PNAS RESEARCH ARTICLE | EVOLUTION OPEN ACCESS

Neanderthal adaptive introgression shaped *LCT* enhancer region diversity without linking to lactase persistence in East Asian populations

Xixian Ma[a,b], Yan Lu[c], Mark Stoneking[d,e], and Shuhua Xu[c,f,1]

Affiliations are included on p. 8.

Edited by Nils Stenseth, Universitetet i Oslo, Oslo, Norway; received March 1, 2024; accepted February 8, 2025

Positive selection at the 2q21.3 enhancer region for lactase gene (*LCT*) expression in Europeans and Africans has long been attributed to selection for lactase persistence (LP), the capacity of adults to digest lactose in milk, presumably because of the benefits associated with milk consumption. While considered a classic example of gene–culture coevolution, recently doubts have been raised about the link between selection at 2q21.3 and LP. Analysis of additional populations could shed further light; here, we demonstrate that a haplotype spanning ~467 kb at the 2q21.3 locus has risen to high frequency in East Asians (~25%) but is absent from Africans and Europeans. This haplotype likely derived from Neanderthals and has been under positive selection in East Asians. The East Asian–specific haplotype is associated with alterations in *LCT* expression and promoter methylation in certain cell types, similar to what is observed with LP-associated haplotypes in Europeans. Moreover, its frequency is comparable to that of LP in East Asians, suggesting a potential association with LP in East Asians. However, it is highly unlikely that selection in East Asians was related to milk-drinking habits. We find that this haplotype impacts the expression of *UBXN4*, *DARS1*, and *DARS1-AS1* in immune cells and is associated with neutrophil and white blood cell counts. Hence, the selection might be linked to certain aspects of immune function. This implies that selection on 2q21.3 has thus either occurred for different reasons in different populations or the selection observed in other populations is also not due to LP.

LCT | adaptive introgression | Neanderthal | East Asian | lactase persistence

Significance

Positive selection at the 2q21.3 locus in Europeans and Africans has long been tied to lactase persistence (LP) and milk-drinking and is a textbook example of convergent evolution and gene–culture coevolution. However, the selection may not have been driven by LP alone. Here, we identified a high-frequency East Asian–specific haplotype at this locus, which came from Neanderthals and has been under positive selection. While this haplotype might be linked to LP, it is highly unlikely that the selection would have been for this reason. Instead, changes in immune function likely explain the selection in East Asians, thus indicating that selection was either for different reasons in different populations or selection was similarly not for LP in other populations.

The 2q21.3 locus harbors one of the strongest signals of positive selection in the human genome in European (1–3), eastern and southern African populations (4–6). The lactase gene (*LCT*) at the 2q21.3 locus encodes the enzyme lactase-phlorizin hydrolase, which is expressed in the small intestine of infants and is responsible for breaking down lactose into glucose and galactose, which can then be absorbed by the body (7, 8). Lactase activity begins to decline after weaning and is subsequently very low in all adult mammals (9)—with the notable exception of some humans. The haplotypes under selection in Europeans and Africans are strongly linked to the expression of intestinal lactase during adulthood, a trait known as lactase persistence (LP) (4).

"The frequency of LP exhibits significant variation among global populations, spanning a range from 5% to almost 100%. Notably, the highest frequencies are found in people of Northern European descent and some populations from West Africa, East Africa, and the Middle East" (10–13). It is generally thought that high frequencies of LP occurred in populations that domesticated animals and have used their milk since the Neolithic Revolution, about 5,000 to 10,000 y ago (11) (i.e., the cultural-historical hypothesis).

Previous studies have shown that LP is inherited as an autosomal dominant trait in Europeans (14). In 2002, a linkage disequilibrium (LD) and haplotype analysis of Finnish families identified the first mutation associated with the LP phenotype: −13910:C>T (rs4988235) (15). The T-13910 allele is 100% associated with LP in the Finnish study and is 86 to 98% associated with LP in other European populations (16). Subsequent studies have revealed at least five regulatory variants in various populations: −13910*T (rs4988235) in Europeans (15), Central Asians (17) and South Asians (18), −13915*G (rs41380347) in Saudi Arabians (19) and Africans (4, 20), and −13907*G (rs41525747), −14009*G (rs869051967), and −14010*C (rs145946881) in eastern and southern Africans (4, 6, 21, 22). All of these variants are located in the regulatory region ~14 kb upstream of the *LCT* gene, in intron 13 of the *MCM6* gene. The variants associated with LP in Europeans are different from those in Africans, providing an example of convergent evolution and gene–culture coevolution, because of the selective pressure resulting from

Author contributions: X.M. and S.X. designed research; X.M. performed research; X.M. and Y.L. analyzed data; and X.M., M.S., and S.X. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: xushua@fudan.edu.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2404393122/-/DCSupplemental.

Published March 10, 2025.

shared cultural practices, including animal domestication and the consumption of milk by adults (4).

However, there are still some unresolved issues concerning this story. One important question is whether the positive selection at the 2q21.3 locus has been driven by selection for LP alone. There is debate about whether and (if so) why being able to consume milk as an adult provides such a strong selective advantage (13, 23). Although the domestication of animals and consumption of milk began with the Neolithic, about 10,000 y ago (24), the LP-associated allele in Europeans only reached appreciable frequencies in the Bronze and Iron Ages (13, 23, 25). In addition, selection near the *LCT* locus predates the emergence of the LP-associated allele by thousands of years (26), suggesting the selection at the *LCT* locus is more complex than previously thought.

Analysis of other populations may provide additional insights into the putative association between selection at 2q21.3 and LP. For example, while the frequency of LP in East Asian populations has been reported to be between 3% and 28% (https://www.ucl.ac.uk/biosciences/gee/molecular-and-cultural-evolution-lab/Global-Lactase-persistence-Association-Database-GLAD), the five known LP-associated alleles are almost absent in East Asian populations (https://www.ucl.ac.uk/biosciences/gee/molecular-and-cultural-evolution-lab/Global-Lactase-persistence-Association-Database-GLAD). Potential variants that might explain LP in East Asian populations have yet to be identified, and the existence of any signal of positive selection on the 2q21.3 locus in East Asians has not been investigated.

We here investigate genetic variants in the 2q21.3 locus in East Asian populations and find that a haplotype at this locus was introgressed from Neanderthals and subsequently influenced by positive selection. We show that this haplotype is likely to be associated with LP in East Asians, but selection for LP is unlikely to explain the signal of positive selection. Instead, we observed genetic and cellular differences that may be associated with immune function. Our results provide insights into the evolutionary history of the 2q21.3 locus in worldwide populations: Either selection at this locus has occurred for different reasons in different populations or the selection at this locus in European or African populations is not associated with LP.

Results

The Known LP-associated Alleles Are Absent in East Asians. We first investigated the frequency distribution of five LP-associated alleles identified previously in other populations in East Asia, using the 1000 Genomes Project Phase 3 dataset (27). The −13907*G (rs41525747) and −14009*G (rs869051967) single nucleotide polymorphisms (SNPs) were not present in this dataset. We, therefore, investigated their frequency in the NCBI Allele Frequency Aggregator (ALFA) dataset (28) and found they were absent in East Asian populations. The three other LP-associated alleles were completely absent in all five East Asian populations from 1000 Genomes [i.e., Han Chinese in Beijing, China (CHB), Southern Han Chinese (CHS), Japanese in Tokyo, Japan (JPT), Kinh in Ho Chi Minh City, Vietnam (KHV), and Chinese Dai in Xishuangbanna, China (CDX)] (*SI Appendix*, Fig. S1). In addition, the −22018*A (rs182549) allele, which is in strong LD with −13910*T (rs4988235) in European populations (15), was also absent in East Asian populations. We further investigated the frequency of these LP-associated alleles in more diverse populations from the Human Genome Diversity Project (29). The LP-associated allele of rs4988235 was still absent in all East Asian populations. Thus, the five known LP-associated alleles are absent in all tested East Asian populations.

