Genetic dietary adaptation in Neandertal, Denisovan and Sapiens revealed by gene copy number variation

Riccardo Vicedomini[1]†, Niccolò Righetti[1]†, Lélia Polit[1]†, Silvana Condemi[2], Laura Longo[3]*, Alessandra Carbone[1]*

[1] Sorbonne Université, CNRS, IBPS, UMR7238, Laboratoire de Biologie Computationnelle et Quantitative (LCQB), Paris, France.
[2] Aix-Marseille Univ, CNRS, EFS, UMR7268, Laboratoire Anthropologie, Droit, Ethique et Santé (ADES), Marseille, France.
[3] Dipartimento di Scienze Ambientali, Informatica e Statistica (DAIS) - University Ca' Foscari, Venice, Italy.

* Corresponding authors: alessandra.carbone@sorbonne-universite.fr, Laura.Longo@unive.it
† These authors contributed equally to this work.

Abstract:

Dietary adaptation involves evolving an efficient system to digest food available in an ecosystem. The diet of archaic humans is traditionally reconstructed by isotopic analyses of human remains combined with the faunal assemblages found on the sites, and, recently, from metagenomic analyses of dental calculus. Here, we propose a new computational approach to find the genetic basis for human dietary adaptation. We searched 15 genomes from Neandertal, Denisovan and Early Sapiens for food digestion genes that tend to have more or fewer copies than the modern human reference genome. We identify 50 genes, including 10 gene clusters, with discernible copy number variation (CNV) trends at the population level, from an analysis of the full set of 20,000 human genes. The genomic variation of 19 of these genes shows how metabolic pathways for carbohydrates, lipids, liver lipids and brown fat in archaic humans adapted to metabolize food from animal or plant sources. The remaining 31 genes are all highly expressed in tissues of the digestive apparatus and are involved in immune response, environmental response and obesity. Analysis of the CNV profiles, compared to 64 modern human individuals belonging to distinct ethnic groups in Eurasia, Africa, Oceania, suggests that *Homo sapiens* may have had an evolutionary advantage compared to Neandertal and Denisovan in adapting to cold and temperate ecosystems.

Significance Statement:

Understanding dietary strategies and foraging behaviors of past populations are among the goals of paleoanthropological and prehistoric studies. Based on a new computational approach, we seek for the genetic basis for human dietary adaptation. Gene copy number variation (CNV) is a major type of structural genome variation that we used as indicator of efficient metabolic processes for food digestion. By analysing billions of short sequences from 15 archaic human genomes and 64 modern human ones, across the full set of 20,000 human genes, we identify 50 genes whose population-wide discernable CNV trends point to lipid metabolism as being crucial for Neandertal, the efficiency in carbohydrate

metabolism for Sapiens' diet, and the importance of lipid metabolism and brown fat metabolism for both Neandertal and early Sapiens compared to modern humans.

Short title: Dietary gene CNV in archaic humans

Preprint availability: an early version, where the CNV analysis of ancient genomes was conducted on the human reference genome hg19, was deposited on BioRxiv on November 2, 2021 (DOI: 10.1101/2021.10.30.466563). The current version, which utilises CNV estimates based on the human reference genome hg38 and includes comparisons with modern human populations, was deposited in BioRxiv in November 2024.

Introduction

After the great dispersion of *Homo sapiens* (*Hs*) from Africa into Eurasia about 60,000 years ago (1), the local Eurasian archaic humans (EAHs), notably the Neandertals (*Hn*) and the Denisovans (*Den*) who populated this geographical region, disappeared. Reasons for the success of *Hs* are debated among specialists, with food intake being one factor to be considered (2,3).

According to the pioneering study of the *AMY1* gene coding for salivary amylase (4,5), high starch diets became staple during the Neolithic, when people became major consumers of cereals and other starch-rich plants, tubers and rhizomes. The inclusion of plants in the diet of *Hs* was demonstrated to provide not only a substantial caloric intake but also essential fatty acids (FAs), proteins, and other vital micronutrients (6–10). By contrast, the diet and ecological niche of EAHs were indicators of low versus high trophic level carnivores (11–14). More recent studies have highlighed that the diet of EAHs is more complex than what was thought for many years and that dietary carbohydrates consumption may have not being totally absent in their diet (15–18).

In this study, we consider differences in diet between EAHs and *Hs* by studying the copy number variation (CNV) of genes involved in digestion and metabolism. The aim of our novel computational approach is to use publicly available complete genomes of archaic humans to identify genomic changes underlying the diverse and evolving ability of humans to efficiently digest and metabolize energetic nutrients. Our CNV data show that humans living in temperate and cold latitudes had access to differentiated food sources. While both *Hn* and *Den* maintained highly carnivory, *Hs* has increased intake of dietary carbohydrates since its oldest occurrence in Eurasia, thus accessing this caloric food long before crop domestication emerged (9,19). This novel information can be interpreted in light of the archaeological and anthropological records of human populations living across Eurasia during the Late Pleistocene.

Genomic CNV has been studied for more than 30 years (20). Initially, it was thought that CNV was rare with a limited impact on the total extent of human genetic variation (21). With recent genomic technologies, thousands of heritable copy number variants (CNVs) within modern populations have been documented, generating considerable interest over the functional significance of gene duplication (21–23). CNV has been demonstrated to influence levels of gene transcription (19,24–26) conferring an adaptive advantage, and some CNVs have been associated with differential susceptibility to complex diseases (27,28). The landscape of past and present CNV in genomes could be of functional significance related to specific phenotypes within and across populations. Only a few genomes of human archaic populations will ever be available, but our CNV analysis shows that genes that maintain their copy number across individuals of a same population while changing copy number across populations can be indicators of adaptation to different environments.

Results

More than 10,000 ancient nuclear genomes of individuals younger than 20,000 years, and only slightly more than 50 genomes from individuals older than 20,000 years, are available to date in the AADR v54.1 reference repository (29), with additional genomes also reported in the broader literature. Of those, we could only estimate gene CNV for 15 of them due to insufficient sequencing coverage, the elimination of repeated regions, and contamination of most sequenced genomes (see **Materials and Methods**). Still, this is the largest pool of ancient nuclear genomes analyzed in this way (**Figure 1A** and **Table S1**). In

this pool are key genomes from individuals at sites spanning temperate and cold latitudes of Eurasia (**Table S2**) including EAHs from the Altai Mountain - Den D3 (30), Altai Neandertal D5 (31) and Chagyrskaya 8 (32) – and the earliest *Hs* from Ust'Ishim (North Siberia) (33). To this Central Asian core sample, we added *Hn* genomes following their east-west distribution, namely Mezmaiskaya 1 (34), Mezmaiskaya 2 (35), Vindija 33.19 (also referred to as Vindija 87) (34), Goyet Q56.1, Spy 94a, and Les Cottés Z4-1514 (35). In addition to Ust'Ishim, modern human genomes belonging to Peştera Muierii 1 (36), Mal'ta (37), to the hunter-gatherers from Motala 12 and Loschbour, and to the earliest northern farmer from Stuttgart (38,39) were analyzed (see **Supplementary Text 1**).

