the homozygous state (100%) that are almost absent in Africans (<1%) and reach a frequency of at least 30% in a given non-African population. Fig. 1B shows that *MUC19* in MXL is an extreme outlier, as no other gene in any non-African population exhibits such a large value of $U_{AFR,B,Denisovan}(1\%, 30\%, 100\%)$. When we compute the same statistic in windows instead of per gene, the *MUC19* region is an outlier only in MXL and is zero for all other non-African populations (*P*-value 72-kb region: <3.284e-5; *P*-value 742-kb region: <3.139e-4; fig. S4 and tables S7 and S8). Furthermore, we compared the windowed $U_{AFR,B,Denisovan}(w = 1\%, x = 30\%, y = 100\%)$ results with their

[1]Department of Anthropology, University of Colorado, Boulder, CO, USA. [2]Department of Ecology, Evolution, and Organismal Biology, Brown University, Providence, RI, USA. [3]Center for Computational Molecular Biology, Brown University, Providence, RI, USA. [4]Institute at Brown for Environment and Society, Brown University, Providence, RI, USA. [5]Division of Medical Genetics, Department of Medicine, University of Washington School of Medicine, Seattle, WA, USA. [6]International Laboratory for Human Genome Research, Universidad Nacional Autónoma de México, Juriquilla, Mexico. [7]Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [8]Center for Human Genetics and Department of Genetics and Biochemistry, Clemson University, Clemson, SC, USA. [9]University of Padova (Italy) - Department of Biology, Padua, Italy. [10]Department of Neurosciences "Rita Levi Montalcini", University of Turin, Turin, Italy. [11]Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA. [12]Center for Computational Biology, University of California, Berkeley, CA, USA. [13]Université Paris-Saclay, CNRS, INRIA, Laboratoire Interdisciplinaire des Sciences du Numérique, 91400, Orsay, France. [14]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. [15]Data Science Institute, Brown University, Providence, RI, USA. ***Corresponding author. Email: emilia_huerta-sanchez@brown.edu**
†These authors contributed equally to this work.

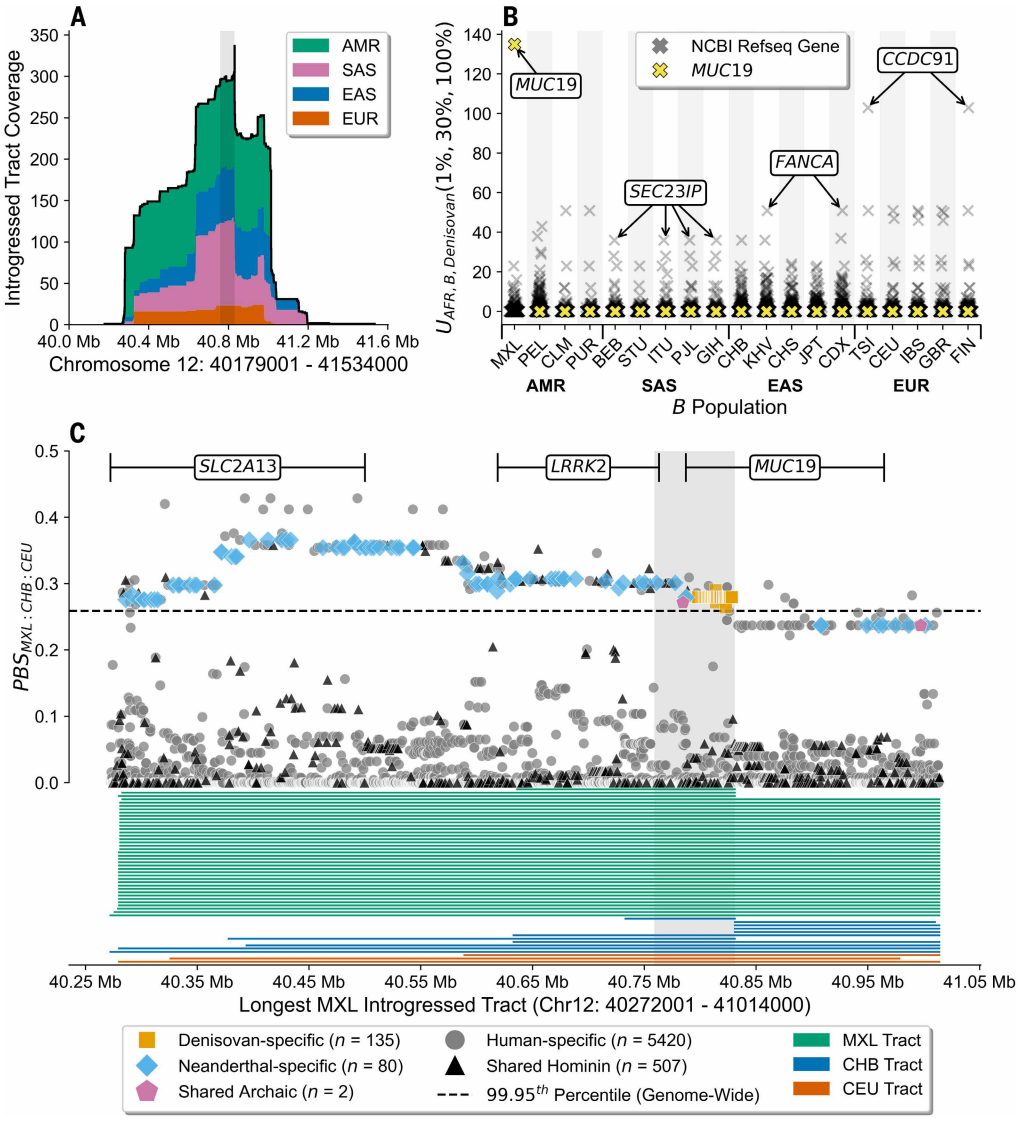


Fig. 1. Signals of adaptive introgression at *MUC19*. (**A**) Density of introgressed tracts inferred using hmmix that overlap *MUC19* for the 1KG (black outline) and stratified by superpopulation: Admixed Americans (AMR) in green, South Asians (SAS) in pink, East Asians (EAS) in blue, and Europeans (EUR) in orange. The gray shaded region corresponds to the focal 72-kb region, which is the densest contiguous region of introgressed tracts longer than 40-kb. (**B**) $U_{AFR,B,Denisovan}(1\%, 30\%, 100\%)$ values for each non-African population, stratified by superpopulation, per NCBI Refseq gene (gray X's), where *MUC19* is denoted as a yellow X. (**C**) Population branch statistics (*PBS*) for the Mexican population (MXL) in the 1KG using the Han Chinese (CHB) and Central European (CEU) populations in the 1KG as control populations ($PBS_{MXL:CHB:CEU}$) for all SNPs in the 742-kb region that correspond to the longest introgressed tract found in MXL. The orange squares represent Denisovan-specific SNPs; the blue diamonds represent Neanderthal-specific SNPs; and the pink pentagons represent shared archaic SNPs. Note that all of these archaic SNP partitions are rare or absent in African populations and present in MXL [see (*28*)]. The black triangles represent SNPs present across both modern human populations and the archaics whereas the gray circles represent SNPs private to modern humans. The black dashed line represents the 99.95th percentile of $PBS_{MXL:CHB:CEU}$ scores for all SNPs genome-wide, and the gray shaded region corresponds to the focal 72-kb region, the same gray shaded region in (A). The *MUC19* and *LRRK2* genes are fully encompassed within the 742-kb region whereas ~65% of *SLC2A13* overlaps the 742-kb region. Below the $PBS_{MXL:CHB:CEU}$ points are the introgressed tracts for MXL (green), CHB (blue), and CEU (orange) sorted from shortest to longest within each population.

corresponding $Q95_{AFR,B,Denisovan}(w = 1\%, y = 100\%)$ value, which quantifies the 95th percentile of the Denisovan allele frequencies found in a given non-African population *B* for the Denisovan alleles found in the homozygous state (100%) and are almost absent in Africans (<1%) (*24*). We find that for both the 72-kb and 742-kb *MUC19* regions, $Q95_{AFR,MXL,Denisovan}(w = 1\%, y = 100\%) = \sim 30\%$, which suggests that both the 72-kb and 742-kb *MUC19* regions exhibit signals consistent with adaptive introgression that are not observed in any other 1KG population (figs. S5 and S6 and table S9).