An East Asian–Specific Haplotype of High Frequency. To characterize genetic diversity at the 2q21.3 locus in Eurasia, we used the fixation index F_{ST} (30) to quantify genetic differentiation between Europeans [represented by Utah Residents with Northern and Western European Ancestry (CEU)] and East Asians (represented by CHB) at the 2q21.3 locus. Some variants showed extremely high differentiation ($F_{ST} \geq 0.7$) between European and East Asian populations, while others showed lower but still high differentiation ($F_{ST} \geq 0.2$) (Fig. 1*A*). To investigate in which population the genetic differentiation occurred, we calculated the locus-specific branch length (LSBL) (31) for each locus within this region for European (i.e., CEU) and East Asian (i.e., CHB) populations with African populations [i.e., Yoruba from Ibadan, Nigeria (YRI)] as an outgroup, respectively (Fig. 1 *B* and *C*). Specifically, one class of variants exhibited extremely high values for the European-specific branch length (LSBL for European populations ≥ 0.4, which is the top 0.1% of significance threshold of the empirical distribution) (Fig. 1*B*). For example, the derived allele of rs4988235 reached high frequencies in northern European populations [e.g., 73.7% in CEU, 72.0% in British in England and Scotland (GBR), 45.8% in Iberian Population in Spain (IBS), and 8.9% in Toscani in Italia (TSI)] but was absent in African and East Asian populations (Fig. 1*D*). This is consistent with positive selection in European populations (32). Another class of variants showed high East Asian–specific branch lengths (LSBL for East Asian populations > 0.24, which is in the top 1% of the empirical distribution) (Fig. 1*C*). For example, the derived allele of rs4988245 reached high frequencies in East Asians (e.g., 25.2% in CHB, 21.9% in CHS, 12.1% in KHV, and 11.8% in CDX) but was absent in African and European populations (Fig. 1*E*). The variants exhibiting high East Asian–specific branch lengths are in LD, and analysis of the phased haplotypes from East Asian populations from 1000 Genomes Project Phase 3 indicates that these alleles were located on one haplotype (Fig. 2*A*). The results showed that the genetic differentiation between East Asians and Europeans was caused not only by some alleles reaching extremely high frequencies in European populations (Fig. 1 *B* and *D*) but also by some alleles reaching high frequencies in East Asian populations (Fig. 1 *C* and *E*).

Neanderthal Origin of the East Asian–Specific Haplotype. We then focused on this ~467 kb highly differentiated region between Europeans and East Asians (chr2: 136,322,676–136,789,259, GRCh37), which was defined by variants with F_{ST} (CEU vs. CHB) ≥ 0.2. We defined variants with LSBL (CHB; CEU, YRI) for East Asians > 0.2 and LSBL (CEU; CHB, YRI) for Europeans < 0.01 as highly differentiated variants in East Asians. We similarly defined variants with LSBL (CEU; CHB, YRI) for Europeans > 0.2 and LSBL (CHB; CEU, YRI) for East Asians < 0.01 as highly differentiated variants in Europeans (Fig. 1). We then analyzed the number of variants with derived alleles absent in African populations (*SI Appendix*, Fig. S2). For 12 of the 156 highly differentiated variants in European populations, the derived alleles were absent in African and East Asian populations, but reached high frequencies in European populations (e.g., ~73.7% in CEU). For 153 of 155 highly differentiated variants in East Asian populations, the derived alleles were absent in African and European populations but reached moderate frequencies in East Asian populations (e.g., 25.7% in CHB). This significantly different pattern (Chi-square test, $P < 2.2 \times 10^{-16}$) suggested that putative adaptive haplotypes may have different origins in European and East Asian populations.

We focused on highly differentiated variants between Europeans and East Asians with the derived allele absent in African populations

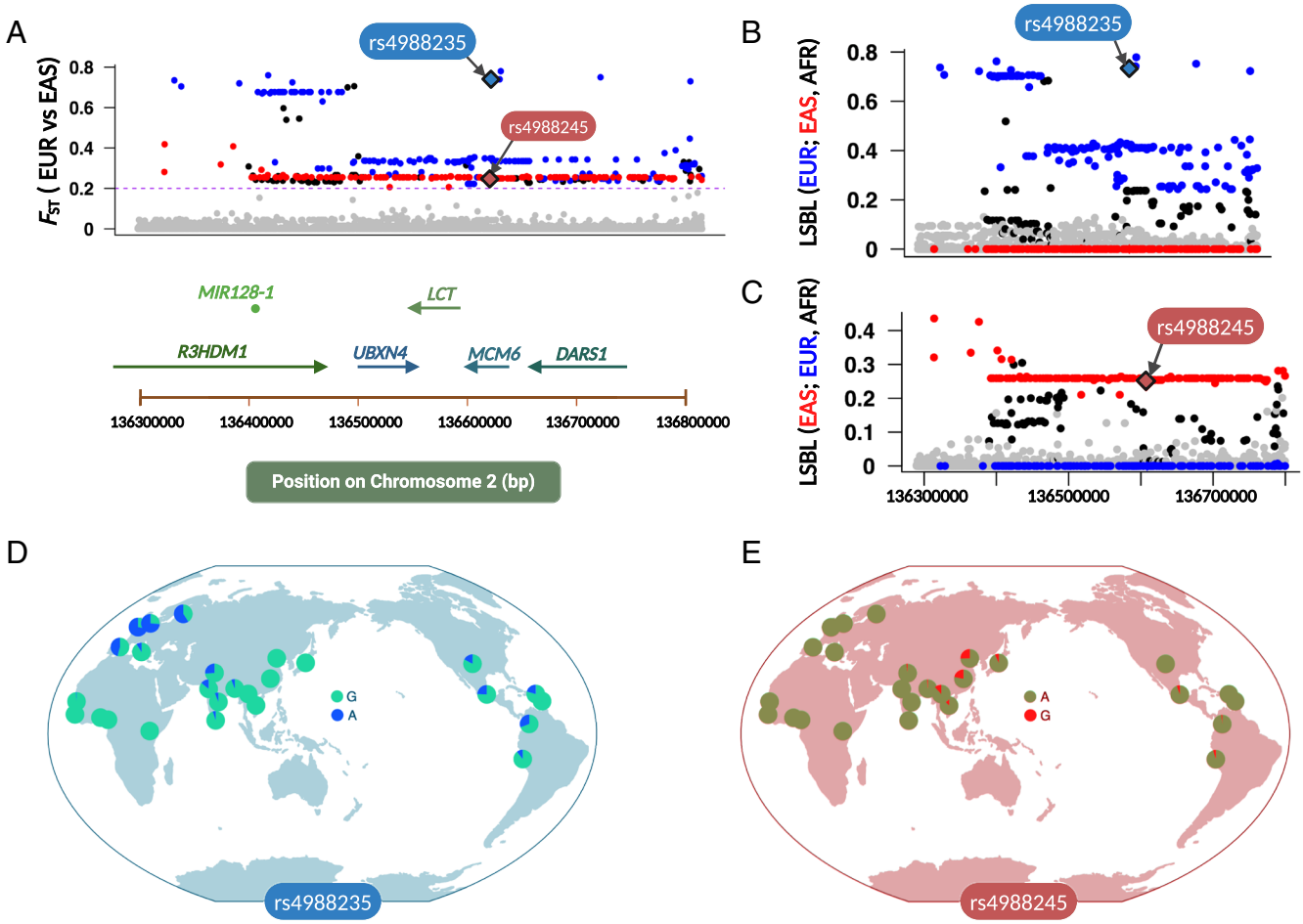

Fig. 1. Genetic differentiation between East Asians and Europeans at the 2q21.3 locus. (*A*) Genetic differentiation between East Asian and European populations. The dashed line indicates F_{ST} = 0.2. The *Bottom* panel shows the protein-coding and micro-RNA genes in this region. The *x* axis indicates the GRCh37 coordinate. (*B*) The genetic differentiation in European populations at the *LCT* region. The *y* axis represents the locus-specific branch length of each locus for European populations. (*C*) The genetic differentiation in East Asian populations at the *LCT* region. The *y* axis represents the locus-specific branch length of each locus for East Asian populations. Each point represents one variant. The blue and red colors indicate variants that have undergone large allele frequency changes in European and East Asian populations (as determined by the locus-specific branch lengths analysis), respectively. (*D*) The allele frequency distribution of rs4988235 in worldwide populations. (*E*) The allele frequency distribution of rs4988245 in worldwide populations.

and tested whether these alleles could match archaic genomes (Fig. 2*A*). None of the highly differentiated variants in Europeans matched either the Altai Neanderthal or the Altai Denisovan genomes. However, out of 150 highly differentiated variants in East Asians, 113 (~75%) matched the Altai Neanderthal (33), and 67 (~45%) matched the Altai Denisovan genome (34). All of the introgressed alleles that matched the Altai Denisovan genome matched the Altai Neanderthal genome. We also used stricter filter criteria: filtering sites with coverage depth below 10, or mapping quality below 25, or that are within tandem repeats or indels, or that have poor mappability (35). With these stricter filters, out of 150 highly differentiated variants in East Asian populations, 79 overlapped the Altai Neanderthal genome and 59 (~75%) matched, while 79 overlapped the Vindija33.19 Neanderthal (36) and 76 (~96%) matched. This result suggests that the East Asian–specific haplotype might reflect gene flow from a Neanderthal-related population.