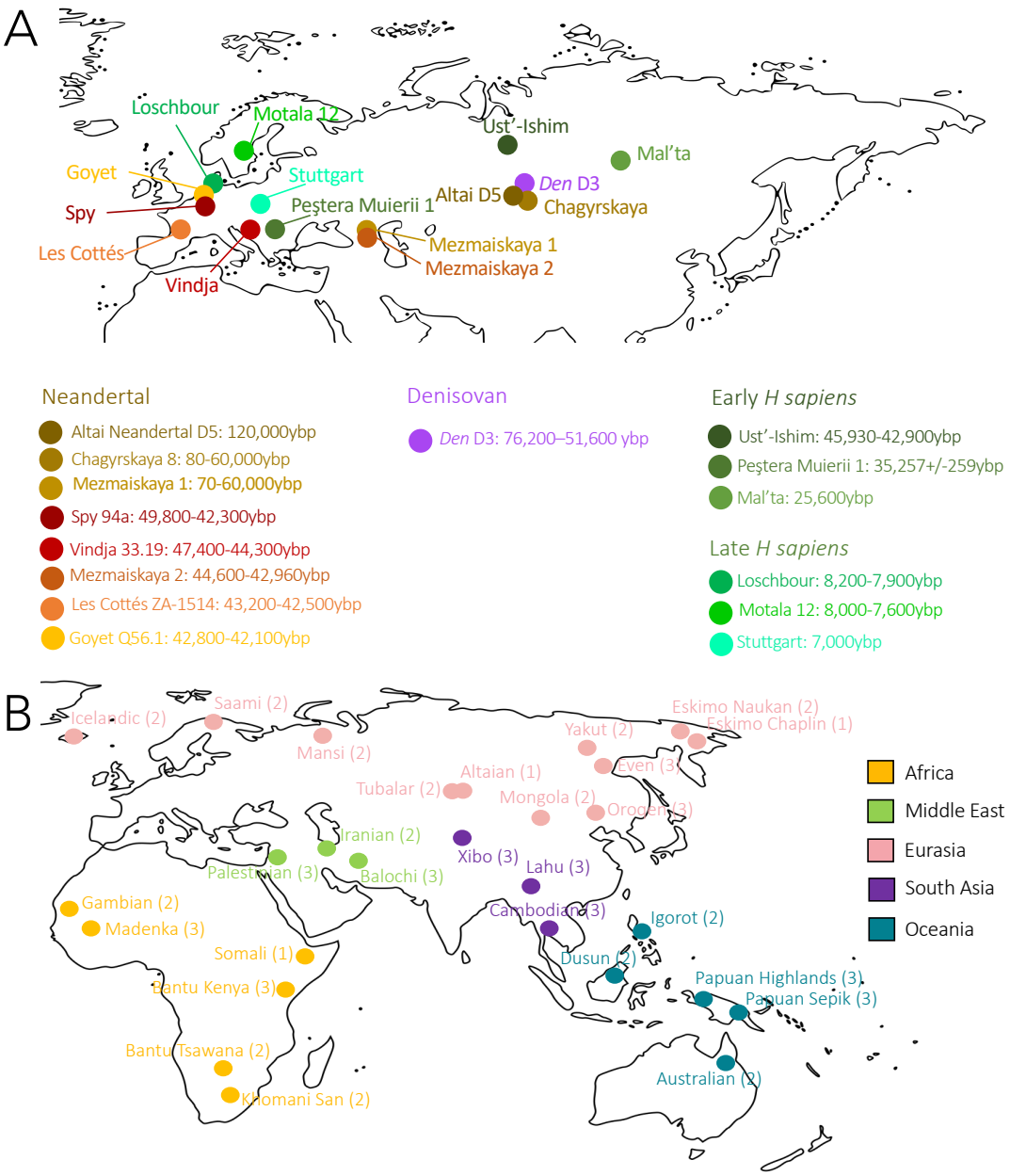
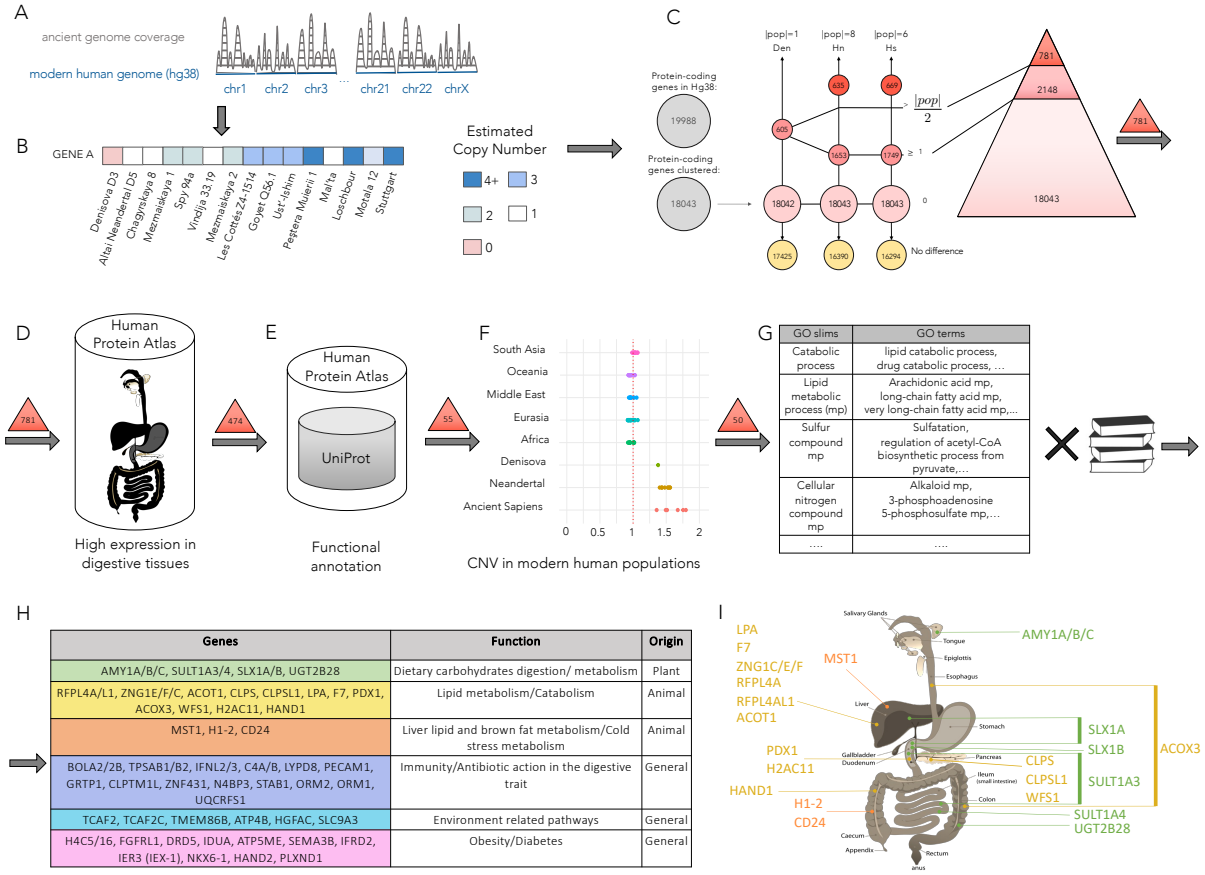


Figure 1. Geographical location of ancient and modern genomes used for gene CNV analysis. A. Ancient genomes are from individuals of early and late *Hs* (greens), *Hn* (browns, reds) and *Den* (purple) populations. Darker shades correspond to older time periods. Dates are calibrated. B. Modern genomes are from individuals from 28 ethnic groups of various geographical origins: Africa (yellow), the Middle East (green), Eurasia (pink), South Asia (purple), and Oceania (cyan). The number of individuals from each group is indicated in parentheses.

Building on this extensive analysis of ancient genomes from key Eurasian sites, we extended our comparison to include modern human populations from a wide geographical range. We analyzed 64 genomes from modern individuals across Africa, Eurasia, the Middle East, South Asia, and Oceania, representing 28 diverse ethnic groups with varying dietary practices (**Figure 1B** and **Table S3**) (40,41). This provides a comprehensive framework for understanding genetic variation across regions. This broader comparison is crucial for properly evaluating gene adaptation, as understanding how CNVs in ancient populations compare to those in present-day human groups enables us to more accurately assess evolutionary changes.



GO slims	GO terms
Catabolic process	lipid catabolic process, drug catabolic process, …
Lipid metabolic process (mp)	Arachidonic acid mp, long-chain fatty acid mp, very long-chain fatty acid mp,…
Sulfur compound mp	Sulfatation, regulation of acetyl-CoA biosynthetic process from pyruvate,…
Cellular nitrogen compound mp	Alkaloid mp, 3-phosphoadenosine 5-phosphosulfate mp,…
….	….

Genes	Function	Origin
AMY1A/B/C, SULT1A3/4, SLX1A/B, UGT2B28	Dietary carbohydrates digestion/ metabolism	Plant
RFPL4A/L1, ZNG1E/F/C, ACOT1, CLPS, CLPSL1, LPA, F7, PDX1, ACOX3, WFS1, H2AC11, HAND1	Lipid metabolism/Catabolism	Animal
MST1, H1-2, CD24	Liver lipid and brown fat metabolism/Cold stress metabolism	Animal
BOLA2/2B, TPSAB1/B2, IFNL2/3, C4A/B, LYPD8, PECAM1, GRTP1, CLPTM1L, ZNF431, N4BP3, STAB1, ORM2, ORM1, UQCRFS1	Immunity/Antibiotic action in the digestive trait	General
TCAF2, TCAF2C, TMEM86B, ATP4B, HGFAC, SLC9A3	Environment related pathways	General
H4C5/16, FGFRL1, DRD5, IDUA, ATP5ME, SEMA3B, IFRD2, IER3 (IEX-1), NKX6-1, HAND2, PLXND1	Obesity/Diabetes	General

Figure 2: A schematic overview of the unsupervised approach identifying 19 strictly diet-related genes involved in three overarching metabolic functions and 31 accessory diet-related genes. Our automatic pipeline is defined by six main steps illustrated in panels A-F (see Methods and **Supplementary Text 2** for details). **A.** Coverage estimation of chromosomes and genes in ancient genomes after read mapping on the modern human reference genome hg38 comprising 19,988 genes, or 18,043 genes if gene clusters are counted instead. **B.** Gene haploid copy number estimation realized for the 15 ancient genomes provides CNV profiles for 18,043 genes in *Hn* and *Hs* and 18,042 in *Den* (see Methods). **C.** For *Den*, *Hn* and *Hs*, the different pink shaded circles detail the number of genes displaying a differential copy number for >n individuals in the population, where n=0…|pop| and |pop| indicates the total number of individuals in a population; for instance, over the total of 18,043 genes, 1,749 genes in *Hs* display a CNV profile different than hg38 on at least one individual, and 669 genes display a CNV profile greater or smaller than hg38 on at least 4 *Hs* individuals for a total of 6. Based on this analysis, the 18,043 clustered human protein genes (grey circle) are split in three subgroups (triangle): 18,043 genes are analysed across the 15 ancient genomes and hg38 (light pink, n=0), 16,294 (16,390) of which display no differential CNV in *Hs* (*Hn*) genomes with hg38 (yellow); 1,749 (1,653) genes have at least one individual across the *Hs* (*Hn*) population with CNV different from hg38 (pink, 1≤n≤|pop|/2); 781 genes have at least one population which displays more than the half of its individuals with a greater or smaller copy number (red, n>|pop|/2). **D**. From the 781 genes selected in **C**, we screened those with primary expression is digestive system organs using data from the Human Protein Atlas (HPA) database (**Table S4**). **E.** These genes were further filtered through a manual review of UniProt functional annotation reported in HPA to confirm their role in digestion, resulting in a set of 55 genes. **F.** These 55 genes were then compared to CNV estimations from five populations across Africa, Eurasia, South Asia, Middle East and Oceania to assess adaptive trends, narrowing the set to 50 genes. **G**. A GO term and literature-based analysis of these 50 genes identified their metabolic functions and specific roles in digesting plant or animal foods, or in "general" digestion-related processes. The results are shown in **H** (see also **Table S5**). **I.** Of these, 19 strictly diet-related genes are mapped to the digestive organs where they are primarily expressed. [The original drawing of the digestive apparatus was released into the public domain by the author LadyofHats.]