We next computed $PBS_{MXL:CHB:CEU}$, where the Han Chinese (CHB) and Central European (CEU) populations were used as control populations for both the region corresponding to the longest introgressed tract in MXL (742-kb) and the 72-kb region in *MUC19*. We find that both regions exhibit statistically significant $PBS_{MXL:CHB:CEU}$ values compared with other 742-kb ($PBS_{MXL:CHB:CEU}$: 0.066; *P*-value: 0.004) and 72-kb ($PBS_{MXL:CHB:CEU}$: 0.127; *P*-value: 0.002) windows of the genome, respectively (fig. S7 and tables S10 and S11). We then computed $PBS_{MXL:CHB:CEU}$ for each SNP in the 742-kb region. Fig. 1C shows that in MXL there are many SNPs with statistically significant *PBS* values in that region (417 out of 6144 SNPs), all which present values above the 99.95th percentile of genome-wide $PBS_{MXL:CHB:CEU}$ values [Benjamini-Hochberg corrected *P*-values: <0.01; see supplementary section S1 in (*28*)]. We note that some SNPs have a larger $PBS_{MXL:CHB:CEU}$ value near the *SLC2A13* gene than within the 72-kb *MUC19* region, but this is due to changes in the archaic allele frequency in CHB and CEU as the introgressed tracts in these populations are more sparse than the introgressed tracts in MXL (see tracts in Fig. 1C). When we partition the MXL population into two demes consisting of individuals with >50% and those with <50% Indigenous American ancestry genome-wide (*27*) and recompute *PBS*, we find that *PBS* values for archaic variants are elevated among individuals with a higher proportion of Indigenous American ancestry, suggesting that this region was likely targeted by selection before admixture with European and African populations (fig. S8 and tables S10 and S11).

To exclude the possibility that demographic events such as a founder effect explain the observed signatures of positive selection, we simulated the best-fitting demographic parameters inferred for the MXL population (*29*) to obtain the expected null distribution of *PBS* values. We first showed that *PBS* has power to detect adaptive introgression

under this demographic model [see supplementary section S1 in (*28*)]. We found that demographic forces alone result in lower *PBS* values compared with what is observed at both the 742-kb and 72-kb regions [see supplementary section S1 in (*28*)], even when we consider a very conservative null model of heterosis. Furthermore, to also consider haplotype-based measures of positive selection, we computed *iHS* for every 1KG population using selscan (*30*) to provide orthogonal evidence of positive selection (*28*). Among all 1KG populations, MXL is the only population with an elevated proportion of SNPs with normalized $|iHS| > 2$ in either the 742-kb (599 out of 2248 SNPs) or 72-kb region (229 out of 425 SNPs; tables S12 and S13). In MXL we find that 130 out of the 135 Denisovan-specific SNPs in the 72-kb region have normalized $|iHS| > 2$, reflective of positive selection [fig. S9 and tables S12 and S13; see supplementary section S2 in (*28*)], which supports our previous allele frequency–based tests of positive selection.

Admixed American individuals exhibit an elevated number of variable number tandem repeats at MUC19

MUC19 contains a 30-bp variable number tandem repeat [(VNTR); hg19, Chr12:-40876395-40885001; fig. S10], located 45.4 kb away from the core 72-kb haplotype but within the larger 742-kb introgressed region. To test whether individuals who harbor an introgressed tract overlapping the repeat region differ in the number of repeats compared with individuals who do not harbor introgressed tracts, we calculated the number of repeats of the 30-bp motif in the 1KG individuals [see (*28*); fig. S11 and tables S14 and S15]. For each individual, we first report the average number of repeats between their two chromosomes. The genomes of the four archaic individuals do not harbor a higher copy number of tandem repeats (Altai Denisovan: 296 copies; Altai Neanderthal: 379 copies; Vindija Neanderthal: 268 copies; and Chagyrskaya Neanderthal: 293 copies). Among all individuals from the 1KG, we identified outlier individuals with elevated number of repeats above the 95th percentile (>487 repeats; dashed line in Fig. 2). We found that MXL individuals have on average ~493 repeats and individuals from the admixed American superpopulation have on average ~417 repeats (Fig. 2A and tables S16 and S17). By contrast, nonadmixed American populations have an average of ~341 to ~365 repeats (Fig. 2A and table S16). Out of all the outlier individuals from the 1KG (>487 repeats), a significant proportion of them (~77%) are from admixed American populations (proportions *Z*-test, *P*-value: 3.971e-17; tables S18 to S21 and fig. S12). Outlier individuals from the Americas also carry a significantly higher copy number of tandem repeats compared with the other outlier individuals from nonadmixed American populations (Mann-Whitney *U* test, *P*-value: 5.789e-7; fig. S12 and tables S18 to S21). In MXL, we find that exactly 50% of individuals exhibit an elevated copy number of tandem repeats (table S16).

Within individuals exhibiting an outlier number of repeats (>487), a significant proportion (~86%) have an introgressed tract overlapping the repeat region and these individuals harbor an elevated number of repeats compared with outlying individuals who do not harbor an introgressed tract overlapping the VNTR region (proportions *Z*-test, *P*-value: 2.127e-29; Mann-Whitney *U* test, *P*-value: 1.398e-06; fig. S13 and tables S18 to S21). All outlying MXL individuals carry at least one

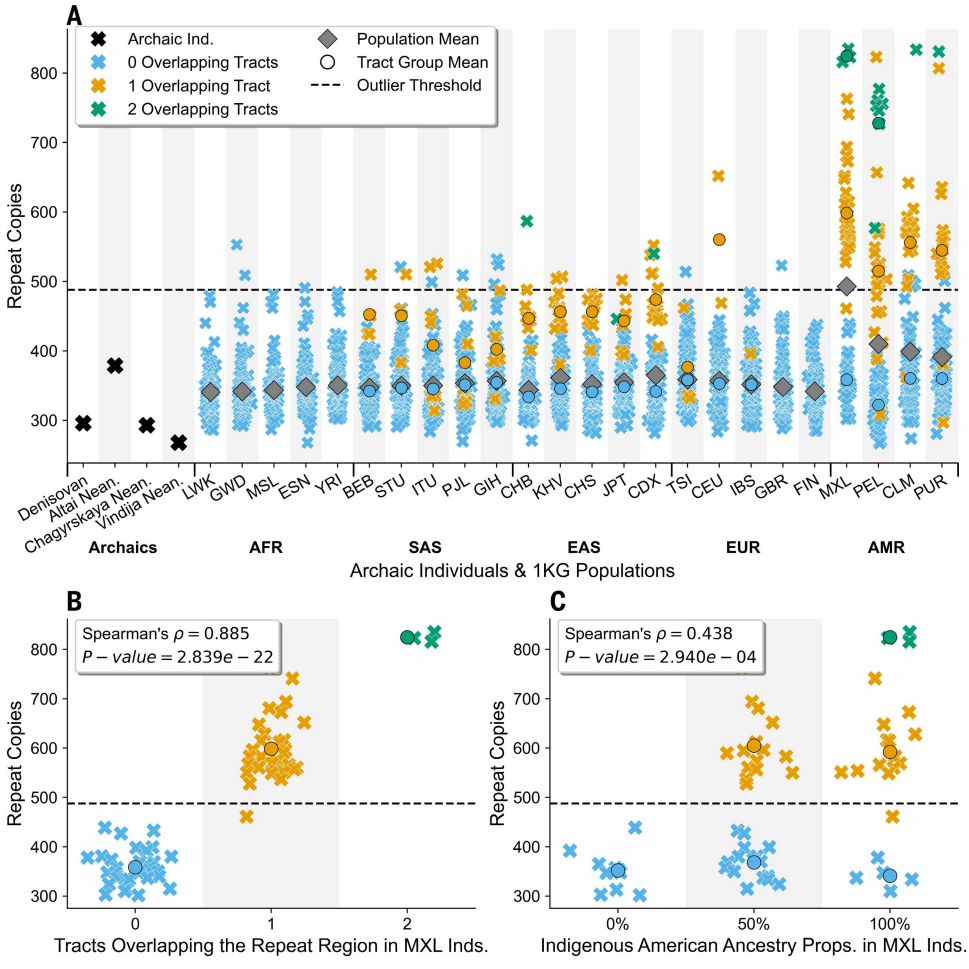


Fig. 2. Copy number variation of a 30-bp variable number tandem repeat motif in the 1KG individuals at *MUC19*. (**A**) Average number of repeat copies between an individual's two chromosomes for archaic individuals (black X's), individuals who do not harbor an introgressed tract (blue X's), individuals with one introgressed tract (orange X's), and individuals with two introgressed tracts (green X's) determined by the number of introgressed tracts inferred using hmmix overlapping the *MUC19* VNTR region, for each population in the 1KG. The mean number of repeat copies stratified by population is denoted by a gray diamond and the average number of repeat copies among individuals who carry exactly zero, one, and two introgressed tracts are denoted by blue, orange, and green circles, respectively, and are stratified by population. The black dashed line denotes the outlier threshold, which corresponds to the 95th percentile of the 1KG repeat copies distribution. Repeat copies appeared similar to the reference human genome (287.5 copies) in the Altai Denisovan (296 copies) and Altai (379 copies), Vindija (268 copies), and Chagyrskaya (293 copies) Neanderthal genomes. (**B**) The relationship between the average number of repeat copies between a MXL individual's two chromosomes and the number of introgressed tracts overlapping the *MUC19* VNTR region. Note that there is a significant positive correlation between the number of repeat copies and the number of introgressed tracts present in an MXL individual (Spearman's ρ: 0.885; *P*-value: 2.839e-22). (**C**) Relationship between the average number of repeat copies between a MXL individual's two chromosomes and the proportion of Indigenous American ancestry at the *MUC19* VNTR region. Note that there is a significant positive correlation between the number of repeat copies and the proportion of Indigenous American ancestry in an MXL individual (Spearman's ρ: 0.438; *P*-value: 2.940e-4).