We further constructed a haplotype network for haplotypes from present-day populations defined by these 161 highly differentiated variants with alleles absent in African populations (Fig. 2*B*). One set of closely related haplotypes carrying the LP-associated allele reached high frequency (~65%) in European populations and was absent in African and East Asian populations. Another set reached high frequency (~15%) in East Asians and was absent in African and European populations; we refer to these collectively as the East Asian–specific haplotype. This East Asian–specific haplotype clustered with the haplotype from the Altai Neanderthal.

To confirm that the East Asian–specific haplotype was from Neanderthal-related populations, we also constructed a neighbor-joining tree (37) for haplotypes using all variants in this highly differentiated region between European and East Asian populations. The neighbor-joining phylogenetic tree was constructed for haplotypes from East Asian (i.e., CHB), European (i.e., CEU), and African (i.e., YRI) populations together with haplotypes from Altai Neanderthal, Altai Denisovan, and chimpanzee, based on nucleotide distance. Some haplotypes from East Asian populations first clustered with Neanderthal haplotypes, suggesting that this East Asian–specific haplotype might be from Neanderthal-related populations (Fig. 2*C*).

Several methods have been developed to detect introgressed segments in present-day modern human genomes (38–41). We also analyzed whether previous studies have detected putatively introgressed segments in East Asian populations in this region. The highly differentiated ~467 kb region overlapped with putatively introgressed segments in East Asians detected by Archaicseeker (~397 kb, 85%) (42), by IBDmix (41) (~261 kb, 55.9%), and by S* (~88 kb, 19%) (39). It also overlapped with putatively introgressed

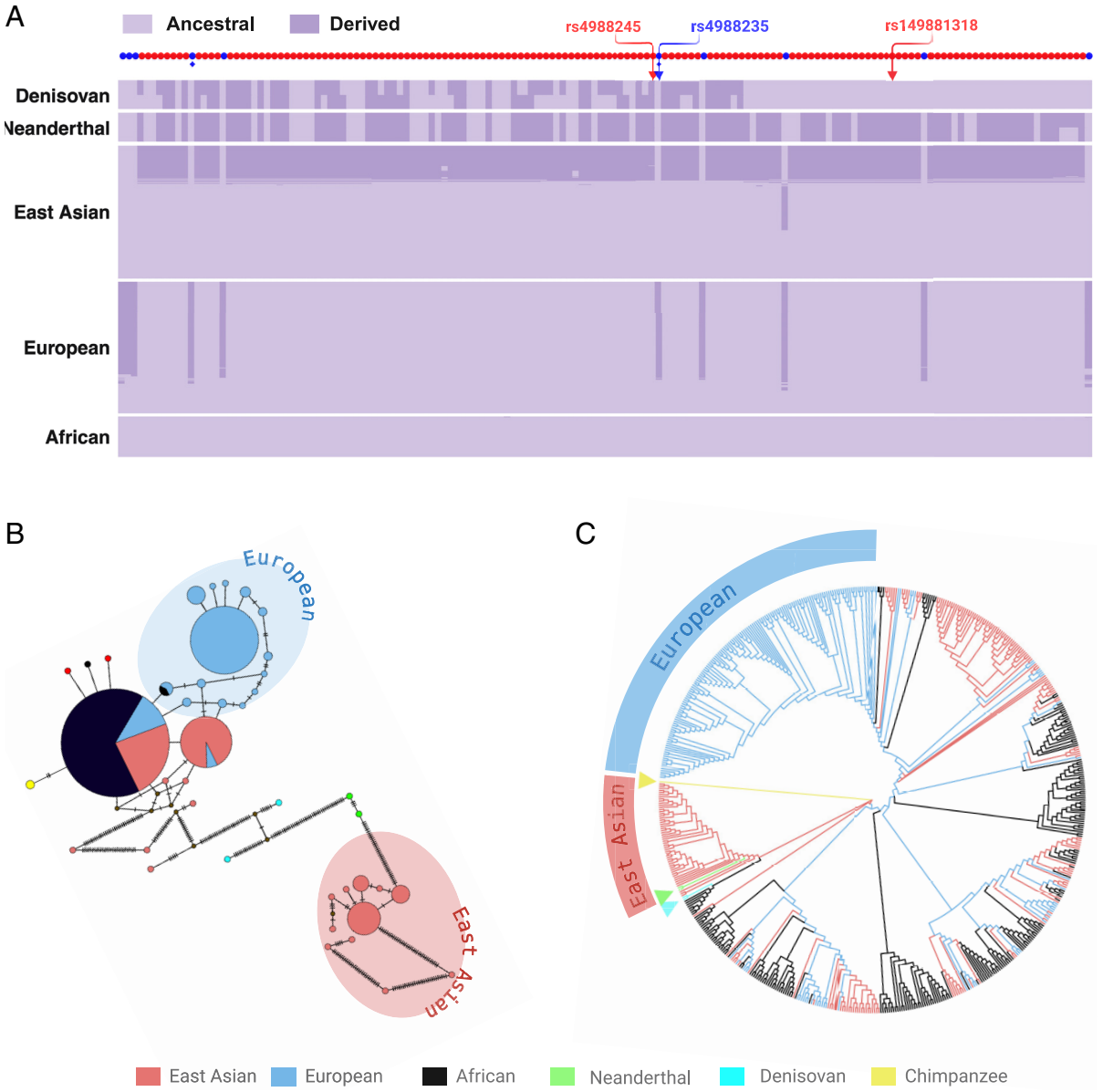

Fig. 2. Neanderthal origin of the East Asian–specific haplotype. (*A*) Allelic states of haplotypes from East Asian (represented by CHB), European (represented by CEU), and African (represented by YRI) populations, together with haplotypes from Altai Neanderthal and Altai Denisovan. These representative haplotypes at the 2q21.3 locus are defined by highly differentiated SNPs, which are at high frequency in European or East Asian populations and carry an allele absent in African populations. The *Top* panel shows variants that were used to define haplotypes. The next panel shows the allele state for these haplotypes. (*B*) The network for haplotypes from present-day modern populations together with Altai Neanderthal, Altai Denisovan, and Chimpanzee. The haplotypes were defined by the highly differentiated SNPs. Each circle represents one unique haplotype, with the size of the circle proportional to the number of chromosomes with that haplotype, colored according to population. (*C*) The neighbor-joining tree for the inferred sequences of modern human haplotypes and the Neanderthal, Denisovan, and Chimpanzee genome sequences. Branches are colored according to population.

segments in East Asians detected by Sankararaman et al. (~312 kb, 67%) (38), whereas SPrime (40) did not detect any archaic introgressed segments in East Asia in this region. As both S* and SPrime use the LD pattern to detect putatively introgressed segments, and SNPs in this region are in strong LD (*SI Appendix*, Fig. S3), these methods may not have enough power to find the putatively introgressed segments. Overall, these results are consistent with the inference that the East Asian–specific haplotype might have introgressed from Neanderthal-related populations.

The Neanderthal-Derived Haplotype Underlying Positive Selection. We next employed some formal statistics to examine whether this region has been under positive selection in East Asians. We first applied the Derived Intra-allelic Nucleotide Diversity (DIND) test (43), which revealed marginal signs of positive selection (empirical P value < 0.05) (Fig. 3*A*). Additionally, we found evidence of extended haplotype homozygosity (EHH) (32) for haplotypes with putatively introgressed alleles (Fig. 3*B*): The putatively introgressed alleles showed longer extended haplotype homozygosity than nonintrogressed alleles. Some haplotypes carrying putatively introgressed alleles also showed a significant signal in the iHS test (44) ($|iHS| > 2$) (*SI Appendix*, Fig. S4). We also performed Tajima's D-test (45) across the region and found that the introgressed haplotypes at the *LCT* region exhibited a significantly negative value (Fig. 3*C*), whereas the flanking regions and nonintrogressed haplotypes did not. All of these results suggest that the introgressed haplotype was under recent positive selection in East Asians.