Definition of population-dependent CNVs. To be certain of capturing the genetic changes in a population rather than specifying the genetic background of any individual, we considered genes with a differential copy number in an ancient population. First, gene copy numbers were determined individually for the genomes of *Hs, Hn, Den*, and the modern reference human genome (**Figure 2AB**; see Methods). Then, the *differential copy number of a gene* was defined for each individual as being positive if the genome had more haploid copies than the modern reference genome (GRCh38/hg38 (42)) and negative if it had fewer copies (**Figure 2C**, center). Next, we identified those genes with a positive or negative CNV in more than the half of the individuals of a population (**Figure 2C**, right; see Methods). Practically, this translates as 5 or more of the 8 *Hn* individuals and 4 or more of the 6 *Hs* individuals. Genes not showing a population-dependent CNV were not selected even though they might have been duplicated in one or more genomes within a population. As there is only one *Den* genome, the list of genes with a differential CNV from *Hn* and *Hs* has been supplemented by the CNV estimation from *Den*.

We used an unsupervised approach to analyze the full set of 20,000 human genes for the 15 genomes (**Figure 1A**) to identify a subgroup of 781 genes based on the differential haploid copy number of the gene for a population (**Figure 2ABC**). These 781 genes are expected to highlight population-dependent characteristics of the genomes and constitute the "tip of the iceberg" of the entire collection of human genes. Here, being interested in investigating the human diet, we further filtered them on the basis of two other criteria: their principal expression in organs of the digestive apparatus (**Figure 2DE** and **Table S4**) and a comparison with CNV estimations from five modern human populations to assess their adaptive trends (see **Materials and Methods**; **Figure 2F**). This process led to the identification of 50 genes, which were then analysed using Gene Ontology classification (43) and digestion-related literature (**Figure 2G** and **Table S5**). Of these, 19 genes were involved in three strictly diet-related overarching metabolic functions: carbohydrate, lipid, and liver lipid/brown fat metabolisms. The remaining 31 were associated with auxiliary roles in diet, including functions related to the immune system, cold response, and obesity (**Figure 2HI**). The computational approach is large-scale and allows genome-wide gene analysis while remaining accurate in estimation, as indicated in the discussion and in the detailed analysis of the *AMY1A/B/C* gene cluster, involved in carbohydrate metabolism, compared to *AMY2A* and *AMY2B* genes reported in **Figure S1**. (See **Materials and Methods** and **Supplementary Text 2** for all details.)

The panel of 50 genes is primarily expressed in digestive organs, with most showing expression in a specific organ (**Figure 2G** and **Table S4**), according to data from the Human Protein Atlas (HPA) (44). However, certain genes, such as *ACOT1*, which is involved in fatty acid metabolism and cellular energy balance, and *MST1*, which plays a role in liver lipid and brown fat metabolism, exhibit a much wider expression across multiple tissues. The CNV observed may therefore be an adaption to processes other than digestive activity. We excluded from our analysis any food metabolism genes (45) which are not principally expressed in the digestive apparatus (44), even if they presented a differential CNV in a population. We also excluded those genes that have not yet been associated with a known biological function (46), even though their high expression in digestive organs and differential CNV suggest they might play a role in, or be influenced by, digestion. The stringency of the expression criterion and the functional criterion used here may mean that the current panel of genes is an underestimate of those potentially responsible for CNV-driven food adaption.

For the 50 genes selected with a differential CNV in a population, we checked whether there was a *CNV trend* in the population, that is in 5 or more of the 8 *Hn* individuals and 4 or more of the 6 *Hs* individuals (see Methods). The *AMY1A/B/C* gene cluster (**Table 1A**; **Table S6**), for example, was selected by a negative differential CNV with 7 out of 8 *Hn* individuals having fewer gene copies than modern humans

(**Table 1A**). The *Hs* population does not show a positive CNV trend, even though 4 of the 6 genomes have at least the same number of *AMY1A/B/C* gene copies than the modern genome (**Table 1A**). Determining the differential CNV trend of a gene in a population revealed several aspects. The same gene may show different trends between ancient populations: *AMY1A/B/C* and *SULT13/4*, for instance, have fewer gene copies in *Hn* and *Den*, but more or on par in *Hs*. Trends may be positive and negative in a single population: in *Hn*, the differential CNV of *ACOT1* is negative, while that of *ACOX3* and *HAND1* is positive. The same trend was shared by populations: for example, for *TPSAB1/B2* and *IFNL2/3* show a negative trend in both *Hn* and *Hs*, whereas *CLPS* and *CLPSL1* show a positive trend in both *Hn* and *Hs*. As expected, CNV estimations in *Den* very often agree with the *Hn* trend (47). The differential CNV captured chronological trends within populations, as shown for the *HAND1* gene in *Hs*, where duplication appears in Mesolithic (late hunter-gatherer) individuals. Another example are the 5 and 6 copies of *AMY1A/B/C* observed in Loschbour and Stuttgart (Neolithic, early farmers) genomes (**Table 1B**), recently highlighted in (48).

Among the 50 genes, we found the well-studied *AMY1A/B/C* gene cluster (4,5,9,10). The negative CNV trend seen in *Hn* and *Den* populations (**Table 1** and **Table S6**) confirms the key role of the cluster in *Hs* adaptation to a different mode of food digestion and metabolism. Some of the remaining 18 genes, such as *ACOT1* (49) and *CLPS* (23), have previously been identified as key players in human dietary adaptation. In this study, we highlight their primary roles and propose an expanded panel of genes for further investigation. The 19 genes cover a broad spectrum of metabolic functions. The most striking feature is that for each overarching metabolic process represented, either all the genes involved in that process present a positive differential CNV for a population or a negative one (see **Table 1A** and **Table S6**, rightmost columns). This is a strong indication that an adaptive force underlies the duplication of these genes in archaic populations. A detailed description of the 50 genes is reported in **Supplementary Text 3.**

It should be highlighted that the dataset of 50 genes has been identified after a comparison with CNV data from modern individuals (**Figures 3** and **Figure S2**). Their CNV was estimated in the same way as for ancient individuals (**Figure 1AB**), and for each gene and each of the five modern human populations, the median CNV distribution (**Tables 1** and **2**) was considered to represent the CNV of the modern population. This comparison revealed that while CNVs in the human reference genome generally align with CNV medians of the modern populations, there are certain populations where they do not. This suggests that the reference human genome alone is insufficient for evaluating CNV variability and that comparison with modern human populations should be integrated into CNV analyses. We excluded five genes showing these differences (**Figure S3ABC**) and identical CNV median for both ancient and modern populations, suggesting no evidence for environmental adaptation for these genes. For instance, the *PGA3/4/5* gene cluster (Pepsinogen A, group 1), involved in the breakdown of animal proteins and highly expressed in the stomach, shows a differential CNV between hg38 and nearly all *Hn* and *Hs* individuals, with more than 4 copies observed as a median in Eurasia, Oceania and South Asia populations, aligning with the high duplication in both *Hn* and *Hs* (**Figure S3CD**). We conclude that this gene is essential for the human diet across both ancient and modern populations; however, its duplication seems unrelated to environmental adaptation in ancient populations.

In **Figure 3A** and **Table 1A** (see **Figure 3C** and **Table 2A** for the 31 accessory diet-related genes), the comparison between CNV distributions between ancient and modern populations reveals three main patterns: 1. Modern populations may align with the CNV estimation from the hg38 reference genome. This is observed in lipid metabolism genes such as *F7, PDX1, WFS1, HCAC11, ACOX3*, and *HAND1*, as well as liver lipid/brown fat metabolism genes *MST1, H1-2, CD24*, and the carbohydrate metabolism

gene *UGT2B28*. In these cases, the CNV median for ancient populations, including *Hn*, *Den* and *Hs,* differs from that of modern populations. 2. Modern populations can exhibit a wide range of CNV values, with their median aligning with hg38, but diverging from the CNV median of at least one ancient population. This pattern is seen in the carbohydrate metabolism gene *AMY1A/B/C* and lipid metabolism genes *ZNG1E/F/C, ACOT1, CLPS*, and *CLPSL1*. 3. In some cases, modern populations show a wide range of CNV values, but their median does not align with the hg38 CNV estimation, while the CNV median in at least one ancient population differs significantly. This is the case for the carbohydrate metabolism genes *SULT1A3/4* and *SLX1A/B* genes, which display lower CNV values in *Hn* compared to modern humans. Similarly, lipid metabolism genes *LPA* and *RFPL4A/L1* show lower and higher CNV medians in *Hs*, respectively, compared to modern populations.