introgressed tract that overlaps with the VNTR region (Fig. 2). MXL has more individuals exhibiting an elevated copy number (>487 repeats) than any other 1KG population, and there is a positive correlation between the number of repeats and the number of introgressed tracts that overlap the VNTR region present in a MXL individual (Spearman's ρ: 0.885; *P*-value: 2.839e-22; Fig. 2B, fig. S14, and table S22). We find that among MXL individuals, the number of repeats and the Indigenous American ancestry proportion at the repeat region is significantly positively correlated (Spearman's ρ: 0.483; *P*-value: 2.940e-4; Fig. 2C, fig. S15, and tables S23 and S24), whereas the African (Spearman's ρ: -0.289; *P*-value: 2.072e-2; fig. S15, table S23 and S24) and European (Spearman's ρ: -0.353; *P*-value: 4.191e-3; fig. S15 and tables S23 and S24) ancestry proportions have a significant negative correlation. Taken together, in MXL we find that an individual's VNTR copy number is highly predicted by the number of introgressed tracts that overlap the VNTR region. To a lesser extent, the VNTR copy number is also predicted by the Indigenous American ancestry proportion in the repeat region, indicating that individuals with elevated VNTR copy number have higher proportions of Indigenous American ancestry and harbor the introgressed haplotype. Individuals who carry an elevated number of the *MUC19* VNTR are likely to also carry the archaic haplotype, especially in admixed American populations where the archaic haplotype of *MUC19* is found at highest frequencies (Mann-Whitney *U* test, *P*-value: 1.597e-87; fig. S13, Fig. 2, and tables S18 to S21).

Given the difficulties of calling the number of repeats from short-read data, we examined long-read sequence data from the Human Pangenome Reference Consortium (HPRC) and Human Genome Structural Variant Consortium (HGSVC) (*31*). These corroborated our findings (figs. S10 and S16), revealing an extra 424 copies of the 30-bp *MUC19* tandem repeat exclusively in admixed American samples, arranged in four additional segments of 106 repeats (at 3171 bp each). This structural variant is exceptionally large; it effectively doubles the size of the ~12-kb coding exon that harbors the tandem repeat (fig. S10).

Introgression introduced missense variants at MUC19

Inspecting the 135 Denisovan-specific SNPs and four Neanderthal-specific SNPs in the core 72-kb region reveals that some modern humans carry two Denisovan-specific synonymous mutations and nine Denisovan-specific nonsynonymous mutations (table S25). We quantified the allele frequencies for these nine Denisovan-specific missense variants in present-day populations and in 23 ancient Indigenous American genomes that pre-date European colonization and the African slave trade (Fig. 3A and tables S26 to S33). In the admixed American superpopulation, we find that the Denisovan-specific missense mutations are segregating at the highest frequencies (frequency range in AMR: ~0.154 to ~0.157) compared with all other 1KG superpopulations (frequency range in non-AMR: ~0 to ~0.108; tables S27

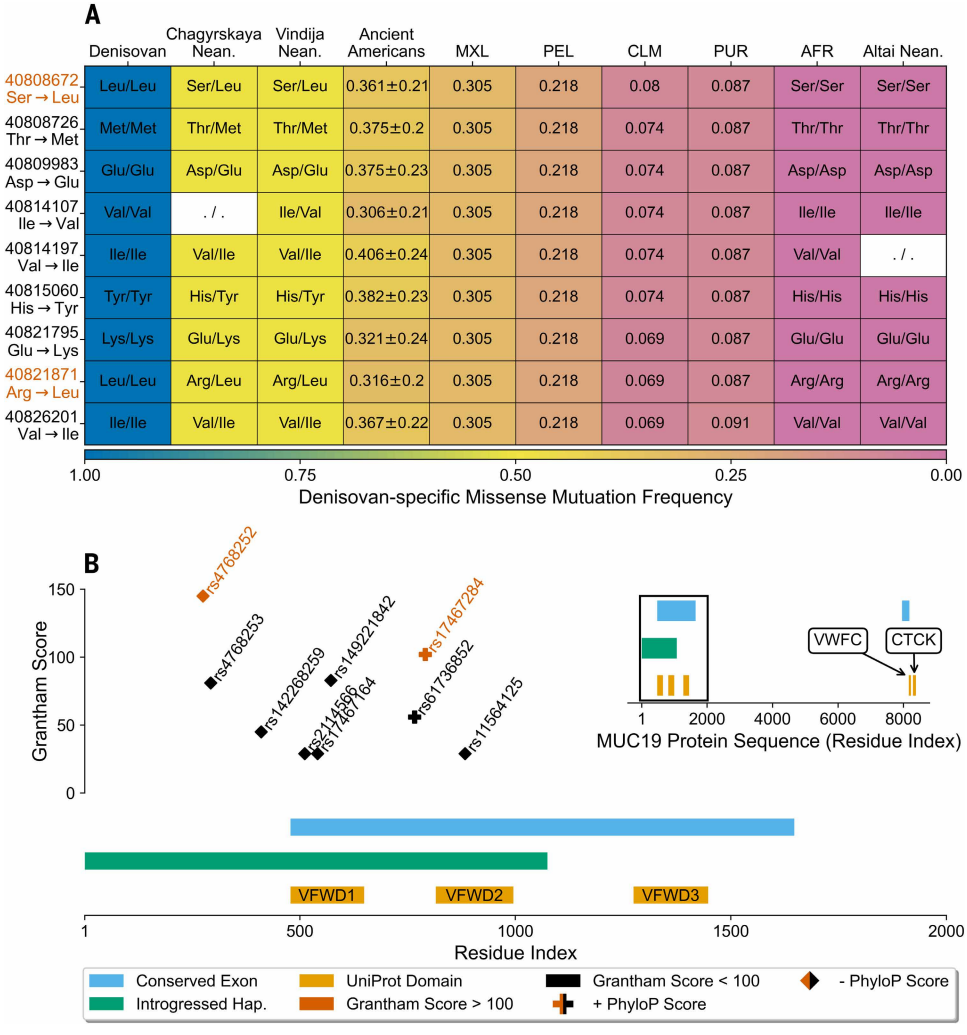


Fig. 3. Frequency and protein sequence context of the nine Denisovan-specific missense mutations at the 72-kb region in *MUC19*. (**A**) Heatmap depicting the frequency of Denisovan-specific missense mutations (columns) among the four archaic individuals ($n = 2$, per archaic individual), 23 ancient pre-European colonization American individuals ($n = 46$), the entire African superpopulation in the 1KG (AFR; $n = 1008$), and admixed American populations in the 1KG; Mexico (MXL; $n = 128$), Peru (PEL; $n = 170$), Colombia (CLM; $n = 188$), and Puerto Rico (PUR; $n = 208$), where n represents the number of chromosomes in each population. The lefthand side of each row denotes one of the nine Denisovan-specific missense mutations with the position and amino acid substitution (hg19 reference amino acid → Denisovan-specific amino acid). The text in each cell represents the Denisovan-specific missense mutation frequency, and for the ancient Americans, we also report the 95% confidence interval. For the archaic individuals, each cell is denoted with the individual's amino acid genotype and each AFR cell is denoted by the homozygous hg19 reference amino acid genotype. (**B**) Denisovan-specific missense mutations in the context of the MUC19 protein sequence are shown. The first 2000 residues are depicted as the main plot and the full protein sequence is displayed in the smaller subplot. Conserved exons are blue and the UniProt domains are orange, and the text corresponds to specific UniProt domain identity: Von Willebrand factor (VFW) D domains, VWFC domain, and C-terminal cystine knot-like (CTCK) domains. Each of the nine Denisovan-specific missense mutations are denoted by their rsID, plotted with respect to residue index on the *x*-axis and their corresponding Grantham score on the *y*-axis. The color of each Denisovan-specific missense mutation denotes whether the mutation has a Grantham score less than 100 (black) or a Grantham score greater than 100 (red, with the marker denoting whether their respective exon has a negative PhyloP score (diamonds) or a positive PhyloP score (crosses).

and S28). When we stratify by population instead of by superpopulation, we find the Denisovan-specific missense mutations are segregating at frequencies between ~0.069 and ~0.305 among admixed American populations, at varying frequencies between ~0.005 and ~0.157 throughout European, East Asian, and South Asian populations, and at the highest frequency in MXL where all nine Denisovan-specific missense mutations are segregating at a frequency of ~0.305 (Fig. 3A and table S29). We find the mean Denisovan-specific missense mutation frequency to be positively correlated with the introgressed tract frequency per population (Pearson's ρ: 0.976; *P*-value: 5.306e-16; fig. S17).