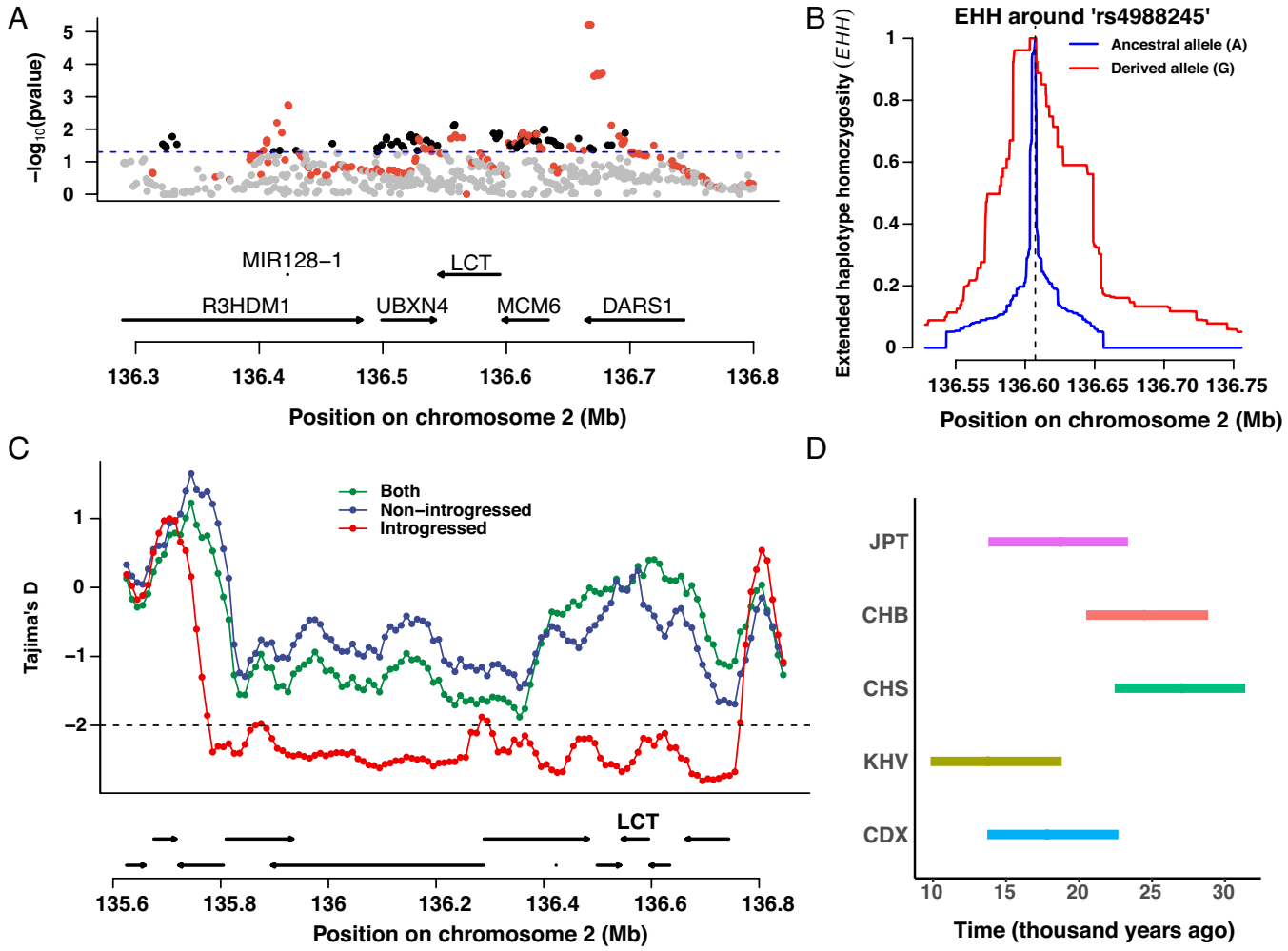

Fig. 3. The signal of positive selection for the East Asian–specific haplotype. (*A*) The DIND test result for the 2q21.3 locus. Each point represents one SNP with the derived allele >0.1. The *y* axis indicates the empirical *P*-value of this SNP in the DIND test. The blue dashed line indicates the 5% empirical distribution threshold. The red color indicates SNPs with introgressed alleles. (*B*) The extended haplotype homozygosity (EHH) results for rs4988245 for all haplotypes from the CHB population. (*C*) The Tajima's D result for the 2q21.3 locus. Each point represents one 50-kb window. For each window, we calculated Tajima's D for all the haplotypes, the introgressed haplotypes, and the nonintrogressed haplotypes from the CHB population. (*D*) The estimate of the selection time for rs3087348 in East Asian populations. The selection time was estimated separately for each population (see *Materials and Methods* for details).

Evolution of the East Asian–Specific Haplotype. We used several methods to infer when the putative positive selection at the 2q21.3 locus occurred in East Asians, based on both present-day genomes and ancient DNA. We employed startmrca (46) to estimate the time of onset of selection at this locus in five East Asian populations. The variant rs3087348 was used to tag the East Asian–specific haplotype (Fig. 3*D*). The estimated selection time was 24,478 (95% CI: 20,481 to 28,841) y ago in the northern Han Chinese population (CHB) and 27,048 (95%CI: 22,431 to 31,382) in the southern Han Chinese population (CHS). The youngest estimate was 13,756 (95%CI: 9,825 to 18,807) in a Vietnam population (KHV), which suggests either a more recent selection pressure appeared in the region, or a more recent arrival of this haplotype in Southeast Asia.

Sequencing ancient DNA from temporally spaced samples makes it possible to directly gauge how frequencies of genetic variants have changed over time. We investigated allele frequency change using the currently available Allen Ancient DNA Resource (AADR) (version 9) dataset (47). For all variants with a derived Neanderthal-like allele and included in AADR "1240 K" dataset, rs3087348 had higher genotype quality. We, therefore, used this variant to tag the East Asian–specific haplotype (Fig. 4*A*). Among 19 ancient samples dated to 10,000 y ago or earlier, eight samples have data for this variant. The East Asian–specific allele at rs3087348 was observed in one individual (AR14K) from the Amur region of China ~14,000 y ago (48). The AR14K-related population is the closest East Asian source for Ancient Paleo-Siberians, who are closely related to contemporary people from Native Americans (49). Notably, this allele was also observed in some American samples, such as one sample (USR1) from the Upward Sun River in Alaska in the United States around 11,425 y ago (50) and another sample (Kennewick Man) from the Columbia River in Washington State in the United States ~8,700 y ago (51). These results suggested this haplotype may have been widely present in East Asian populations ~14,000 y ago. Because some ancient samples are pseudodiploid, we calculated the proportion of carriers of this haplotype rather than directly calculating the frequency. Between 8,000 and 3,000 y ago, the proportion of carriers of this haplotype was ~10%. Subsequently, it has risen in frequency to ~20% until 3,000 to 1,000 y ago. The current proportion of carriers of this haplotype in East Asians is 28.9%. The lack of ancient DNA data from older times precludes estimating when the allele arose, but it seems to have been present at high frequencies by ~8,000 y ago.

We also inferred the change in frequency of these putative adaptive alleles over time, using all haplotypes from five different East Asian populations separately, with RELATE (52) and CLUES (53).