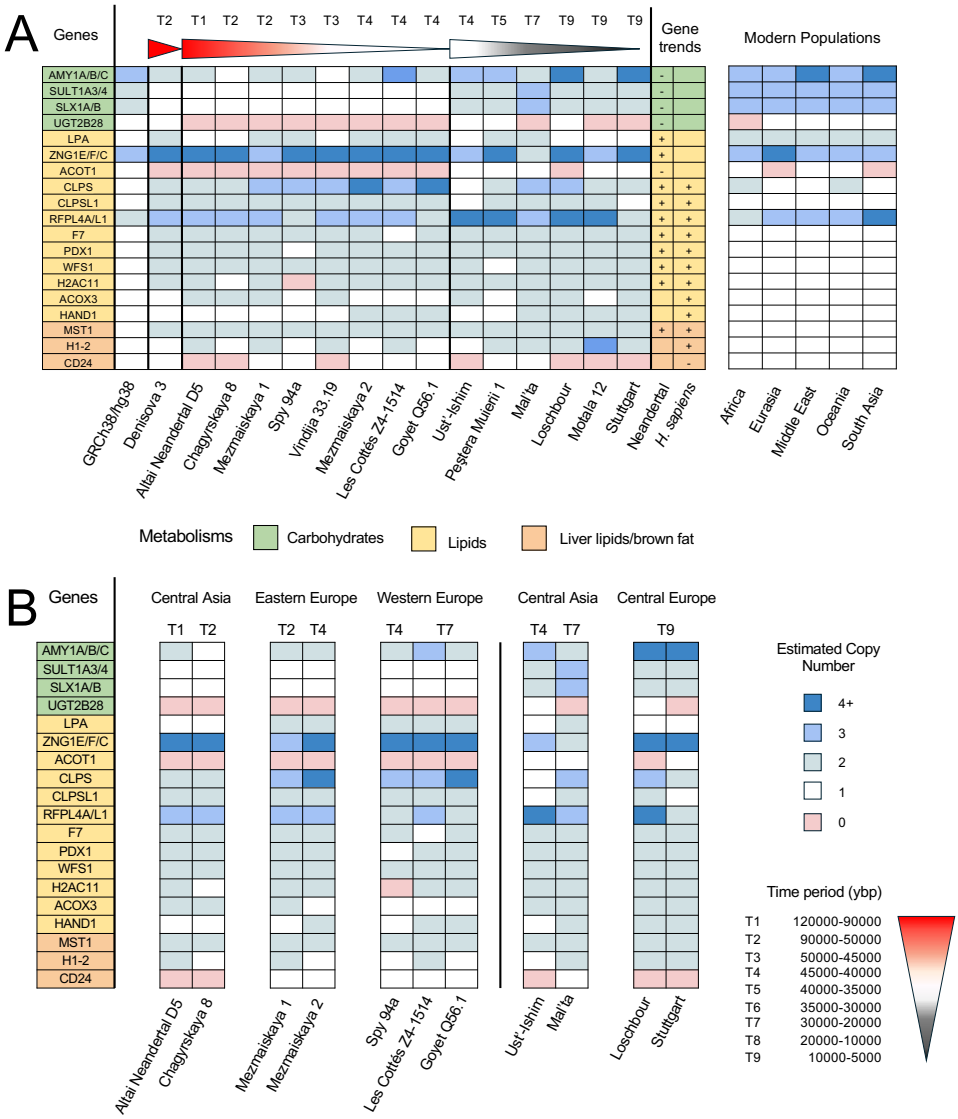


Table 1. Estimated haploid copy number for genes involved in digestion in ancient genomes and differential CNV analysis in ancient and modern populations. A. Copy number of 19 genes (rows) in 15 ancient genomes and the modern human reference (columns). Individuals are grouped according to their population, from left to right, the GRCh38/hg38 reference, *Den, Hn* and *Hs*. Within a population, the order reflects time period and geographical origin. The estimated copy numbers computed for each gene/genome (see **Supplementary Text 2**) are rounded to the closest integer (see **Table S6** for specific values). The order and the color shading for gene names correspond to those used in **Figure 2A**. On the right, the differential CNV trends of each human population are scored for each gene as positive (+), negative (-), not determined (NA) or no differential trend (no symbol). The far right table displays the CNV medians computed for each gene in modern populations across 5 different geographical areas, each representing various ethnic groups (**Table S3** and **Table S7**). Compare to **Figure 3**. **B.** CNV profiles of ancient genomes in **A** grouped by geographical area.

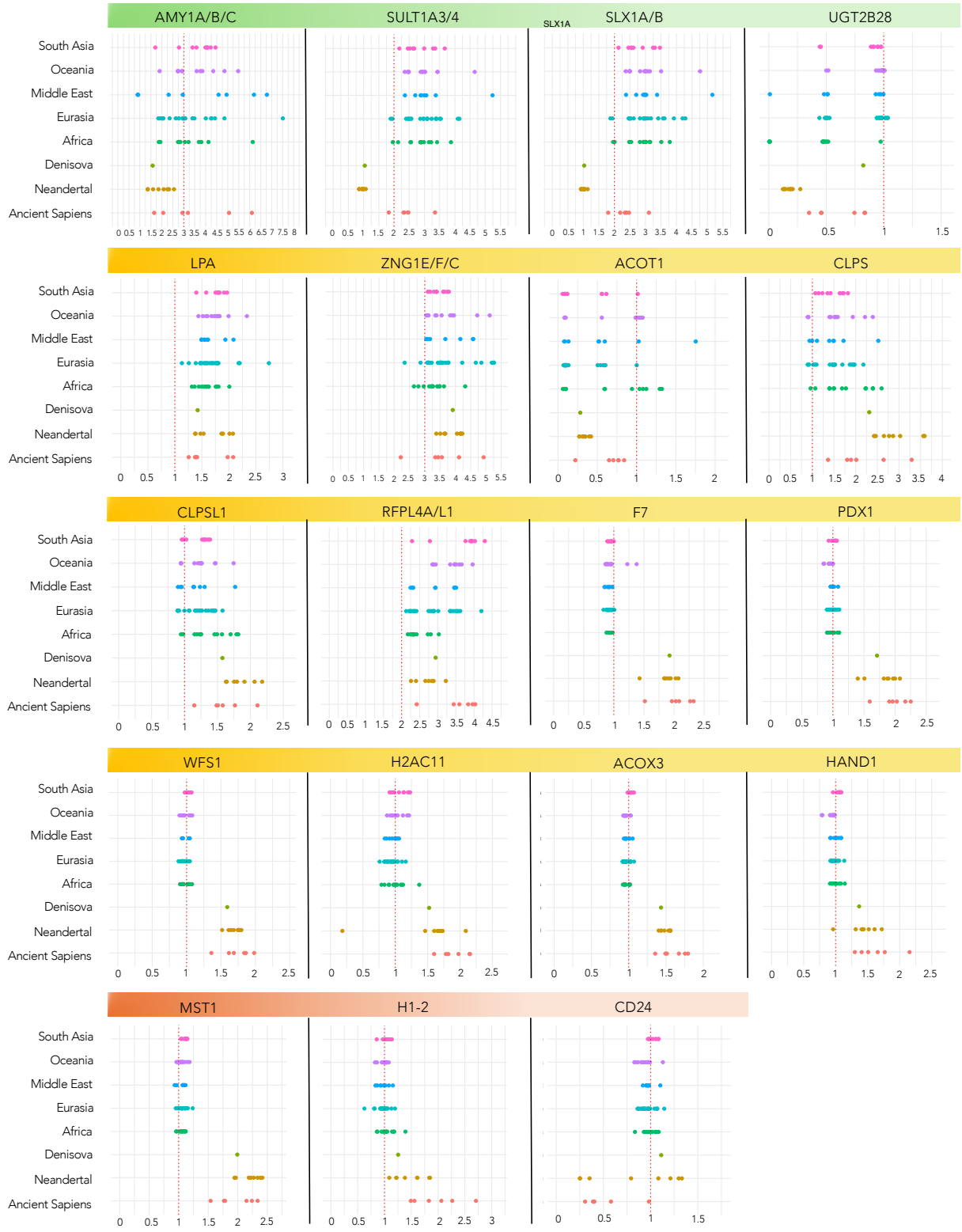


Figure 3. Estimated haploid copy number for the 19 strictly diet-related genes in ancient and modern genomes. For each gene in **Table 1** and for each individual in modern and ancient populations, the estimated CNV is provided. See **Figure S2** for the full set of 50 diet-related genes. The red vertical bar indicates the CNV in the reference genome hg38. Green, yellow and red gene labels correspond to the carbohydrate, lipid, liver lipid/brow fat metabolisms, respectively.