We then evaluate the frequency of the nine Denisovan-specific missense mutations in 23 ancient pre-European colonization American individuals, and find that each of the nine Denisovan-specific missense mutations are segregating at higher frequencies than in any admixed American population in the 1KG, but at statistically similar frequencies with respect to MXL [see (*28*); Fig. 3A and tables S29 to S32]. These ancient individuals were sampled from a wide geographic and temporal range [fig. S18 and table S26; (*28*)] and do not comprise a single population, yet we detect the presence of the Denisovan-specific missense mutations in sampled individuals from AK, MT, CA, ON, Central Mexico, Peru, and Patagonia (table S30). When we quantify the frequency of these mutations in 22 unadmixed Indigenous Americans from the Simons Genome Diversity Project (SGDP), we find that all nine Denisovan-specific missense variants are segregating at a frequency of ~0.364, which is statistically similar to the ancient American frequencies [see (*28*); tables S31 and S32] and higher than any admixed American population in the 1KG, albeit at statistically similar frequencies with respect to MXL (tables S31 and S32). Given that all nine of the missense mutations are found within a ~17.5-kb region, we quantified the frequency of the Denisovan-specific missense mutation at position Chr12:40808726 in both the ancient individuals and admixed Americans in the 1KG, as this position has genotype information in 20 out of the 23 ancient American individuals (table S30). We then assessed the relationship between Indigenous American ancestry proportion at the 72-kb region, and this Denisovan-specific missense mutation frequency. We find a positive and significant relationship (Pearson's ρ: 0.489; *P*-value: 1.982e-23; fig. S19) between an individual's Indigenous American Ancestry proportion and their respective Denisovan-specific missense mutation frequency, which suggests that recent admixture in the Americas may have diluted the introgressed ancestry at the 72-kb region. We also quantify the frequency of these variants in 44 African individuals from the SGDP and find all nine Denisovan-specific missense variants at a frequency of ~0.011, in a single chromosome from a Khomani San individual (table S33).

To estimate the potential effect of these missense mutations on the MUC19 protein, we relied on Grantham scores (*32*). One of the Denisovan-specific missense mutations found at position Chr12:40821871 (rs17467284 in Fig. 3B) results in an amino acid change with a Grantham score of 102. This substitution is classified as moderately radical (*33*) and suggests that the amino acid introduced through introgression is likely to affect the translated protein's structure or function. This Denisovan-specific missense mutation falls within an exon that is highly conserved across vertebrates (PhyloP score: 5.15, *P*-value: 7.08e-6; Fig. 3B) (*34*), indicating that this amino acid residue is likely functionally important, and that the amino acid change introduced by the Denisovan-specific missense mutation may have a substantial structural or functional impact. Furthermore, this missense mutation falls between two Von Willebrand factor D domains, which play an important role in the formation of mucin polymers and gel-like matrices (*35*). Our results suggest that this Denisovan-specific missense mutation is a potential candidate for affecting its translated protein and may affect the polymerization properties of *MUC19* and the viscosity of the mucin matrix.

Identification of the most likely donor of the introgressed haplotype at MUC19

To identify the most likely archaic donor, we investigated the patterns of haplotype divergence at *MUC19* by comparing the modern human haplotypes in the 1KG at the 72-kb region (see Supplementary Methods and shaded region in Fig. 1A) with the high-coverage archaic humans. We calculated the sequence divergence—the number of pairwise differences normalized by the effective sequence length—between all haplotypes in the 1KG and the genotypes for the Altai Denisovan and the three high-coverage Neanderthal individuals (figs. S20 to S22 and tables S34 and S35). Haplotypes from the Americas exhibit a bimodal distribution of sequence divergence for affinities to the Altai Denisovan, which we do not observe for the African haplotypes (Fig. 4A), as expected for an introgressed region. When comparing with all four high-coverage archaic genomes at the 72-kb region (Fig. 4B), there is a clear pattern of sequence divergence for the introgressed haplotypes found in the admixed American superpopulation of the 1KG (AMR). Although, Fig. 4B shows that African haplotypes are closer in sequence divergence to the Altai Neanderthal than to the Altai Denisovan, these values are not statistically significant [see Dataset 1 in (*28*, *36*)]. The Altai Neanderthal itself is significantly more distant than expected from the Altai Denisovan (sequence divergence: 0.003782, *P*-value: 0.002, fig. S23 and table S36), and this larger than expected divergence explains why African haplotypes appear closer to the Altai Neanderthal in this region. We corroborate the pattern observed in Fig. 4 using principal component analysis to visualize the haplotype structure in this region (fig. S24).

Despite our $U_{AFR,MXL,Denisovan}(1\%, 30\%, 100\%)$ and archaic SNP density results demonstrating that the introgressed haplotype at the 72-kb region shares the most alleles with the Altai Denisovan (Fig. 4B), we find that this region is not statistically significantly closer to the Altai Denisovan individual than expected from the genomic background of sequence divergence (sequence divergence: 0.00097, *P*-value: 0.237, fig. S25, table S37). However, this is not unusual, given that the Altai Denisovan is not genetically closely related to Denisovan introgressed segments in modern humans [see Section S5 in (*28*)], which might suggest that the Denisovan donor population of the 72-kb region in *MUC19* is not closely related to the Altai Denisovan individual. Furthermore, the 72-kb region is also not statistically significantly closer to Neanderthals than expected from the genomic background of sequence divergence (sequence divergence from the Altai Neanderthal: 0.003648, *P*-value: 0.995; Chagyrskaya Neanderthal: 0.001818, *P*-value: 0.811; Vindija Neanderthal: 0.001816, *P*-value: 0.806; fig. S25 and table S37).

As an additional approach, we used the *D*+ statistic to assess which archaic human exhibits the most allele sharing with the introgressed haplotype at the 72-kb region in *MUC19* (*37*, *38*). We performed *D*+ (P1, P2, P3, outgroup) tests with the following configurations: the Yoruban population (YRI) as P1, the focal MXL individual (NA19664) with two copies of the introgressed haplotype with an affinity to the Altai Denisovan as P2, and one of the four high-coverage archaic genomes as P3. We use the Enredo-Pecan-Ortheus ancestral allele calls from the six primate alignments as the outgroup. We exclusively observe a positive and significant *D*+ value (*D*+: 0.743, *P*-value: 1.386e-5; fig. S26 and table S38) when the Altai Denisovan is used as P3 (the putative donor population). Conversely, when any of the three Neanderthals are used as P3, we observe non-significant *D*+ values (P3: Altai Neanderthal, *D*+: -0.622, *P*-value: 0.999; P3: Chagyrskaya Neanderthal, *D*+: 0.175, *P* -value: 0.183; P3: Vindija Neanderthal, *D*+: 0.182, *P*-value: 0.174; fig. S26 and table S38). These *D*+ results suggest that the introgressed haplotype at the 72-kb *MUC19* region shares more alleles with the Altai Denisovan, which is not observed with any of the three Neanderthals and provides evidence that the introgressed haplotype found in modern humans is Denisovan-like.

When we consider the 742-kb region in MXL, we find that it is closest to the Chagyrskaya and Vindija Neanderthals, and significantly