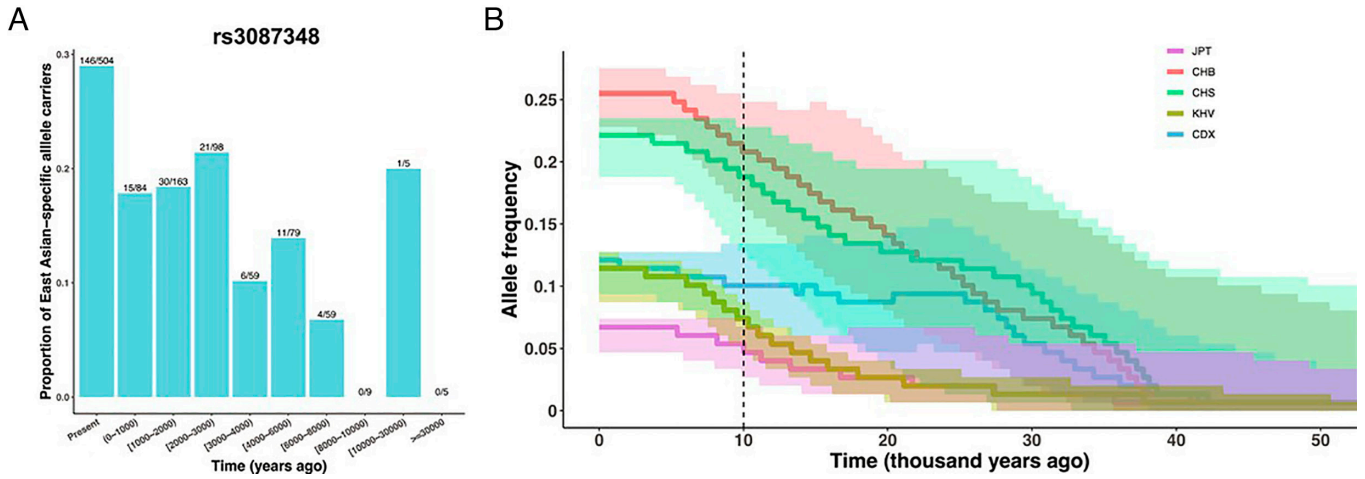

Fig. 4. Allele frequency change over time. (*A*) The proportion of carriers of the East Asian–specific alleles at rs3087348 across time. The number above each bar represents the number of samples carrying the East Asian–specific allele and the total number of samples with known genotypes. (*B*) The allele frequency changes over time for rs3087348. The solid line indicates the allele frequency change inferred by RELATE and CLUES using all haplotypes from different East Asian populations. The filled region indicates the 95% CI.

The variant rs3087348 was used to tag this East Asian–specific haplotype (Fig. 4*B*); for the CHB population, the frequency of the derived allele T reached 1.3% about 41,006 y ago, increased to 10.1% about 25,897 y ago, and reached 20% about 12,097 y ago. Thus, all of the analyses suggest this haplotype began rising in frequency more than 10,000 y ago and has continued to increase up to the present.

Is the East Asian–Specific Haplotype Associated with LP? To investigate this question, we attempted to obtain DNA samples from East Asians who had been tested for the LP phenotype, however, we were unable to identify a source of such samples. We, therefore, looked for expression and methylation differences at the 2q21.3 locus that have been previously associated with LP in European populations (54, 55).

We acquired gene expression data for 28 distinct immune cell subsets from Japanese populations using the ImmuNexUT (Immune Cell Gene Expression Atlas from the University of Tokyo) dataset (56). According to the GTEx portal (57), the *LCT* gene is highly expressed in the small intestine but expressed only at very low levels (if at all) in other tissues. However, we found that the expression of the *LCT* gene was higher in plasmacytoid dendritic cells (pDCs) than in other immune cell subtypes (*SI Appendix*, Fig. S5). The derived allele T of rs149881318 reached high frequencies in East Asians (e.g. 26.2% in CHB, 22.4% in CHS, 7.2% in JPT, 14.0% in CDX, and 13.6% in KHV) and is absent in Europeans and Africans. Individuals carrying the East Asian–specific alleles exhibited significantly higher levels of the *LCT* expression (Fig. 5*A*). Importantly, our result suggested that the East Asian–specific haplotype was associated with increased expression of the *LCT* gene.

DNA methylation levels at cg20242066 in intestinal tissue have been linked to LP and correlated with the genotype of rs4988235 in Europeans (54). We analyzed methylation data from whole blood samples from 3,523 Chinese individuals (58). Alleles on the East Asian–specific haplotype are associated with increased methylation levels at cg20242066 at the *LCT* promoter (Fig. 5*B*). Hence, both the expression and methylation results indicate that the East Asian–specific haplotype has the same regulatory effect on the *LCT* gene as is observed for the haplotype that is associated with LP in Europeans, suggesting that the East Asian–specific haplotype is also associated with LP.

The Regulatory Effect of the East-Asian Specific Haplotype on Other Genes at the 2q21.3 Locus. This East Asian–specific haplotype covers several protein-coding and microRNA genes, including *R3HDM1*, miR-128-1, *UBXN4*, *LCT*, *MCM6*, *DARS1,* and *DARS1-AS1* et al. Using data from ImmuNexUT (56), we found that individuals carrying the East Asian–specific alleles exhibited significantly higher levels of *UBXN4* in some immune cell types (e.g., pDCs) (*SI Appendix*, Fig. S6). Individuals carrying the East Asian–specific alleles exhibited significantly lower levels of *DARS1* in some immune cell types (e.g., T follicular helper cells) (*SI Appendix*, Fig. S7). Conversely, individuals carrying the East Asian–specific alleles exhibited significantly higher levels of *DARS1-AS1* in some immune cells (*SI Appendix*, Fig. S8). Additionally, we analyzed gene expression data from whole blood from 39 Han Chinese people (59). The *DARS1* gene showed significantly lower expression in individuals homozygous for the East Asian–specific alleles (*t* test, $P < 0.05$) (*SI Appendix*, Fig. S9); however, expression levels of other genes did not differ among genotypes. The results indicated that the East Asian–specific haplotype enhances the expression of the *UBXN4* gene, reduces the expression of the *DARS1* gene, and increases the expression of the *DARS1-AS1* gene in some immune cell types.

We further assessed the functional consequences of this East Asian–specific haplotype using data from genome-wide association studies (60). For 150 SNPs with East Asian–specific alleles, seven SNPs are significantly associated with neutrophil count and white blood cell count (61–63). These SNPs are all located in the *DARS1* gene. For example, the SNP rs149881318 has been reported to be associated with neutrophil count (61). The derived allele is at a high frequency in East Asian populations (e.g., 26% in CHB) and was absent in European and African populations.

No Link Between the −13838*A Allele in Tibetans and the East Asian–Specific Haplotype. One previous study suggested that the SNP −13838:G>A (chr2: 136,608,574, C>T) might be associated with LP in Tibetan populations (64). To investigate the relationship between the −13838*A allele and the East Asian–specific haplotype, we obtained whole-genome sequencing data from 33 Tibetan samples (65). The derived T allele was about 7.6% in the Tibetan population and was absent in Han Chinese, African, and European populations, which is consistent with the previous finding that the frequency of the T allele was about 6.6%, ranging from 1.9 to 20.8% in Tibetan populations (66).

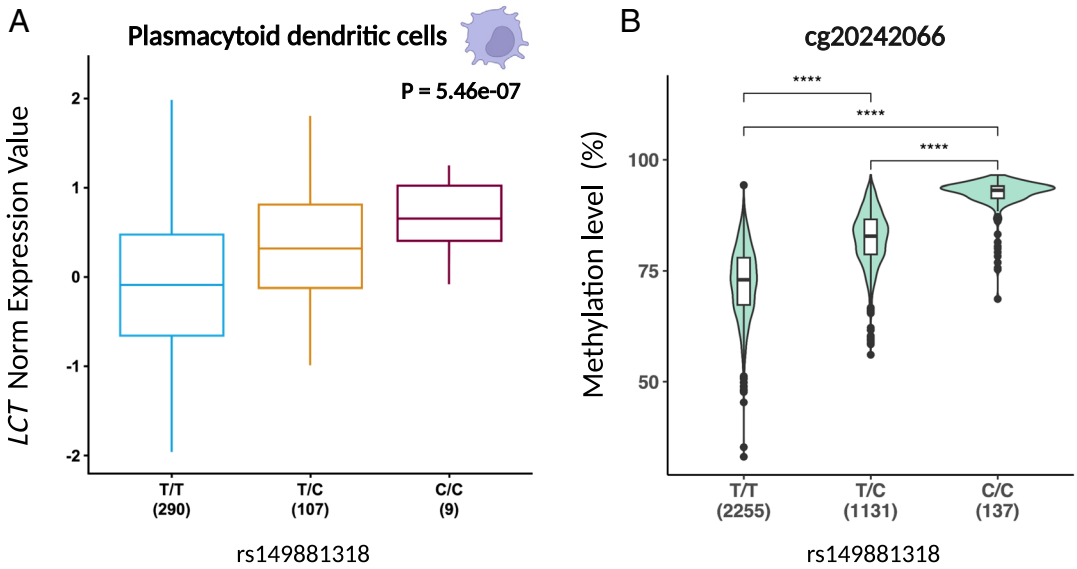

Fig. 5. Expression and methylation change for the *LCT* gene/region associated with the East Asian–specific haplotype. (*A*) The relationship between the expression of the *LCT* gene and the genotype at rs149881318 in pDCs in Japanese. (*B*) The relationship between the methylation level at cg20242066 at *LCT* promoter and the genotype at rs149881318 in whole blood in Han Chinese.