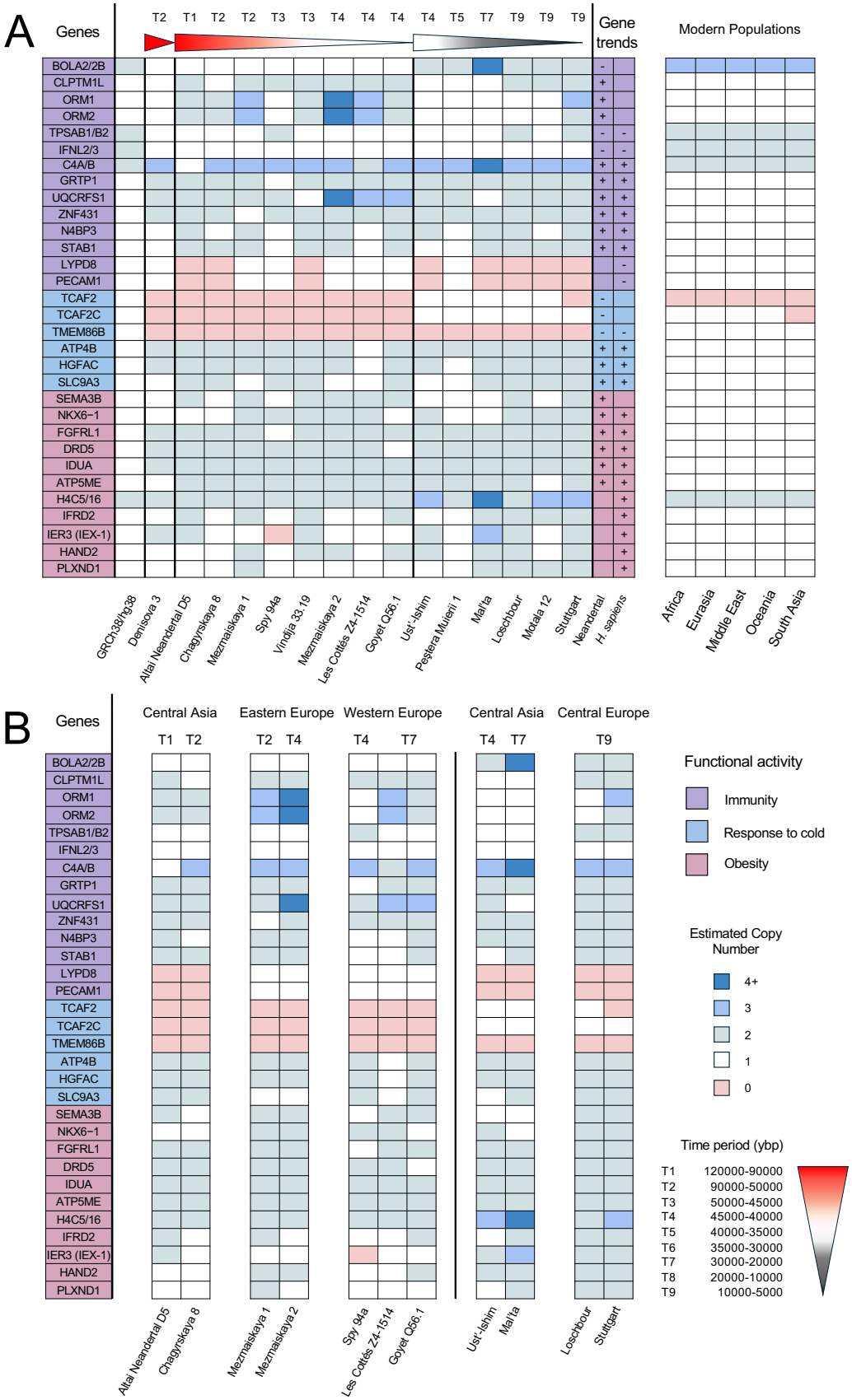


Table 2. Estimated haploid copy number for digestion-related genes in ancient genomes and differential CNV analysis in ancient and modern populations. Description of the 31 accessory genes related to digestion. Legend as in **Table 1**.

A multidimensional genetic space defined by CNV. We further test our main hypothesis that dietary changes directly cause a significant reshaping of individual genomes within populations, and that this phenomenon can be decoded today through large-scale analysis of gene copy number. For this we conducted a PCA analysis, which reveals how ancient and modern genomes can be grouped together in gene space. Based on the CNV of the 19 genes involved in digestion listed in **Table 1A**, we visualized and compared the 15 genomes by plotting them in a 19-dimensional space, where each dimension corresponds to a gene. Formally, we represent a genome of an individual with its CNV profile as described by the relevant column in **Table 1A**. In this multidimensional space, two individuals are close when their CNV profiles are similar. To aid visualization, we reduced the space by principal component analysis (PCA) (50). The PCA projection (**Figure 4A)** guarantees the largest variation in the data points through an optimized linear combination of the 19 original dimensions (see **Supplementary Text 2**). We built two 19-dimensional spaces, with (**Figure S4A**) or without (**Figure 4A**) the modern human hg38 profile included.

In **Figure 4A**, the CNV data clearly separate *Hn* and *Hs* (see the first component PC1 in **Figure 4A**). Den D3 localises close to *Hn*. Classification based on geographical origin (**Figure S5A**) or time period (**Figure S6B**) shows coherent subgroupings. Comparing CNV profiles of the individuals living in the same geographical area (**Table 1B**) illustrates how they are close in the high-dimensional space even when they belong to very different time periods. For instance, Mezmaiskaya 1 and Mezmaiskaya 2 both lived in Eastern Europe (Caucasus) and their similar CNV profiles (**Table 1B**) bring them close together in the CNV space (**Figure S5A**, light blue dots), even though they were not contemporary, living in T2 and T4 respectively.

Analysis of the PCA axes of **Figure 4A** reveals that the most important factors explaining the first dimension (PC1) of the CNV projection separating *Hn* from *Hs*, are genes involved in carbohydrate digestion, *SULT1A3/4* and *SLX1A/B*, and to a minor yet statistically significant extent genes involved in liver lipids metabolism, *H1-2* and *CD24*, and genes involved in lipid metabolism, *ACOX3, PDX1, ACOT1*, and *H2AC11* (**Figure 4B**). The second principal component of the CNV space is mainly accounted for by the gene *UGT2B28* involved in carbohydrate digestion, and the genes involved in lipid metabolism (WFS1, *CLPS, CLPLS1, ACOX3, RFPL4A/L1, HAND1*). Finally, the liver lipid/brown fat metabolism gene *CD24* and the lipid metabolism genes *LPA, CLPSL1, HAND1*, and *CLPS*, make a significant contribution to the third dimension.

In **Figure S4A**, the CNV data clearly separate *Hn, Hs* from the modern genome hg38 (PC1 in **Figure S4AC**) and also separate *Hn* from *Hs* (PC2 in **Figure S4AC**). The first principal component (PC1), separating modern from ancient genomes, is mainly explained by several lipid metabolism genes (*ACOX3, PDX1, WFS1, F7, HAND1, H2AC11, CLPS*), and liver lipid metabolism genes (*MST1* and *H1-2*). The second component (PC2), separating *Hn* and *Hs* genomes, is explained by genes involved in carbohydrate digestion (*SLX1A/B, SULT1A3/4* and UGT2B28). This ensemble of contributions, taken together, underscore the significant dietary shift of modern populations compared to ancient ones, who primarily consumed high-fat and protein-rich animal products (11–15,17,18), whereas modern diets are characterized by a more diverse range of nutrients that include a growing rate of dietary carbohydrates.

The dietary habits of ancient and modern populations are further reflected at the genomic level through a signature based on an independent set of 31 accessory genes. This suggests that the genomic differences between *Hn* and *Hs* related to their diets extend beyond immediate metabolic differences, strongly influencing other biological processes such as immune response, adaptation to cold environments, and obesity. We run our analysis over the complementary set of 31 accessory genes in

Table 2A. These genes are highly expressed in the digestive apparatus. This analysis provided the same genome split for the first and second PCA components (**Figure 4B**) identified by the set of 19 strictly diet-related genes. Interestingly, the separation of *Hn* from *Hs* (PC2; **Figure 4BD**) is obtained with the main contribution of genes which are involved in immunity (*BOLA2/2B*, *ORM2, UQCRFS1, ORM1, PECAM1, C4B*), response to cold (*TCAF2C, TCAF2*), and obesity (*H4C16*, *IER3*). It is noteworthy that the main factors contributing to the splitting between modern humans from ancient humans (PC1 in **Figure S4BD**) are principally involved in obesity (*IDUA, PLXND1, HAND2, FGLR1*) and immune response (*N4BP3, CLPTM1L*).