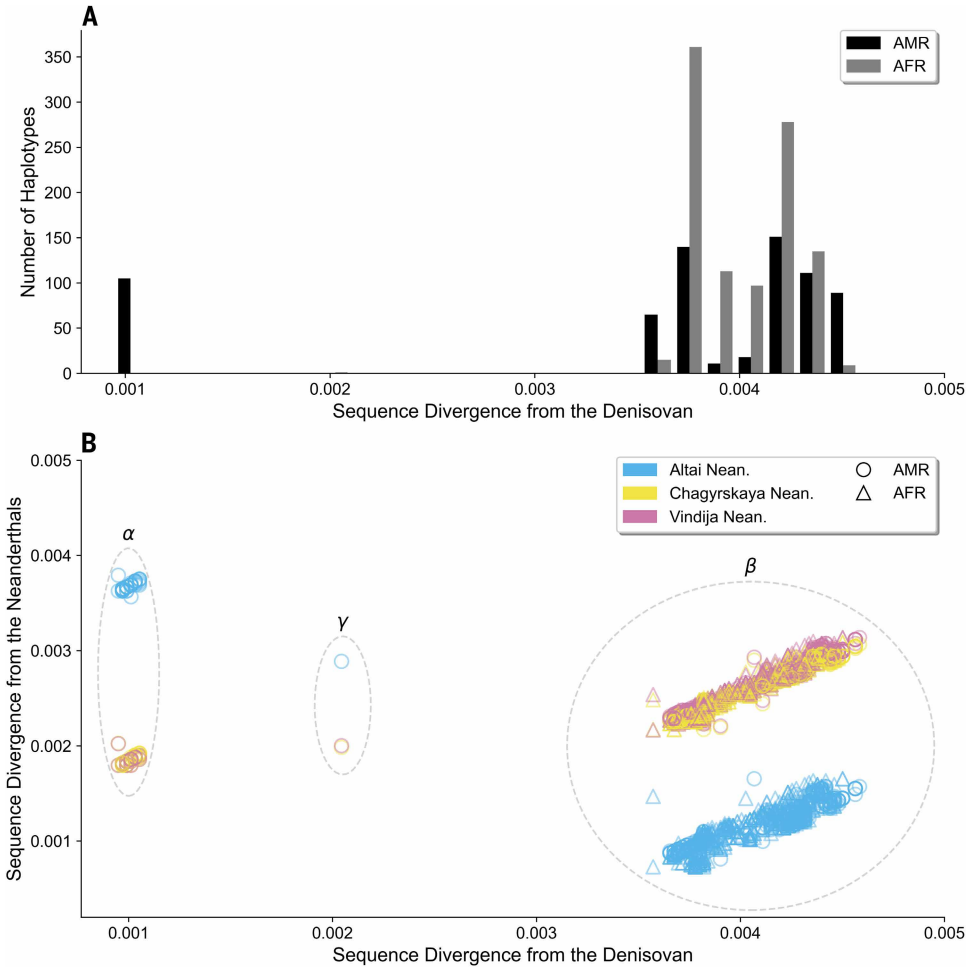


Fig. 4. Haplotype divergence at the 72-kb region in *MUC19*. (**A**) Distribution of haplotype divergence—number of pairwise differences between a modern human haplotype and an archaic genotype normalized by the effective sequence length—with respect to the Altai Denisovan for all individuals in the Admixed American (AMR, black bars) and African (AFR, gray bars) superpopulations. (**B**) Joint distribution of haplotype divergence from the Altai Denisovan (*x*-axis) and the Neanderthals (*y*-axis)—Altai Neanderthal in blue, Chagyrskaya Neanderthal in yellow, and Vindija Neanderthal in pink—for all individuals in the AMR (circles) and AFR (triangles) superpopulations. The three gray ellipses (α, β, and γ) represent the three distinct haplotype groups segregating in the 1KG. The α ellipse represents the introgressed haplotypes that exhibit a low sequence divergence from the Altai Denisovan, a high sequence divergence from the Altai Neanderthal, and an intermediate sequence divergence—higher compared with the Altai Denisovan but lower compared to the Altai Neanderthal—with respect to the Chagyrskaya and Vindija Neanderthals. The β ellipse represents the non-introgresssed haplotypes which exhibit a high sequence divergence from the Altai Denisovan, a low sequence divergence from the Altai Neanderthal, and an intermediate sequence divergence—lower compared with the Altai Denisovan but higher compared with the Altai Neanderthal—with respect to the Chagyrskaya and Vindija Neanderthals. Note that the AMR haplotype within the γ ellipse is positioned at intermediate sequence divergence values with respect to the α and β ellipses, which represents one of seven recombinant haplotypes segregating in the 1KG [see fig. S44 in (*28*)].

closer than expected from the genomic background [sequence divergence from the Chagyrskaya Neanderthal: 0.000661, *P*-value: 0.006; from the Vindija Neanderthal: 0.000656, *P*-value: 0.007; figs. S27 to S30 and tables S39 to S41; Dataset 2 (*28*, *36*)]. We also tested whether this region is statistically significantly closer to the Altai Denisovan than expected from the genomic background and found that this tract in MXL is also significantly closer than expected to the Altai Denisovan, albeit not as close when compared with the Chagyrskaya and Vindija Neanderthals (sequence divergence from the Altai Denisovan: 0.000806, *P*-value: 0.019; figs. S27 to S30 and tables S39 to S41). We then performed *D*+ analyses for the 742-kb region with identical configurations as for the 72-kb region, and observed positive and significant *D*+ values when P3 was Chagyrskaya (*D*+: 0.381, *P*-value: 7.375e-6; fig. S31 and table S42) and Vindija Neanderthals (*D*+: 0.383, *P*-value: 7.505e-6; fig. S31 and table S42), but not when the Altai Neanderthal was P3 (*D*+: 0.091, *P*-value: 1.442e-1; fig. S31 and table S42). *D*+ was, however, significant when the Altai Denisovan was P3 (*D*+: 0.377, *P*-value: 9.889e-8; fig. S31 and table S42). These *D*+ results are consistent with our sequence divergence results, which indicate that the introgressed haplotype at the 742-kb *MUC19* region has a high affinity for the Altai Denisovan and the two late Neanderthals, but not the Altai Neanderthal (figs. S20 to S31 and tables S34 to S42).

Given the high density of Denisovan-specific alleles (fig. S2 and table S4), the sequence divergence, and *D*+ results for the 72-kb and 742-kb regions, the most parsimonious explanation is that a Denisovan population could have introduced this haplotype into non-Africans. However, our 742-kb results also suggest that a Neanderthal population could have introduced the introgressed haplotype. This is further supported by the sequence divergence results at the 72-kb region, in which late Neanderthals exhibit intermediate distance to the introgressed haplotype (Fig. 4B), suggesting they harbor some of the Denisovan alleles.

Neanderthals introduce Denisovan-like introgression into non-African modern humans

Based on sequence divergence, the Chagyrskaya and Vindija Neanderthals carry a 742-kb haplotype that is most similar to the Altai Neanderthal, with the exception of the 72-kb region. To understand why the Chagyrskaya and Vindija Neanderthals exhibit intermediate levels of sequence divergence with the introgressed haplotype present in MXL at the 72-kb region in *MUC19* relative to the Altai Denisovan and Altai Neanderthal (see the α ellipse in Fig. 4B), we computed the number of heterozygous sites for each archaic human. Because the Chagyrskaya and Vindija Neanderthals present intermediate sequence divergences, we expected these two individuals to have more heterozygosity than the Altai Neanderthal. At the 72-kb region in *MUC19*, we observed that the Chagyrskaya and Vindija Neanderthals carry an elevated number of heterozygous sites (168 Chagyrskaya heterozygous sites, *P*-value: 2.307e-4; 171 Vindija heterozygous sites, *P*-value: 3.282e-4; Fig. 5A, fig. S32 and table S43) that is higher than those of the Altai Neanderthal (one heterozygous site, *P*-value: 0.679; Fig. 5A, fig. S32, and table S43) and the Altai Denisovan (six heterozygous sites, *P*-value: 0.455; Fig. 5A, fig. S32, and table S43). The Chagyrskaya and Vindija Neanderthals carry a higher number of heterozygous sites than all African individuals (~75, *P*-value: 0.424; Fig. 5A, fig. S33, and table S44) and have a more similar pattern

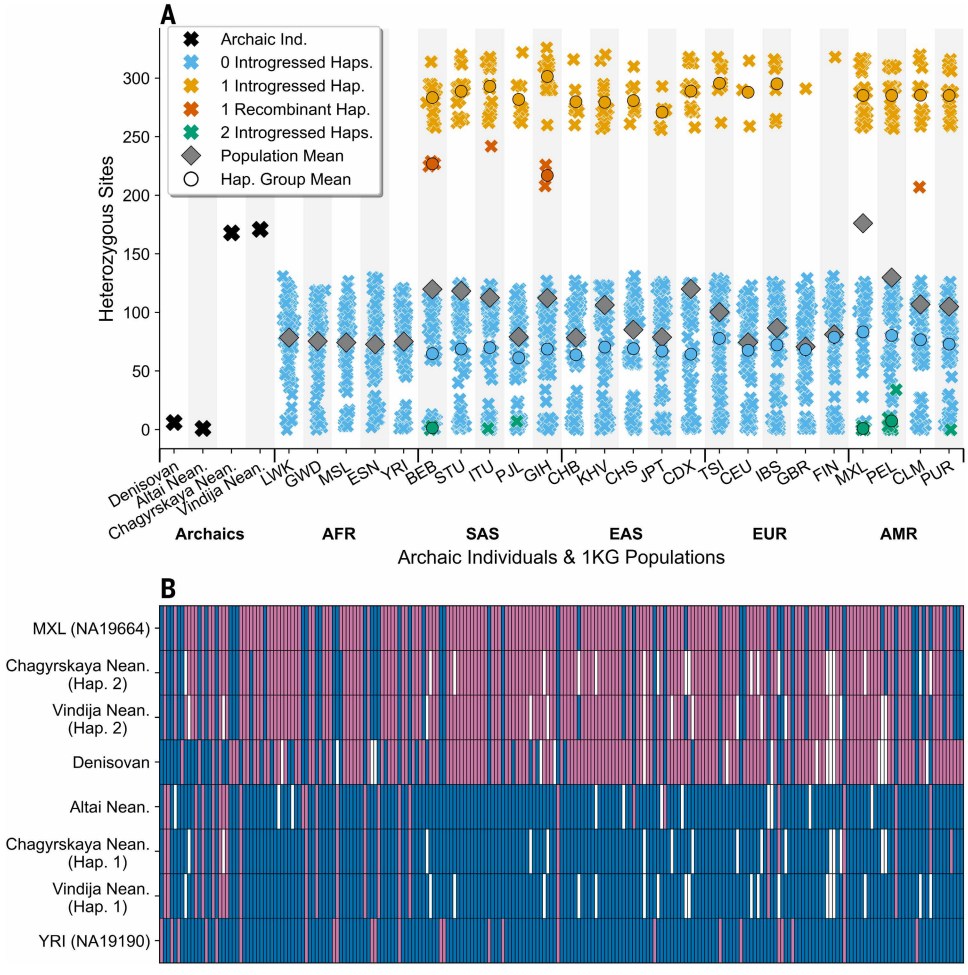


Fig. 5. The high levels of heterozygosity in the Chagyrskaya and Vindija Neanderthals are explained by Denisovan-like ancestry at the 72-kb region in *MUC19*. (**A**) Number of heterozygous sites at the 72-kb region in *MUC19* per archaic individual (black X's), 1KG individuals without the introgressed haplotype (blue X's), 1KG individuals with exactly one copy of the introgressed haplotype (yellow X's), 1KG individuals with a recombinant introgressed haplotype (red X's), and 1KG individuals with two copies of the introgressed haplotype (green X's). The average number of heterozygous sites stratified by population are denoted by the gray diamonds and the average number of heterozygous sites among individuals who carry exactly zero, one, and two introgressed haplotypes are denoted by blue, yellow, and green circles, respectively, and are stratified by population. (**B**) Haplotype matrix of the 233 segregating sites (columns) among the focal MXL individual (NA19664) with two copies of the introgressed haplotype, the focal YRI individual (NA19190) without the introgressed haplotype, the Altai Denisovan, the Altai Neanderthal, and the two phased haplotypes for the Chagyrskaya and Vindija Neanderthals, respectively. Cells shaded blue denote the hg19 reference allele, cells shaded pink denote the alternative allele, and cells shaded white represent sites that did not pass quality control in the given archaic individual. Note that the focal MXL and YRI individuals are homozygous for every position in the 72-kb region in *MUC19* and that the heterozygous sites for the Altai Denisovan and Altai Neanderthal—six and one heterozygous sites, respectively—are omitted.