We used SNP rs4988245 (chr2: 136,607,300) to tag the East Asian–specific haplotype, which is closest to rs4988235 (chr2: 136,608,646), associated with LP in Europeans. The frequency of the derived G allele of rs4988245 is higher in Tibetan (i.e., 27%) than in other East Asian populations. The haplotype structure results demonstrate that the −13838*A allele is not located on the East Asian–specific haplotype and that the East Asian–specific haplotype is also present at high frequency in Tibetans (*SI Appendix*, Fig. S10) and is, therefore, a potential explanation for LP in this population.

Discussion

The 2q21.3 locus, harboring the *LCT* gene, harbors the strongest signals of positive selection in European and African populations (1–6, 67). Despite being a textbook example of convergent regulatory evolution and gene–culture coevolution, the story of positive selection at this locus remains somewhat elusive, especially in non-European populations (13).

In this study, we first investigated the genetic differentiation between East Asian and European populations at the 2q21.3 locus. The genetic differentiation was caused not only by some alleles reaching extremely high frequencies in European populations but also by some alleles reaching high frequencies in East Asian populations. Our investigations suggest that this East Asian–specific haplotype has experienced positive selection in East Asians. Furthermore, we find that this East Asian–specific haplotype likely reflects gene flow from Neanderthals, adding further evidence for the importance of adaptive introgression in the evolutionary history of our species (68–70).

The frequency of the East Asian–specific haplotype in East Asians (6.7 to 25%) is similar to the frequency of LP of 3 to 28% (https://www.ucl.ac.uk/biosciences/gee/molecular-and-cultural-evolution-lab/Global-Lactase-persistence-Association-Database-GLAD), suggesting that this haplotype might confer LP in East Asians. To investigate this further, we explored gene expression and methylation data, as differences in both have been reported for European variants associated with LP (54, 55, 71), Notably, the East Asian–specific haplotype is associated with an increase in the expression of the *LCT* gene in pDCs from Japanese, and also with an increase in methylation levels at the cg20242066 region in whole blood from Han Chinese (Fig. 5). These results are consistent with expression changes associated with the European LP-associated allele −13910*T in intestinal tissue (71), but this allele is associated with reduced DNA methylation levels at cg20242066 in intestinal tissue in Europeans (54). While overall these results suggest a possible link between the East Asian–specific haplotype and LP, more work is needed to substantiate this link.

The frequency trajectory indicated that positive selection likely occurred within the ancestral East Asian hunter-gatherer populations, and it is highly unlikely that the positive selection is related to LP phenotype. A recent study also proposed that the selection signal on the *LCT* region predates the emergence of the major LP allele (rs4988235-A) in Europeans and the ancestry-stratified analysis revealed that the positive selection signal of the major LP allele in Europeans is driven by sweeps in hunter-gatherer ancestry [Eastern hunter-gatherer (EHG) and Caucasus hunter-gatherer (CHG)], not Neolithic ancestry [Anatolian Farmers (ANA)] (26). This provided evidence that the selection on the *LCT* region in Europeans may not be associated with LP phenotype alone but be associated with an ancient adaptation to famine or increased pathogen exposure (23, 72).

Our results suggest that this East Asian–specific haplotype is associated with the enhanced expression of the *UBXN4* gene, decreased expression of the *DARS1* gene, and increased expression of the *DARS1-AS1* gene. The *UBXN4* gene encodes an integral membrane protein of the endoplasmic reticulum (ER) that binds valosin-containing protein and promotes ER-associated protein degradation. Interestingly, according to the GTEx portal (57), the European-specific haplotype associated with LP also enhances the expression of *UBXN4*, reduces the expression of the *DARS1* gene, and increases the expression of the *DARS1-AS1* gene in multiple tissues (*SI Appendix*, Fig. S9). For example, the A allele of rs4988235 is associated with enhanced expression of the *UBXN4* gene in skin not exposed to sun [normalized effect size (NES) = 0.12, $P = 7.3 \times 10^{-7}$], reduced expression of the *DARS1* gene in skin not exposed to sun (NES = –0.12, $P = 3 \times 10^{-5}$) and increased expression of the *DARS1-AS1* in whole blood (NES = 0.22, $P = 4.8 \times 10^{-9}$) (*SI Appendix*, Fig. S9). While it is intriguing that variants at the 2q21.3 locus that were independently selected for in

European and East Asian populations reduce *DARS1* expression, the phenotypic consequences are unknown. *DARS1* is a member of the aminoacyl-tRNA synthetases (ARS) family, enzymes that are essential for protein synthesis as they catalyze the attachment of specific substrate amino acids to their cognate tRNAs. In addition to their classical function in translation, ARSs also function as regulators of cellular processes by sensing and responding to different cellular conditions. ARSs are integral to the maturation, transcription, activation, and recruitment of immune cells, thus playing a crucial role in immune cell development. As regulators and signaling molecules, ARSs influence various immune diseases, such as autoimmune diseases, infectious diseases, and tumor immunity (73). *DARS1* encodes the enzyme aspartyl-tRNA synthetase1, mainly expressed in the cell membrane and cytoplasm. Both *DARS1* and its antisense RNA (*DARS1-AS1*) have gained significant attention for their involvement in tumor progression (74–76). Notably, a recent study has reported a significant association between *DARS1* expression and immune cell composition in myeloproliferative neoplasms (MPNs) patients, suggesting an interplay between *DARS1* and the immune microenvironment (77). Furthermore, the protein encoded by *DARS1* has been shown to promote gastrointestinal tumor progression by regulating immune cells such as CD4+ T cells and related signaling pathways (78). These findings indicate that *DARS1* and *DARS1-AS1* play a critical role in different types of immune responses, potentially explaining the *LCT* region selection in East Asians. Moreover, as there are many other genes at this locus, the target of selection in East Asians requires further investigation.

To conclude, we show that an East Asian–specific haplotype that was introgressed from Neanderthals has experienced positive selection beginning 25 to 28 thousand years ago and that there is suggestive evidence associating this haplotype with LP in East Asians. Given the origin of this haplotype from Neanderthals and the old age for the onset of selection, it is highly unlikely that the reason for positive selection involves LP (and even more unlikely if it turns out that the haplotype is not associated with LP). We suggest the *UBXN4*, *DARS1,* and *DARS1-AS1* genes are candidate targets of selection, but there are others, so this requires further investigation. Nonetheless, we can conclude that either there has been selection on the 2q21.3 locus for different reasons in different populations, or the selection in European and African populations is also not associated with LP. In either event, the classic story of selection on the 2q21.3 locus as reflecting gene–culture coevolution becomes more complicated (and hence, more interesting).

Materials and Methods

Additional details are in *SI Appendix, Materials and Methods*.

Dataset. To investigate the genetic diversity and frequency distribution of SNPs of interest, we collected currently available whole-genome sequencing and genotyping data. We used the 1000 Genomes Project Phase 3 dataset, which includes 2,504 low-coverage (~4×) whole-genome sequencing data from 26 populations (27). The sample size for each population ranges from 61 to 113, with a median of 99, which is large enough for haplotype analysis. The high-coverage (~30×) whole-genome sequencing 1000 Genomes Project dataset (79) was also included. We also obtained 929 high-coverage (~30×) whole-genome sequences from 54 populations from the Human Genome Diversity Project (HGDP) (29), which contains more diverse populations. The sample size ranges from 6 to 46 with a median of 13. We also obtained 33 high-coverage (~30×) whole-genome sequences from the Tibetan population (65). This dataset was phased using Beagle 5.0 with all default parameters. To investigate how the allele frequencies of SNPs of interest changed over time, we employed available ancient DNA data from the 1240 K dataset [Harvard Dataverse, 9.0 data release (September 16, 2024)] (47). We clustered these samples into different groups by time and geography (*SI Appendix*, Table S1). The high-coverage reference genomes for the Altai Neanderthal (33), Altai Denisovan (34), Chagyrskaya Neanderthal (80), and Vindija 33.19 Neanderthal (36) were included in some analyses.