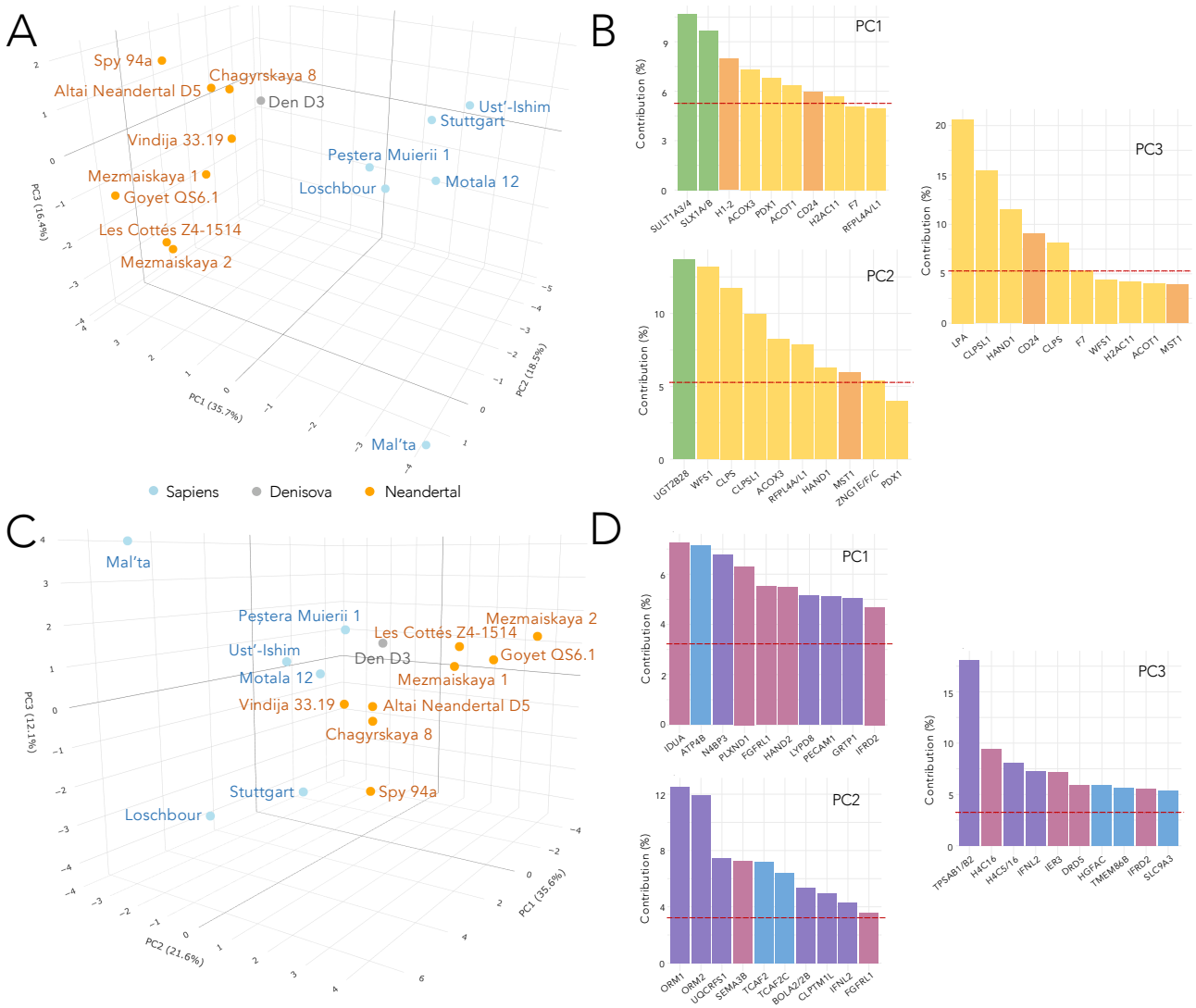


Figure 4. Comparative CNV analysis of *Hn*, *Den* and *Hs* living in the European belt. **A.** Fifteen *Hn*, *Den* and *Hs* individuals are represented in a three-dimensional space obtained by PCA from the 19-dimensional space defined from differential CNV values of the 19 genes involved in metabolic processes (see **Supplementary Text 2**). View of the 3D space where each individual (points) is colored with respect to human populations: *Hn* (orange), *Den* (grey), and *Hs* (light blue). PC1 explains the 35.7% of the variance, PC2 the 18.5% and PC3 the 16.4%. **B.** Contributions of factors for each principal component of the 3D space in **A**. Colored bars indicate the contributions from CNV of genes acting in the three overarching metabolic processes (colors as in **Figure 2F**). **C.** Same analysis as in **A** but based on the differential CNV values of the 31 accessory genes. PC1 explains the 35.6% of the variance, PC2 the 21.6%, and PC3 12.1%. **D.** Contributions of factors for the definition of each principal component of the 3D space, obtained by PCA for the 31 genes analysis in **C**. See legend in **B**.

To summarize, our analysis of the 19 genes primarily involved in digestive functions and displaying a differential copy number provides an unprecedented basis from which to deduce the adaptive changes between EAHs and *Hs* regarding dietary strategies. This study underscores the critical role of three distinct metabolic processes in shaping the behavior of ancestral populations and suggests that these 19 genes could serve as reliable markers for tracking changes in dietary adaptation over time. The CNV genetic trend (**Figure 4A)** points to lipid metabolism as being crucial for both *Hn* and *Hs*, while efficiency in carbohydrate metabolism influences the diet of *Hs*. Lipid and liver lipid/brown fat metabolisms must have been important for both *Hn* and *Hs* compared to modern humans and were probably already present in the common ancestral lineage.

Discussion

In this study, we analyzed 15 complete ancient genomes to identify CNV trends in genes across three human populations present in the vast Eurasian territory during Late Pleistocene— *Hn*, *Den* and early *Hs*. By examining the largest number of publically available full genomes suitable for CNV analysis, our study significantly extends the list of genes with unusual copy number in archaic genomes. By remapping ancient genome reads to the modern human reference genome hg38, we accessed the most up-to-date CNV estimations for comparative analysis. Our genome-wide CNV analysis identified all genes with increased or decreased CNVs in ancient populations relative to the human reference genome, uncovering 781 genes of high interest through an unsupervised search. Among these, we focused on genes involved in food digestion and metabolism, revealing the complexity of nutritional adaptation among the three human populations that shared the same environmental conditions in Eurasia at the end of MIS 4 and during MIS 3. From the genomic data, three overarching metabolic pathways related to diet emerged as overrepresented in the 781-gene dataset: carbohydrate metabolism, lipid metabolism, and liver lipid and brown fat metabolism. These findings provide evidence of distinct adaptive nutritional pathways that may underlie the dietary habits of these populations. Most importantly, the study offers a novel perspective to look at longstanding questions, challenging established results with new genetic evidence. It contributes unique insights into the physiology and biology of past humans that cannot be gleaned from the fossils and archeological records. However, it should be noted that functional complexity, epigenetic regulation, and microbiotic interaction make it challenging to precisely determine the roles of these genes in metabolic activities, even in present-day human populations.

A genome-wide methodological approach based on multiple selective criteria. Distinguishing between adaptive from neutral evolution, especially in hominids which have low effective population sizes and whose available genomes are limited, has been a major challenge (21,23,22,38,49,51,22,52). In this article, we highlight the difficulty in such a selection and propose that a combination of criteria can lead to the identification of a set of candidate genes to play a key role in the dietary adaptation of ancient populations. We search for a pool of genes with differential copy number in ancient populations by analysing, at large-scale, 20 000 human genes with four main selection criteria:

1. a CNV differentiation in ancient populations (not individuals). Focusing on "populations" rather than "individuals" helps identify CNVs that are consistent across a group, implying a broader evolutionary significance of the gene copies. CNV differentiation across ancient populations could highlight long-term selection pressures that shaped the genetic landscape over time. The intuition being that CNVs that show marked differentiation in ancient populations might reflect early environmental pressures (e.g., climate, diet, or pathogens) that shaped genetic adaptations crucial for survival.
2. a comparison of modern populations to ancient ones, to track how certain CNVs have persisted, expanded, or diminished over time. This comparison helps in distinguishing adaptive CNVs that

were beneficial in the past and might still be advantageous in some modern environments, or those that were selected against and have decreased. CNVs that have been maintained or show a selective sweep in modern populations likely reflect ongoing adaptive advantages, such as responses to new environmental factors (e.g., agriculture, modern diets, urbanization, or exposure to new pathogens).

3. the high expression of the gene in specific tissues from digestive organs. Tissue-specific high-expression of genes, especially in critical systems like digestive organs, suggests a key functional role. If the CNVs influence gene expression in tissues tied to nutrient absorption, metabolism, or immune responses in the gut, this could indicate an adaptive response to dietary changes, such as the ability to process different food sources (e.g., starch, lactose), or microbial challenges, to fend off gut pathogens.
4. the functional annotation of the gene. Functional annotation helps clarify the biological role of the genes affected by CNVs. By linking CNVs to specific functions (e.g., immune response, metabolism, etc.), we can pinpoint whether these genetic changes have played a role in specific traits. For example, if a CNV is associated with genes involved in immune function, this might suggest adaptation to pathogen pressures, which is a key driver of human evolution.

The combination of these criteria strengthens the argument that CNVs can serve as markers for human adaptation. Note that the criteria for gene selection under these multiple markers were very strict and allowed to identify 50 genes related to diet which likely contributed to adaptive changes of ancient populations. The stringency of the criteria may also mean that the current panel of genes is an underestimate of those potentially responsible for CNV-driven food adaption. Our 50 genes should be considered as the core set for digestion.

High read coverage and CNV estimation in ancient genomes. Our computational approach underscores the critical need of sequencing complete ancient genomes at high coverage, as emphasized by previous studies (48,53). Read coverage analysis of the 50 diet-related genes (**Figures S7 and S8**) demonstrates that while CNV estimation is feasible, mapping sensitivity is highly influenced by coverage depth. High-coverage genomes, including Den D3, Altai Neandertal D5, Chagyrskaya 8, Ust'Ishim, Peştera Muierii, Vindija, Loschbour, and Stuttgart (**Table S1**) exhibit significantly reduced mapping fluctuations compared to lower-coverage genomes, as illustrated by the smoother curves in **Figure S7** for the CLPS gene and **Figure S8** for the AMY1A/B/C cluster. Moreover, sequencing entire genomes, rather than focusing solely on SNPs or regions, enables to uncover significant aspects of human genome evolution, such as gene copy number variations and mutations in coding or regulatory regions—some of which may not be observable in modern populations. Comprehensive sequencing is thus more essential than ever for advancing our understanding of human evolutionary history.