to non-African individuals carrying exactly one Denisovan-like haplotype (~287, *P*-value: 3.157e-4; yellow X's in Fig. 5A, fig. S33, and table S44). This observation runs opposite to the genome-wide expectation for Neanderthals, as archaic humans have much lower heterozygosity than modern humans (genome-wide heterozygosity is ~0.00014 to ~0.00017 for the Neanderthals, ~0.00019 for the Denisovan and ~0.001 for African modern humans; fig. S34 and table S45).

Within modern humans, we find that individuals carrying exactly one Denisovan-like haplotype at the 72-kb region harbor significantly more heterozygous sites at *MUC19* compared with the rest of their genome (average number of heterozygous sites: ~287, *P*-value: 3.157e-4; fig. S33 and table S44), which surpasses the number of heterozygous sites at *MUC19* of any African individual (Fig. 5A). Individuals carrying two Denisovan-like haplotypes harbor significantly fewer heterozygous sites than expected at *MUC19* relative to the rest of their genome (average number of heterozygous sites: ~4, *P*-value: 6.945e-4; fig. S33 and table S44), whereas African individuals harbor the expected number of heterozygous sites (average number of heterozygous sites: ~75, *P*-value: 0.424; fig. S33 and table S44). Given that the Chagyrskaya and Vindija Neanderthals and non-African individuals who harbor one copy of the Denisovan-like haplotype exhibit an excess of heterozygous sites at the 72-kb region, we hypothesized that the Chagyrskaya and Vindija Neanderthals also harbor one Denisovan-like haplotype. This arrangement would explain the elevated number of heterozygous sites and the intermediary sequence divergences with respect to the introgressed haplotype.

To test this hypothesis, we first performed additional tests for gene flow between the archaic individuals using the *D*+ statistic within the 72-kb *MUC19* region that provided evidence that the Chagyrskaya and Vindija Neanderthals harbor one copy of the Denisovan-like haplotype. For these comparisons, the Altai Neanderthal is P1, either the Chagyrskaya or Vindija Neanderthals are P2, and the Altai Denisovan is P3, and we observe significant and positive *D*+ values supporting gene flow between the Denisovan and the Chagyrskaya (*D*+: 0.783; *P*-value: 0.029) and Vindija (*D*+: 0.819; *P*-value: 0.018) Neanderthals (fig. S35 and table S46). To further investigate whether the Chagyrskaya and Vindija Neanderthals harbor one Denisovan-like haplotype in the 72-kb region, we used BEAGLE to phase the 72-kb region. As no phasing has been done for archaic humans, we tested the reliability of using the 1KG as a reference panel by constructing a synthetic 72-kb region. We sampled one allele from the Altai Neanderthal and one allele from the Altai Denisovan at heterozygous sites in either the Chagyrskaya or Vindija Neanderthals. We found that we could phase the synthetic individual perfectly at this region [see supplementary sections S3 and S4 in (*28*)]. Encouraged by these results, we phased the Chagyrskaya and Vindija Neanderthals at the 72-kb region and confirmed that they carry one haplotype similar to the Altai Neanderthal and one haplotype similar to the Denisovan-like haplotype in MXL. Relative to the Altai Neanderthal, the Chagyrskaya Neanderthal-like haplotype exhibits 3.5 differences and the Vindija exhibits 4 differences (Fig. 5B and table S47). Relative to the Altai Denisovan, the Chagyrskaya Denisovan-like haplotype exhibits 43 differences and the Vindija haplotype exhibits 41 differences (Fig. 5B and table S47). As expected, the phased Denisovan-like haplotype in these two Neanderthals is closest to the Denisovan-like haplotype in MXL;

the Chagyrskaya exhibits five differences and the Vindija Neanderthal exhibits four differences (Fig. 5B and table S48). We show that in the 72-kb region, the introgressed haplotype in MXL is statistically significantly closer to the phased Denisovan-like haplotype present in Chagyrskaya and Vindija Neanderthals [sequence divergence from Chagyrskaya Neanderthal haplotype: 0.000104, *P*-value: 0.003; sequence divergence from Vindija Neanderthal haplotype: 0.000083, *P*-value: 0.002; fig. S36 and table S48; Dataset 3 in (*28*, *36*)]. Due to the potential introduction of biases when phasing ancient DNA data, to investigate whether the Chagyrskaya and Vindija Neanderthals carry a Denisovan-like haplotype we developed an approach called Pseudo-Ancestry Painting (*PAP*) (*28*) to assign the two alleles at a heterozygous site to two source individuals. We found that using MXL (NA19664) and YRI (NA19190) individuals as sources maximizes the number of heterozygous sites in the Chagyrskaya (*PAP* score: 0.94, *P*-value: 3.683e-4) and Vindija (*PAP* score: 0.929, *P*-value: 8.679e-5) Neanderthals (fig. S37and table S49).

In sum, our analyses suggest that some non-Africans carry a mosaic region of archaic ancestry: a small Denisovan-like haplotype (72-kb) embedded in a larger Neanderthal haplotype (742-kb), that was inherited through Neanderthals who themselves acquired Denisovan ancestry from an earlier introgression event (fig. S38). This is consistent with the literature, where Denisovan introgression into Neanderthals is rather common (*39*, *40*). Thus, we refer to the mosaic haplotype found in modern humans as the archaic haplotype.

Discussion

The study of adaptive archaic introgression has illuminated candidate genomic regions that affect the health and overall fitness of global populations. In this study, we pinpointed several aspects of the gene *MUC19* that highlight its importance as a candidate to study adaptive introgression; one of the haplotypes that span this gene in modern humans is of archaic origin, as modern humans inherited this haplotype from Neanderthals who in turn inherited it from Denisovans, and the haplotype introduced nine missense mutations that are at high frequency in both Indigenous and admixed American populations. Individuals with the archaic haplotype carry a massive coding VNTR expansion relative to the nonarchaic haplotype and their functional differences may help explain how mainland Indigenous Americans adapted to their environments, a concept that remains underexplored. This study adds an example to the growing literature of natural selection acting on archaic alleles at coding sites, or possibly an example of natural selection acting on human VNTRs, a developing research frontier (*41*).

A larger implication of our findings is that archaic ancestry could have been a useful source of standing genetic variation as the early Indigenous American populations adapted to new environments, with genes like *MUC19* and other mucins possibly mediating important fitness effects (*42*). The variation in the *MUC19* coding VNTR in global populations dovetails with this idea and adds to a growing body of evidence for the important role of structural variants in human genomics and evolution (*31*, *43*). In admixed American populations, particular haplotypes carrying the most extreme copy numbers were selected and are now relatively frequent. This VNTR expansion effectively doubles the functional domain of this mucin, indicating an adaptive role driven by environmental pressures particular to the Americas. However, we cannot know whether it is the nonsynonymous variants or the VNTR driving natural selection as they are linked in haplotypes, and our evidence for positive selection is tied to SNP variation and not to the VNTR itself.

Another interesting aspect of *MUC19* is the evolutionary history of the introgressed region. Our observation of a 72-kb Denisovan-like haplotype found in Neanderthals and non-African modern humans nested within a larger Neanderthal haplotype suggests that the smaller Denisovan haplotype was first introgressed into Neanderthals, who later admixed with modern humans to introduce the full 742-kb haplotype. Although the Altai Neanderthal does not harbor the Denisovan-like haplotype at the 72-kb region, the other two chronologically younger Neanderthals (Chagyrskaya and Vindija) do. We phased these younger Neanderthals [see Supplementary Sections S3 to S5 in (*28*)] and showed that they harbor exactly one Denisovan-like haplotype, which explains why they exhibit an excess of heterozygosity. The Denisovan-like haplotype in the younger Neanderthals is also statistically significantly closer to the archaic haplotype present in MXL (fig. S36 and table S48), providing additional evidence that modern humans obtained this haplotype through an interbreeding event with Neanderthals. Despite the introgressed archaic haplotype having an excessive amount of shared alleles with the Altai Denisovan at the 72-kb region, the Altai Denisovan harbors several private mutations—14 and 6 mutations in the homozygous and heterozygous state, respectively—that are absent across all 287 Denisovan-like haplotypes in the 1KG, suggesting that the introgressing Denisovan population may not be closely related to Altai Denisovan [see supplementary section S5 in (*28*)]. Indeed, the introgressed haplotype in the 72-kb region is present at low frequencies in other non-African populations including Papuans, where the genome-wide Denisovan ancestry of Papuans has been estimated to originate from a population of Denisovans that was not closely related to the Altai Denisovan (*44*). Finding two highly divergent haplotypes maintained in polymorphism in two Neanderthal individuals and finding the archaic haplotype at high frequencies in admixed American populations—but not at fixation—may point to a balanced polymorphism (*45*). More generally, the evolutionary history of this region suggests a complex history that involves recurrent introgression and natural selection, and it parallels complex introgression patterns from other regions of the genome (*46*–*48*).