Statistical Analysis. An unbiased estimate of F_{ST} value was calculated for each locus using Weir and Cockerham's calculation (30). The locus-specific branch length (LSBL) (31) was used to calculate the population-specific differentiation. The extended haplotype homozygosity (EHH) (32) and the integrated haplotype score (iHS) (44) were calculated with the R package REHH 3.0 (81) to detect the signal of positive selection. For each SNP, the ancestral and derived states were retrieved from the Ensembl database release 71, which was determined according to the alignment for six primates (82). The Derived Intra-allelic Nucleotide Diversity (DIND) test (43) and Tajima's D test (45) were also used to search for signals of positive selection. We obtained segments identified by previous studies (38–41), such as SPrime (40) and IBDmix (41). We also used ArchaicSeeker 2.0 (42) to detect archaic segments in East Asian populations from the 1000 Genomes Project Phase 3 dataset. The default values were used for all parameters as suggested in the protocol paper (83). We constructed a haplotype network, using POPART (84) with the median-joining network strategy (85), for the haplotypes of individuals from East Asian (i.e., CHB), European (i.e., CEU), and African (i.e., YRI) populations from the 1000 Genomes Project Phase 3 together with the Altai Neanderthal (33), Altai Denisovan (34) genome sequences, and the Chimpanzee genome sequence (86). We also constructed a neighbor-joining phylogenetic tree (37) using FastME (87) and the substitution model of Tamura-Nei (88). The tree was visualized using the "ggtree" package (89). We used startmrca (46) to estimate the selection-onset time of the inferred beneficial alleles. For each locus, we used a 1-Mb region surrounding it and the HapMap combined recombination map (90) in this flanking region. We used RELATE (52) and CLUES (53) to infer the frequency trajectories of the alleles at SNPs of interest at this locus. We analyzed gene expression using published data from whole blood from 39 Han Chinese samples (59). We also queried whether SNPs of interest could affect the expression in the ImmuNexUT dataset (Immune Cell Gene Expression Atlas from the University of Tokyo) (56). We also investigated whether SNPs of interest could affect the DNA methylation level using data from whole blood from 3,523 Chinese individuals (58). Variants were annotated using the Ensembl Variant Effect Predictor (VEP) (91) with the corresponding VEP-compiled annotation database (v109_GRCh38). We also analyzed whether these variants have regulatory effects on genes using the Ensembl regulatory annotation database (92) and investigated the phenotypic legacy of the putative adaptive alleles using the association results from the GWAS catalog (60) (accessed on 15 December 2020). Figs. 1, 2, and 5 were created with BioRender.com.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*. Previously published data were used for this work (27, 29, 33, 34, 36, 39–41, 65, 80).

ACKNOWLEDGMENTS. We are grateful to Prof. Sijia Wang and Dr. Qianqian Peng for their help in preparing the methylation data. This study was supported by the National Key Research and Development Program of China (No. 2023YFC2605400), the National Natural Science Foundation of China Grants (32030020, 32288101, and 32470649), the Shanghai Science and Technology Commission Program (23JS1410100), and the Office of Global Partnerships (Key Projects Development Fund). The computational work in this research was performed using the Computing for the Future at Fudan (CFFF) platform of Fudan University. The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Author affiliations: [a]Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China; [b]Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen 518055, China; [c]State Key Laboratory of Genetic Engineering, Human Phenome Institute, Zhangjiang Fudan International Innovation Center, Center for Evolutionary Biology, School of Life Sciences, Department of Liver Surgery and Transplantation Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai 200032, China; [d]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig D04103, Germany; [e]Biométrie et Biologie Évolutive, Unité mixte de recherche 5558, CNRS & Université de Lyon, Lyon 69622, France; and [f]Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200032, China