Another key aspect of our approach is the estimation of the haploid CNV values, which are rounded to the nearest integer and used as indicators of duplication. These estimations are provided and can support further analysis of CNV. A key question arises regarding the method's ability to distinguish between closely related but distinct gene copies. For instance, the highly similar *AMY1* and *AMY2* genes can be correctly differentiated using our haploid genome approach. **Figure S1A** demonstrates our method's estimation of CNV for *AMY1A/B/C, AMY2A,* and *AMY2B* in both modern and ancient populations. The amylase locus, known for its genotyping challenges, underscores the importance of verifying CNV accuracy, particularly for distinguishing *AMY1* from *AMY2* in diploid genomes. Our results align with previous studies (54), showing that modern populations consistently have one copy each of *AMY2A* and *AMY2B*, while *AMY1A/B/C* is typically duplicated with a median of three copies. Further supporting the accuracy of our haploid estimations, the distribution of paired and unpaired copy

numbers in corresponding diploid CNV analyses agrees with earlier studies on modern genomes, as illustrated in **Figure S1B**.

Comparison to modern populations. Another critical aspect highlighted in this study is the importance of the comparison with modern humans (23,55), which helps us better assess whether gene duplication truly represents adaptation, as argued above, by taking into consideration the *PGA3/4/5* gene. Additional information, such as whether a gene is located in copy number variation hotspots across multiple primates, is also crucial for evaluating the gene's likelihood of being duplicated (56). This was the case in (57), which explored the duplication of the *PGA3/4/5* gene. In our analysis, we excluded this gene because its duplication in modern populations—facilitated by its location in a hotspot—does not distinguish ancient genomes from modern ones (**Figure S3**). Our large-scale approach, comparing modern and ancient genomes, indirectly captures aspects at the genome organisation level.

On the generality of the method. While diet plays a significant yet partial role in the evolutionary history of ancient populations, our computational approach is broad and applicable across the entire set of 20,000 genes in our 15 ancient genomes. It provides a foundation for future research into identifying differential CNVs for other functional domains, such as cognition, morphology, sensory perception, or immune response, although further development will be needed to fully explore this potential.

Unsupervised identification of genes presenting common CNV trends. The 19 strongly diet-related genes identified through our unsupervised approach are involved in three key metabolic processes: carbohydrate metabolism, lipid metabolism, and liver lipid/brown fat metabolism. Each metabolic group includes genes with either positive or negative CNV trends within a population. In our view, this is a strong indication that an adaptive force acted on the duplication of these genes in archaic populations, shaping their metabolic capacities with consequences on dietary habits and adaptation to diverse environmental conditions. Our data also indicate that gene copy numbers can differ between earlier and later individuals within the same population, as seen with *AMY1A/B/C* in *Hn* genomes and *LPA* in *Hs* genomes. These chronological trends can be clarified as more high-coverage genomes become available to this analysis. The 31 accessory diet-related genes fall into three main functional categories: immune response, thermoregulation, and obesity management. Many of these genes show higher or lower copy numbers in both *Hn* and *Hs* compared to modern humans, indicating that different levels of gene expression were likely required in these population.

Two independent classifications lead to the same signal of CNV trends. The clear separation of *Hn* individuals from *Hs* individuals in the multidimensional spaces defined by our two independent datasets of diet-related genes—categorized as "strictly" and "accessory" diet-related genes—strongly suggests that dietary preferences and adaptations are encoded at the genomic level, even when considering small gene sets comprising only a few dozen genes. Notably, both gene sets also distinguish ancient populations from the modern reference genome, which occupies a separate region in the multidimensional spaces, suggesting a significant shift in dietary patterns in modern humans. Importantly, our computational approach opens new perspectives on the possibility to reconstruct the evolution of the human diet by analysing genomic data over the past 50,000 year, extending evidence of the inclusion of starchy foods in *Hs'* diet well before crops domestication. This aligns with the view that the transition from hunting-gathering to domestication was occurring across protracted conditions, shaped by ecological and behavioural factors that left their mark at the population level rather than on individuals.

Genomically distinct metabolisms in ancient human populations. Due to limited number of available genomes where the estimation of CNV was possible (20), we mainly focused on *Hn* and *Hs*, with inferences for *Den* where possible. The different CNV trends observed for *Hn* and *Hs* reflect that these hominins relied on different diets, and align well with their evolutionary histories. The genes involved in efficient digestion and **metabolism of dietary carbohydrates** (*AMY1A/B/C, SULT1A3/4, SLX1A* and *UGT2B28*) appear important for *Hs* whereas most genes involved in **lipid metabolism** (*CLPS, CLPSL1, RFPL4A/L1, F7, PDX1, WFS1, H2AC11*) show positive trends for both *Hn* and *Hs*. A few genes involved in lipid metabolism (*LPA, ZNG1E/F/C*, and *ACOT1*) exhibit trends specific to *Hn*, while others (*ACOX3*, *HAND1*) are unique to *Hs*. Conversely, the MST1 gene involved in **liver lipid and brown fat metabolism**, so important for the survival of newborns, appears to be of importance for both *Hn* and early *Hs*, both displaying a positive trend compared to hg38. These CNV trends provide key insights into ancient populations, but integrating them with modern genomic and clinical data allows us to deepen our understanding.

Comparing CNV trends in ancient and modern populations reveals two primary distribution patterns, clearly visible in **Figure 3**: in one, both modern and ancient individuals display limited CNV variation, with modern individuals having a single copy, while ancient individuals have higher copy numbers (*F7, PDX1, WFS1, H2AC11, ACOX3, HAND1*). In the second pattern, *Hs* and/or *Hn* exhibit narrow CNV variation (*ZNG1E/F/C, ACOT1, CLPS1, LPA*), while modern individuals show a broader CNV range. The *CLPS* gene deviates from these patterns, showing a wide CNV range in both modern and ancient populations: modern individuals range from 1 to 3 copies, mostly centered around 1 and 2, whereas *Hn* and *Hs* range from 1 to 4 copies, centered around 2 copies for *Hs* and 3 for *Hn*. The narrow CNV range observed in *Hn* across most genes confirms once more their great homogeneity. Overall, the elevated CNV in archaic individuals aligns with the role of these genes in enhancing nutrient processing and the ability to consume high-fat diets, as seen with *ACOX3* and *PDX1* (**Supplementary Text 3**).

The functional effects of CNV in modern humans—whether their copy numbers match those found in ancient genomes or their expression is altered by environmental factors such as high-fat diets—enable us to hypothesize about their adaptive significance in ancient populations. By investigating CNVs in genes involved in key metabolic processes, we can trace how these variations shaped dietary strategies and survival mechanisms across diverse environments.

Carbohydrate metabolism played a pivotal role in the dietary adaptations of *Hs*, as evidenced by the *AMY1A/B/C* genes. Here, broadening the number of genomes investigated since the pivotal studies by Perry and co-authors (4,5) provides further evidence that *AMY1A/B/C* CNV played a vital role in equipping *Hs* to digest and transform dietary carbohydrates and thus benefit from caloric energy for cost-efficient homeostasis. It is relevant here to estimate when the *AMY1A/B/C* duplication occurred (9,10) as the earliest *Hs* in Eurasia, represented in this work by the Ust'-Ishim fossil, already bore three copies per haploid genome. An ancient *AMY1A/B/C* duplication matches the archaeological evidence and provides molecular corroboration that cooked rhizomes could have been consumed as early as 170 ybp in South African caves (58), dovetailing with the occurrence of ground stones used to mechanically tenderize rhizomes and tubers during the Early Upper Paleolithic in Eurasia (15,59–62). Based on our data, it is parsimonious to consider that the emergence of an efficient strategy for starch digestion and metabolism is related to the *Hs* lineage already carrying the *AMY1A/B/C* gene duplication when they moved out of Africa, thus well before crop domestication (9,19). On the other hand, our findings reveal that the earliest *Hn*, such as the Altai Neandertal D5 (dated to 120,000 years ago) and Den D3 (dated to 76,000–51,000 years ago), possess only two copies of the *AMY1* gene (see **Tables 1A** and **S6**). Among the eight *Hn* individuals analyzed, all have either one or two copies. This confirms once again the great

homogeneity of *Hn*, which has already been demonstrated many times from an anatomical point of view in terms of bone and dental characters, and suggests that this character was also present in *Den*. The only exception is Les Cottès (dated to 43,000 years ago), which has three copies—potentially due to interbreeding with *Hs*, who may have introduced the third copy, as *Hs* were already present in Europe at that time. As shown in **Figure 3**, the CNV distribution in ancient and modern Sapiens spans a broader range of 1 to more than 5 copies across populations. In ancient individuals, the three copies were already present in the earliest *Hs*, with higher copy numbers becoming more prevalent in later Mesolithic individuals. All five modern populations exhibit a wide range of CNV variation. This stands in sharp contrast to the limited variation in *Hn* and *Den*, where the CNV is consistently concentrated between 1 and 2 copies.