Finally, we find a single San individual who carries the nine Denisovan missense variants in heterozygous form, uniquely among all African individuals considered here. The sequence divergence between this San haplotype and the archaic MXL haplotype at the 72-kb region is high (0.001342), further supporting the origin of the archaic haplotype in non-Africans as introgressed. Khoe-San populations are estimated to have diverged from other African groups 120,000 years ago (*49*). Finding a divergent haplotype in the San is consistent with a previous study (*50*), as ~1% of their ancestry can be attributed to lineages that diverged from the main human lineage more than 1 million years ago. We note that this San individual does not harbor an elevated number of copies of the VNTR (301 copies), which further supports the importance of the VNTR expansion in the Americas. Furthermore, we cannot determine whether this variant found its way into the San through modern admixture of non-African ancestry into Sub-Saharan populations.

Perhaps the largest knowledge gap regarding why the archaic haplotype of *MUC19* would be under positive selection is its underlying function. Mucins are secreted glycoproteins responsible for the gel-like properties and viscosity of mucus (*51*), and are characterized by proline, threonine, and serine (PTS) tandem repeats, which in *MUC19* are structured into 30-bp tandem repeats. The massive difference in copy numbers of the 30-bp PTS tandem repeat domains carried by individuals harboring the Human-like and archaic haplotypes strongly suggests that *MUC19* variants differ in function as a consequence of different molecular binding affinities between variants. This is the case in other mucins, such as *MUC7*, in which variants carrying different numbers of PTS repeats exhibit different microbe-binding properties (*42*). If the two variants of *MUC19* also have differential binding properties, this would lend support to why positive selection would increase the frequency of the archaic haplotype in Indigenous American populations. However, there is limited medical literature associating variation in *MUC19* with human fitness. Further experimental validation of how VNTRs and the Denisovan-specific missense mutations affect MUC19 function is necessary to understand the effects the archaic haplotype

may exert on the translated *MUC19* protein, and how it modifies its function during the formation of mucin polymers.

Methods developed in evolutionary biology can be useful for identifying candidate variants underlying biological functions. Future functional and evolutionary studies of the *MUC19* region will not only provide insight into specific mechanisms of how variation at this gene confers a selective advantage, but also specific evolutionary events that occurred in the history of humans. Beyond improving our understanding of how archaic variants facilitated adaptation in novel environments, our findings also highlight the importance of studying archaic introgression in understudied populations, such as admixed populations from the Americas (*52*). Genetic variation in admixed American populations is less well-characterized than other global populations, as it is difficult to deconvolve Indigenous ancestries from European, African, and to a lesser extent, South Asian ancestries, following 500 years of European colonization (*29*). This knowledge gap is exacerbated by the high cost of performing genomic studies, building infrastructure, and generating scientific capacity in Latin America, but it is a worthwhile investment as our study shows that leveraging these populations can lead to the identification of candidate loci that can expand our understanding of adaptation from archaic standing variation.

REFERENCES AND NOTES

1. K. D. Ahlquist *et al.*, Our tangled family tree: New genomic methods offer insight into the legacy of archaic admixture. *Genome Biol. Evol.* **13**, evab115 (2021). doi: 10.1093/gbe/evab115; pmid: 34028527
2. R. Sharon, B. L. Browning, Y. Zhou, S. Tucci, J. M. Akey, Analysis of human sequence data reveals two pulses of archaic Denisovan admixture. *Cell* **173**, 53–61 (2018). doi: 10.1016/j.cell.2018.02.031; pmid: 29551270
3. F. A. Villanea, J. G. Schraiber, Multiple episodes of interbreeding between Neanderthal and modern humans. *Nat. Ecol. Evol.* **3**, 39–44 (2019). doi: 10.1038/s41559-018-0735-8; pmid: 30478305
4. D. Enard, D. A. Petrov, Evidence that RNA viruses drove adaptive introgression between Neanderthals and modern humans. *Cell* **175**, 360–371 (2018). doi: 10.1016/j.cell.2018.08.034; pmid: 30290142
5. S. Sankararaman, S. Mallick, N. Patterson, D. Reich, The combined landscape of Denisovan and Neanderthal ancestry in present day humans. *Curr. Biol.* **26**, 1241–1247 (2016). doi: 10.1016/j.cub.2016.03.037; pmid: 27032491
6. M. Petr, S. Pääbo, J. Kelso, B. Vernot, Limits of long-term selection against Neandertal introgression. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1639–1644 (2019). doi: 10.1073/pnas.1814338116; pmid: 30647110
7. X. Zhang, B. Kim, K. E. Lohmueller, E. Huerta-Sánchez, The impact of recessive deleterious variation on signals of adaptive introgression in human populations. *Genetics* **215**, 799–812 (2020). doi: 10.1534/genetics.120.303081; pmid: 32487519
8. F. Racimo, S. Sankararaman, R. Nielsen, E. Huerta-Sánchez, Evidence for archaic adaptive introgression in humans. *Nat. Rev. Genet.* **16**, 359–371 (2015). doi: 10.1038/nrg3936; pmid: 25963373
9. X. Zhang *et al.*, MaLAdapt reveals novel targets of adaptive introgression from Neanderthals and Denisovans in worldwide human populations. *Mol. Biol. Evol.* **40**, msad001 (2023). doi: 10.1093/molbev/msad001; pmid: 36617238
10. S. Fan, M. E. Hansen, Y. Lo, S. A. Tishkoff, Going global by adapting local: A review of recent human adaptation. *Science* **354**, 54–59 (2016). doi: 10.1126/science.aaf5098; pmid: 27846491
11. F. L. Mendez, J. C. Watkins, M. F. Hammer, A haplotype at *STAT2* introgressed from Neanderthals and serves as a candidate of positive selection in Papua New Guinea. *Am. J. Hum. Genet.* **91**, 265–274 (2012). doi: 10.1016/j.ajhg.2012.06.015; pmid: 22883142
12. S. Sankararaman *et al.*, The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014). doi: 10.1038/nature12961; pmid: 24476815
13. B. Vernot, J. M. Akey, Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**, 1017–1021 (2014). doi: 10.1126/science.1245938; pmid: 24476670
14. R. M. Gittelman *et al.*, Archaic hominin admixture facilitated adaptation to out-of-Africa environments. *Curr. Biol.* **26**, 3375–3382 (2016). doi: 10.1016/j.cub.2016.10.041; pmid: 27839976
15. A. J. Sams *et al.*, Adaptively introgressed Neandertal haplotype at the *OAS* locus functionally impacts innate immune responses in humans. *Genome Biol.* **17**, 1–15 (2016). doi: 10.1186/s13059-016-1098-6; pmid: 27899133
16. M. Dannemann, J. Kelso, The contribution of Neanderthals to phenotypic variation in modern humans. *Am. J. Hum. Genet.* **101**, 578–589 (2017). doi: 10.1016/j.ajhg.2017.09.010; pmid: 28985494
17. R. González-Buenfil *et al.*, Genetic signatures of positive selection in human populations adapted to high altitude in Papua New Guinea. *Genome Biol. Evol.* **16**, evae161 (2024). doi: 10.1093/gbe/evae161; pmid: 39173139
18. E. Huerta-Sánchez *et al.*, Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014). doi: 10.1038/nature13408;pmid: 25043035
19. F. Racimo *et al.*, Archaic adaptive introgression in *TBX15/WARS2*. *Mol. Biol. Evol.* **34**, 509–524 (2017). doi: 10.1093/molbev/msw283; pmid: 28007980
20. X. Zhang *et al.*, The history and evolution of the Denisovan-*EPAS1* haplotype in Tibetans. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020803118 (2021). doi: 10.1073/pnas.2020803118; pmid: 34050022
21. E. Tamm *et al.*, Beringian standstill and spread of Native American founders. *PLOS ONE* **2**, e829 (2007). doi: 10.1371/journal.pone.0000829; pmid: 17786201
22. H. E. Beck *et al.*, Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* **5**, 180214 (2018). doi: 10.1038/sdata.2018.214; pmid: 30375988
23. K. E. Witt, A. Funk, V. Añorve-Garibay, L. Lopez Fang, E. Huerta-Sánchez, The impact of modern admixture on archaic human ancestry in human populations. *Genome Biol. Evol.* **15**, evad066 (2023). doi: 10.1093/gbe/evad066; pmid: 36711776
24. F. Racimo, D. Marnetto, E. Huerta-Sánchez, Signatures of archaic adaptive introgression in present-day human populations. *Mol. Biol. Evol.* **34**, 296–317 (2017). doi: 10.1093/molbev/msw216; pmid: 27756828
25. A. W. Reynolds *et al.*, Comparing signals of natural selection between three Indigenous North American populations. *Proc. Natl. Acad. Sci.* **116**, 9312–9317, (2019). doi: 10.1093/gbe/evad066; pmid: 30988184
26. L. Skov *et al.*, Detecting archaic introgression using an unadmixed outgroup. *PLOS Genet.* **14**, e1007641 (2018). doi: 10.1371/journal.pgen.1007641; pmid: 30226838
27. A. R. Martin *et al.*, Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017). doi: 10.1016/j.ajhg.2017.03.004; pmid: 28366442
28. Materials and methods are available as supplementary materials.
29. S. G. Medina-Muñoz *et al.*, Demographic modeling of admixed Latin American populations from whole genomes. *Am. J. Hum. Genet.* **110**, 1804–1816 (2023). doi: 10.1371/journal.pgen.1007641; pmid: 37725976
30. Z. A. Szpiech, selscan 2.0: Scanning for sweeps in unphased data. *Bioinformatics* **40**, btae006 (2024). doi: 10.1093/bioinformatics/btae006; pmid: 38180866
31. P. Ebert *et al.*, Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* **372**, eabf7117 (2021). doi: 10.1126/science.abf7117; pmid: 33632895
32. R. Grantham. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864, 1974. doi: 10.1126/science.185.4154.862; pmid: 4843792
33. W. H. Li, C. I. Wu, C. C. Luo, A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**, 150–174 (1985). doi: 10.1093/oxfordjournals.molbev.a040343; pmid: 3916709
34. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010). doi: 10.1101/gr.097857.109; pmid: 19858363
35. G. Javitt *et al.*, Assembly mechanism of mucin and Von Willebrand factor polymers. *Cell* **183**, 717–729.e16 (2020). doi: 10.1016/j.cell.2020.09.021; pmid: 33031746
36. L. Ongaro, E. Huerta-Sánchez, A history of multiple Denisovan introgression events in modern humans. *Nat. Genet.* **56**, 2612–2622 (2024). doi: 10.1038/s41588-024-01960-y; pmid: 39501127
37. F. A. Fernando *et al.*,The MUC19 gene: An evolutionary history of recurrent introgression and natural selection, Version v4 (Zenodo, 2025); doi: 10.5281/zenodo.15042423
38. L. Lopez Fang, D. Peede, D. Ortega-Del Vecchyo, E. J. McTavish, E. Huerta-Sánchez, Leveraging shared ancestral variation to detect local introgression. *PLOS Genet.* **20**, e1010155 (2024). doi: 10.1371/journal.pgen.1010155; pmid: 38190420
39. D. Peede, D. O.-D. Vecchyo, E. Huerta-Sánchez, The utility of ancestral and derived allele sharing for genome-wide inferences of introgression. bioRxiv 2022.12.02.518851 [Preprint] (2022); doi: 10.1101/2022.12.02.518851
40. V. Slon *et al.*, The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature* **561**, 113–116 (2018). doi: 10.1038/s41586-018-0455-x; pmid: 30135579
41. B. M. Peter, 100,000 years of gene flow between Neandertals and Denisovans in the Altai mountains. bioRxiv 2020.03.13.990523 [Preprint] (2020); doi: 10.1101/2020.03.13.990523
42. E. G. Plender *et al.*, Structural and genetic diversity in the secreted mucins *MUC5AC* and *MUC5B*. *Am. J. Hum. Genet.* **111**, 1700–1716 (2024). doi: 10.1016/j.ajhg.2024.06.007; pmid: 38991590
43. D. Xu *et al.*, Archaic hominin introgression in Africa contributes to functional salivary *MUC7* genetic variation. *Mol. Biol. Evol.* **34**, 2704–2715 (2017). doi: 10.1093/molbev/msx206; pmid: 28957509
44. S. M. Yan *et al.*, Local adaptation and archaic introgression shape global diversity at human structural variant loci. *eLife* **10**, e67615 (2021). doi: 10.7554/eLife.67615; pmid: 34528508
45. L. H. Viscardi *et al.*, Searching for ancient balanced polymorphisms shared between Neanderthals and Modern Humans. *Genet. Mol. Biol.* **41**, 67–81 (2018). doi: 10.1590/1678-4685-gmb-2017-0308; pmid: 29658973
46. C. Posth *et al.*, Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. *Nat. Commun.* **8**, 16046 (2017). doi: 10.1038/ncomms16046; pmid: 28675384