1. S. R. Grossman *et al.*, Identifying recent adaptations in large-scale genomic data. *Cell* **152**, 703–713 (2013).
2. B. F. Voight, S. Kudaravalli, X. Wen, J. K. Pritchard, A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
3. T. Bersaglieri *et al.*, Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genetics* **74**, 1111–1120 (2004).
4. S. A. Tishkoff *et al.*, Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40 (2007).
5. E. Macholdt, M. Slatkin, B. Pakendorf, M. Stoneking, New insights into the history of the C-14010 lactase persistence variant in Eastern and Southern Africa. *Am. J. Phys. Anthropol.* **156**, 661–664 (2015).
6. E. Macholdt *et al.*, Tracing pastoralist migrations to southern Africa with lactase persistence alleles. *Curr. Biol.* **24**, 875–879 (2014).
7. T. Sahi, Dietary lactose and the aetiology of human small-intestinal hypolactasia. *Gut* **19**, 1074 (1978).
8. D. M. Swallow, Genetics of lactase persistence and lactose intolerance. *Annu. Rev. Genet.* **37**, 197–219 (2003).
9. G. Flatz, "Genetics of lactose digestion in humans" in *Advances in Human Genetics*, H. Harris, K. Hirschhorn, Eds. (Springer US, Boston, MA, 1987), pp. 1–77, 10.1007/978-1-4757-0620-8_1.
10. T. Sahi, Genetics and epidemiology of adult-type hypolactasia. *Scand. J. Gastroenterol.* **29**, 7–20 (1994).
11. C. J. E. Ingram, C. A. Mulcare, Y. Itan, M. G. Thomas, D. M. Swallow, Lactose digestion and the evolutionary genetics of lactase persistence. *Hum. Genet.* **124**, 579–591 (2009).
12. R. D. McCracken, Lactase deficiency: An example of dietary evolution. *Curr. Anthropol.* **12**, 479–517 (1971).
13. L. Ségurel, C. Bon, On the evolution of lactase persistence in humans. *Annu. Rev. Genomics Hum. Genet.* **18**, 297–319 (2017).
14. T. Sahi, M. Isokoski, J. Jussila, K. Launiala, K. Pyörälä, Recessive inheritance of adult-type lactose malabsorption. *Lancet* **302**, 823–826 (1973).
15. N. S. Enattah *et al.*, Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30**, 233–237 (2002).
16. P. Ridefelt, L. D. Håkansson, Lactose intolerance: Lactose tolerance test versus genotyping. *Scand. J. Gastroenterol.* **40**, 822–826 (2005).
17. E. Heyer *et al.*, Lactase persistence in Central Asia: Phenotype, genotype, and evolution. *Hum. Biol.* **83**, 379–392 (2011).
18. I. Gallego Romero *et al.*, Herders of Indian and European cattle share their predominant allele for lactase persistence. *Mol. Biol. Evol.* **29**, 249–260 (2012).
19. N. S. Enattah *et al.*, Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am. J. Hum. Genet.* **82**, 57–72 (2008).
20. C. J. E. Ingram *et al.*, A novel polymorphism associated with lactose tolerance in Africa: Multiple causes for lactase persistence? *Hum. Genetics* **120**, 779–788 (2007).
21. S. Torniainen *et al.*, Screening of variants for lactase persistence/non-persistence in populations from South Africa and Ghana. *BMC Genet.* **10**, 31 (2009).
22. G. Breton *et al.*, Lactase persistence alleles reveal partial East African ancestry of southern African khoe pastoralists. *Curr. Biol.* **24**, 852–858 (2014).
23. R. P. Evershed *et al.*, Dairying, diseases and the evolution of lactase persistence in Europe. *Nature* **608**, 336–345 (2022).
24. D. E. MacHugh, G. Larson, L. Orlando, Taming the past: Ancient DNA and the study of animal domestication. *Annu. Rev. Anim. Biosci.* **5**, 329–351 (2017).
25. J. Burger *et al.*, Low prevalence of lactase persistence in Bronze Age Europe indicates ongoing strong selection over the last 3,000 years. *Curr. Biol.* **30**, 4307–4315.e4313 (2020).
26. E. K. Irving-Pease *et al.*, The selection landscape and genetic legacy of ancient Eurasians. *Nature* **625**, 312–320 (2024).
27. 1000 Genomes Project Consortium *et al.*, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
28. L. Phan *et al.*, *ALFA: Allele Frequency Aggregator* (U.S. National Library of Medicine, National Center for Biotechnology Information, 2020), https://www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/.
29. A. Bergström *et al.*, Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).
30. B. S. Weir, C. C. Cockerham, Estimating F-STATISTICS for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
31. M. D. Shriver *et al.*, The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum. Genomics* **1**, 274–286 (2004).
32. P. C. Sabeti *et al.*, Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
33. K. Prüfer *et al.*, The complete genome sequence of a Neanderthal from the Altai mountains. *Nature* **505**, 43–49 (2014).
34. M. Meyer *et al.*, A high-coverage genome sequence from an Archaic Denisovan individual. *Science* **338**, 222–226 (2012).
35. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
36. K. Prüfer *et al.*, A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* **358**, 655–658 (2017).
37. N. Saitou, M. Nei, The neighbor-joining method–A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
38. S. Sankararaman *et al.*, The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).
39. B. Vernot, J. M. Akey, Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**, 1017–1021 (2014).
40. S. R. Browning, 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).
41. L. Chen, A. B. Wolf, W. Fu, L. Li, J. M. Akey, Identifying and interpreting apparent Neanderthal ancestry in African individuals. *Cell* **180**, 677–687.e616 (2020).
42. K. Yuan *et al.*, Refining models of archaic admixture in Eurasia with ArchaicSeeker 2.0. *Nat. Commun.* **12**, 6232 (2021).
43. L. B. Barreiro *et al.*, Evolutionary dynamics of human toll-like receptors and their different contributions to host defense. *PLoS Genet.* **5**, e1000562 (2009).
44. B. F. Voight, S. Kudaravalli, X. Q. Wen, J. K. Pritchard, A map of recent positive selection in the human genome. *PLoS Biol.* **4**, 446–458 (2006).
45. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
46. J. Smith, G. Coop, M. Stephens, J. Novembre, Estimating time to the common ancestor for a beneficial allele. *Mol. Biol. Evol.* **35**, 1003–1017 (2018).
47. S. Mallick *et al.*, The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Sci. Data* **11**, 182 (2024).
48. X. Mao *et al.*, The deep population history of northern East Asia from the Late Pleistocene to the Holocene. *Cell* **184**, 3256–3266.e3213 (2021).
49. M. Sikora *et al.*, The population history of northeastern Siberia since the Pleistocene. *Nature* **570**, 182–188 (2019).
50. J. V. Moreno-Mayar *et al.*, Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* **553**, 203–207 (2018).
51. M. Rasmussen *et al.*, The ancestry and affiliations of Kennewick Man. *Nature* **523**, 455–458 (2015).
52. L. Speidel, M. Forest, S. Shi, S. R. Myers, A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019).
53. A. J. Stern, P. R. Wilton, R. Nielsen, An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. *PLoS Genet.* **15**, e1008384 (2019).
54. M. N. Leseva *et al.*, Differences in DNA methylation and functional expression in lactase persistent and non-persistent individuals. *Sci. Rep.* **8**, 5649 (2018).
55. V. Labrie *et al.*, Lactase nonpersistence is directed by DNA-variation-dependent epigenetic aging. *Nat. Struct. Mol. Biol.* **23**, 566–573 (2016).
56. M. Ota *et al.*, Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* **184**, 3006–3021.e3017 (2021).
57. The GTEx Consortium *et al.*, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
58. Q. Peng *et al.*, Analysis of blood methylation quantitative trait loci in East Asians reveals ancestry-specific impacts on complex traits. *Nat. Genet.* **56**, 846–860 (2024).
59. Z. Ning *et al.*, Expression profiles of East–West highly differentiated genes in Uyghur genomes. *Nat. Sci. Rev.* **10**, nwad077 (2023).
60. D. Welter *et al.*, The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
61. M. Kanai *et al.*, Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
62. M.-H. Chen *et al.*, Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* **182**, 1198–1213.e1114 (2020).
63. S. Sakaue *et al.*, A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
64. M.-S. Peng *et al.*, Lactase persistence may have an independent origin in Tibetan populations from Tibet, China. *J. Hum. Genet.* **57**, 394–397 (2012).
65. D. S. Lu *et al.*, Ancestral origins and genetic history of Tibetan highlanders. *Am. J. Hum. Genet.* **99**, 580–594 (2016).
66. M.-S. Peng *et al.*, Genetic and cultural adaptations underlie the establishment of dairy pastoralism in the Tibetan Plateau. *BMC Biol.* **21**, 208 (2023).
67. Y. Field *et al.*, Detection of human adaptation during the past 2000 years. *Science* **354**, 760–764 (2016).
68. R. M. Gittelman *et al.*, Archaic hominin admixture facilitated adaptation to out-of-Africa environments. *Curr. Biol.* **26**, 3375–3382 (2016).
69. F. Racimo *et al.*, Archaic adaptive introgression in TBX15/WARS2. *Mol. Biol. Evol.* **34**, 509–524 (2017).
70. X. Ma, S. Xu, Archaic introgression contributed to the pre-agriculture adaptation of vitamin B1 metabolism in East Asia. *iScience* **25**, 105614 (2022).
71. J. T. Troelsen, J. Olsen, J. Møller, H. SjÖstrÖm, An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology* **125**, 1686–1694 (2003).
72. L. Wang *et al.*, A microRNA linking human positive selection and metabolic disorders. *Cell* **183**, 684 (2020).
73. A. Nie, B. Sun, Z. Fu, D. Yu, Roles of aminoacyl-tRNA synthetases in immune regulation and immune diseases. *Cell Death Dis.* **10**, 901 (2019).
74. C. Zheng *et al.*, Multiomics analyses reveal DARS1-AS1/YBX1–controlled posttranscriptional circuits promoting glioblastoma tumorigenesis/radioresistance. *Sci. Adv.* **9**, eadf3984 (2023).
75. Q. Dang, LncRNA DARS-AS1 in human cancers: A comprehensive review of its potency as a biomarker and therapeutic target. *Gene* **923**, 148566 (2024).
76. B. Liu *et al.*, Functional genomics screening identifies aspartyl-tRNA synthetase as a novel prognostic marker and a therapeutic target for gastric cancers. *J. Pathol.* **258**, 106–120 (2022).
77. H. Xiong *et al.*, DARS expression in BCR/ABL1-negative myeloproliferative neoplasms and its association with the immune microenvironment. *Sci. Rep.* **14**, 16711 (2024).
78. L. Tan *et al.*, NIFK as a potential prognostic biomarker in colorectal cancer correlating with immune infiltrates. *Medicine* **102**, e35452 (2023).
79. M. Byrska-Bishop *et al.*, High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426–3440.e3419 (2022).
80. F. Mafessoni *et al.*, A high-coverage Neandertal genome from Chagyrskaya Cave. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 15132–15136 (2020).
81. M. Gautier, R. Vitalis, rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176–1177 (2012).
82. B. Paten *et al.*, Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.* **18**, 1829–1843 (2008).
83. R. Zhang, K. Yuan, S. Xu, Detecting Archaic introgression and modeling multiple-wave admixture with ArchaicSeeker 2.0. *STAR Protoc.* **3**, 101314 (2022).
84. J. W. Leigh, D. Bryant, POPART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).
85. H. J. Bandelt, P. Forster, A. Rohl, Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).

86. R. H. Waterson, E. S. Lander, R. K. Wilson, The Chimpanzee Sequencing and Analysis Consortium, Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
87. V. Lefort, R. Desper, O. Gascuel, FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* **32**, 2798–2800 (2015).
88. K. Tamura, M. Nei, Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526 (1993).
89. G. C. Yu, D. K. Smith, H. C. 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).
90. International HapMap Consortium *et al.*, A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–853 (2007).
91. W. McLaren *et al.*, The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
92. D. R. Zerbino, S. P. Wilder, N. Johnson, T. Juettemann, P. R. Flicek, The ensembl regulatory build. *Genome Biol.* **16**, 56 (2015).