47. M. Petr *et al.*, The evolutionary history of Neanderthal and Denisovan Y chromosomes. *Science* **369**, 1653–1656 (2020). doi: 10.1126/science.abb6460; pmid: 32973032
48. S. Peyrégne, J. Kelso, B. M. Peter, S. Pääbo, The evolutionary history of human spindle genes includes back-and-forth gene flow with Neandertals. *eLife* **11**, e75464 (2022). doi: 10.7554/eLife.75464; pmid: 35816093
49. A. P. Ragsdale *et al.*, A weakly structured stem for human origins in Africa. *Nature* **617**, 755–763 (2023). doi: 10.1038/s41586-023-06055-y; pmid: 37198480
50. K. Wang, I. Mathieson, J. O'Connell, S. Schiffels, Tracking human population structure through time from whole genome sequences. *PLOS Genet.* **16**, e1008552 (2020). doi: 10.1371/journal.pgen.1008552; pmid: 32150539
51. P. Pajic *et al.*, A mechanism of gene evolution generating mucin function. *Sci. Adv.* **8**, eabm8757 (2022). doi: 10.1126/sciadv.abm8757; pmid: 36026444
52. F. A. Villanea, K. E. Witt, Underrepresented populations at the archaic introgression frontier. *Front. Genet.* **13**, 821170 (2022). doi: 10.3389/fgene.2022.821170; pmid: 35281795
53. D. Peede, David-Peede/MUC19: Accepted MUC19 code, Version v1, Zenodo (2025); doi: 10.5281/zenodo.15610938
54. F. A. Fernando *et al.*, Data for: The MUC19 gene: An evolutionary history of recurrent introgression and natural selection, DRYAD (2025); doi: 10.5061/dryad.z612jm6pj

ACKNOWLEDGMENTS

We would like to thank A. Funk for contributing to the development of the *PBS* analysis, R. Pornmongkolsuk for early visualizations of global frequencies of *MUC19*, and D. O. del Vecchyo and P. Provero for their insightful comments and discussion. We also thank the Crawford and Ramachandran laboratories, especially R. Vinod, J. Stamp, C. Nwizu, C. Williams, and L. Darwin for their invaluable feedback and support throughout the duration of this project. Part of this research was conducted using computational resources and services at the Center for Computation and Visualization at Brown University. **Funding:** This work was funded by the following: The Leakey Foundation (to F.A.V.); National Institutes of Health (1R35GM128946-01 to E.H.S.); Alfred P. Sloan Foundation (to E.H.S.); Blavatnik Family Graduate Fellowship in Biology and Medicine (to D.P.); Brown University Predoctoral Training Program in Biological Data Science (NIH T32 GM128596 to D.P. and E.T.C.); National Institutes of Health Center of Biomedical Research Excellence (1P20GM139769 to K.E.W.); National Science Foundation (CAREER 2338710 to P.M.); Burroughs Wellcome Fund (Career Award at the Scientific Interface to P.M.); National Institutes of Health (R01NS122766 to P.N.V.); and Human Frontier Science Program (to E.H.S., F.J., and M.A.A.). **Author contributions:** Conceptualization: F.A.V., D.P., and E.H.S. Formal analysis: F.A.V., D.P., E.J.K., V.A.G., K.E.W., V.V.I., R.Z., D.M., P.M., F.J., P.N.V., M.A.A., and E.H.S. Supervision: D.M., P.M., F.J., P.N.V., M.A.A., and E.H.S. Writing – original draft: F.A.V., D.P., and E.H.S. Writing – review & editing: E.J.K., V.A.G., K.E.W., V.V.I., R.Z., D.M., P.M., F.J., P.N.V., and M.A.A. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** The 1000 Genomes Project Phase III, Simons Genome Diversity Project, high-coverage archaic genomes, Human Pangenome Reference Consortium, and Human Genome Structural Variant Consortium datasets are all publicly available. Ancient American genomes are available after signing data agreements from the original publications. All software used in this study is publicly available and all statistical tests are described in the supplementary methods. All the information needed to reproduce the results in this study is also described in the Supplementary Methods. Additionally, the original code and final results can be found at (*53*), intermediary files used to produce our final results can be found at (*54*), and the introgressed tracts, repeat information, phased late Neanderthal haplotypes, and datasets S1 to S4 can be found at (*36*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adl0882
Materials and Methods; Supplementary Text; Figs. S1 to S60; Tables S1 to S67; Datasets S1 to S4; References (*55–100*); MDAR Reproducibility Checklist

Submitted 27 September 2023; resubmitted 1 December 2024; accepted 9 June 2025
10.1126/science.adl0882