A similar pattern of variation is observed in the *SULT1A3/4* and *SLX1A/B* gene clusters, where *Hn* display only one copy, while modern and ancient *Hs* show a broader CNV range. This indicates that gene duplications likely occurred after the EAH lineages split (9,19). We suggest that these duplications may have been driven by the nutrition changes faced by *Hs* in African territories compared to the temperate and cold environments where *Hn* and *Den* foraged their prey. The CNV of these genes may have made efficient lipid metabolism possible as a response to the cold stress during MIS 3 and moderating susceptibility to ketosis in adults (17)(p. 256). *SULT1A3/4*, involved in carbohydrate metabolism and in thyroid hormone activity, may have contributed to this adaptive trend by increasing cold-induced thermogenesis, to avoid the effects of "polar T_3 syndrome" (63). The involvement of thyroid hormone in vital pathways warrants further investigation in relation to *SULT1A3/4* gene cluster emergence, chronology and duplication. Overall, CNV in genes involved in oxygen and lipid metabolism provides valuable insights into adaptation to temperate and cold ecosystems. This evidence can help deduce or confirm thermal ranges and their implications for differential nutritional strategies among the three populations.

Genes *MST1* and *H1.2*, which are involved in **liver lipid and brown fat metabolism**, play essential roles in regulating adipose tissue function, crucial for thermogenesis and energy balance. *MST1*, also known as *STK4*, protects the liver from damage due to fasting or high-fat diets, stimulates adipose tissue differentiation (64), and importantly, acts as a physiological suppressor of mitochondrial capacity in brown (BAT), beige, and white (WAT) adipose tissues. Genetic inactivation of *MST1* increases mitochondrial mass and function, providing resistance to metabolic dysfunction induced by high-fat diets (65). These results suggest that, in archaic humans, increased *MST1* CNVs likely enhanced mitochondrial efficiency and metabolic resilience in challenging environmental conditions. *MST1* duplications are consistently observed in Den D3, and across *Hn* and *Hs* genomes, underscoring their significance in metabolic adaptation. On the other hand, the *H1.2* gene, a linker histone variant, regulates the browning of WAT and thermogenic gene expression in beige and brown fat cells, particularly in response to cold environments (66). Increased *H1.2* copy numbers in archaic humans likely provided a finely tuned mechanism for energy storage and release, beneficial for cold survival (66). Notably, higher copy number of H1.2 is seen in all *Hs* genomes except for the oldest out-of-Africa Ust'Ishim genome, indicating that these duplications became more widespread as *Hs* adapted to colder Eurasian climates. Additional genomic changes in *Hs,* such as *CD24* loss, associated with a reduction of WAT (67), *TCAF2* diversification, believed to be a result of positive selection in response to cold climates or dietary adaptations (68), and *MST1* duplication, which also plays a role in physiological responses to oxygen depletion (69) (see **Supplementary Text 3**), further support enhanced brown fat metabolism, providing flexibility and resilience in diverse environments.

Expanding Nutritional Insights: high-fat diet and nutrient utilisation in *Hn*, *Den* and *Hs*. New genetic evidence identified in this work (**Figure 3**) sheds light on the metabolic adaptations of *Hn*, *Den*, and *Hs* to high-fat diets and nutrient utilization, challenging established perspectives. Genes associated with micronutrient metabolism, including vitamins and minerals (70), exhibit notable duplications in both *Hn* and *Hs*. While animal products and oysters are the richest sources of zinc (71), plant-based sources such as pine nuts, hemp seeds, sesame seeds, chickpeas, and lentils, which were accessible to both *Hn* and *Hs*, also provide significant amounts of zinc. The *RFPL4A/L1* gene cluster produces a protein that binds zinc, and the *ZNG1E/F/C* cluster, which is vital for zinc transfer, exhibits a notable trend of increased CNV in *Hn* and certain *Hs*. Calcium is another important example of a food-related minerals. *WFS1* encodes wolframin, a protein believed to regulate cellular calcium levels, which is essential for various critical cellular functions. The association of *WFS1* with diet, particularly high-fat diets, has been demonstrated (72), showing that having only one functional copy of this gene can cause metabolic impairments. The presence of multiple copies of *WFS1* may have provided an advantage to individuals such as *Hn* and *Hs*, who followed high-fat dietary regimes.

New information on the *Hn* diet. Our work highlights that *Hn* metabolism was more efficient in the digestion of lipids than carbohydrates. The genetic adaptation to fat metabolism is an archaic pattern present in *Hn*, who were adapted to temperate and cold climates. The CNV trend of genes involved in fat metabolism (**Table 1**) undoubtedly emerged under cooler climatic conditions (even before the timing of Altai Neandertal D5, the oldest Neandertal in our sample) when the *Hn* differentiated. As already pointed out for *Hn* anatomical characteristics (i.e., their low stature resulting from adaptation to the cold) (13,73,74) and their cold adaptation genes (75,76), these adapted traits persisted even during interglacial periods, putatively due to the low genetic diversity of *Hn* (31,34). We found a trend in *Hn,* going from the oldest Altai Neandertal to the later Spy 94a individual, confirming independently what has already been noted from fossil analysis — that Altai Neandertals were highly carnivorous (77–79) while Spy 94a might have added proteins of vegetable origin to his diet (80). The CNV population trend of genes involved in fat and protein metabolism could even explain the peculiar skeletal morphology observed in *Hn*. A large bell-shaped lower thorax and a wide pelvis shaped to accommodate the larger liver and kidneys are an adaptation to a high fat, high protein diet. These organs are highly solicited for lipid and protein metabolism so the genetic evidence coincides with the specific anatomy (81) and high energy needs of *Hn* (13,82).

Hn shows a negative trend for genes involved in dietary carbohydrate digestion compared to the modern human reference genome and *Hs*. Nonetheless, the late *Hn* in our sample, Les Cottés ZA-1514, has 3 copies of AMY1A/B/C like the human reference genome, indicative of a less restricted foraging strategy being adopted by this late *Hn* due to the different environmental conditions (8,12,15,16,18,83). A plausible interpretation is that some late *Hn* may have interbred with early dispersals of *Hs*. Further analyses will be feasible once additional high-coverage genomes from southern or coastal *Hn* become available for study.

CLPS, CLPSL1, and *LPA* positive CNV trends in *Hn* have been reported to relate to fasting hepatic FA metabolism because they control the lipids balancing oxidative flux and capacity (84). Fats can be digested to form FA, and proteins can be partially transformed into glucose through gluconeogenesis (85,86). Gluconeogenesis is an efficient pathway to allow glucose to directly entering the bloodstream, to fuel the high demand of organs, notably the brain, which in *Hn* is very large (around 1450 cc, much larger than the average brain today) (13). Although gluconeogenesis demands much energy (87), within a highly carnivorous regime it can (i) provide glucose to the brain for a limited time, and (ii) become a

profitable pathway to maximize the benefit of protein intake (3) by effectively reducing the amount of nitrogen to avoid protein poisoning (3,85,88) so limiting ketosis in adults (8).

Regarding the beneficial role of FA in the diet, it is well known that nuts have always been present in human diet as shown by the Acheullian site in Israel 700 kyrs ago (89). Yet pine nuts were also available in the Altai Mountains refugia. In certain geographical regions (e.g coastal Mediterranean), it is possible that *Hn* relied on a diet much richer in omega-3 present in wild fish (90). FA are consumed at higher levels in diets incorporating more animal proteins and are critical for brain development, brain metabolism, and inflammatory responses (91).

A glance into the *Den* diet. The CNV analysis was possible on only one *Den* genome. Den D3 shows an *Hn-like* trend, particularly in genes related to lipid digestion, which is expected given their shared ancestry. Indeed, according to genetic estimates, the *Hs* and *Den* lineages separated from the *Hs* lineage some 800 ka ago, before separating from each other around 400 ka ago (30,47). It is acknowledged that *Den* in Altai are a western extension of a much larger population originating in central and southwestern Asia (92,93) with a less small size population that *Hn* (92). Further analyses will be possible when more *Den* genomes become available.

A new picture of the *Hs* diet. If it holds true that plants were part of the hominid diet for millions of years during their tropical forest foraging niche (94), the physiology required for efficiently digesting and metabolizing dietary carbohydrates, particularly starch, took time to adapt, likely through the redesigning of core metabolic and physiological processes (7). This took place during the evolution of *Hs* in Africa, where under favorable environmental conditions, they may have experienced the gathering and transformation of plant organs rich in starch (58,95). Our study provides three main contributions linked to dietary carbohydrates metabolism. First, *Hs* and the modern genome hg38 have a higher copy number for the gene *UGT2B28*, and the *AMY1A/B/C, SULT1A3/4* and *SLX1A/B* gene clusters compared to *Hn* and *Den*. Second, our data are consistent with the proposed early emergence of *AMY1* coupling – between 273-475 ybp – therefore after the divergence from *Hn* and well before crop domestication (9,10,19). Third, we identified the role of the *SULT1A3/4* and *SLX1A* gene clusters and of gene *UGT2B28* as main contributing factors explaining the first two dimensions of the CNV space of individuals in **Figure 4**. These three results prompt new questions centered on the importance of dietary carbohydrates in the diet of archaic humans and, in particular, in *Hs.*

From diet to survival: broader roles of key genes. Lastly, it is important to highlight three genes—*HAND1, BOLA2/2B*, and *F7*—that extend beyond diet and play significant roles in reproduction and survival. *HAND1*, which is duplicated in most *Hs*, has been shown to inhibit the conversion of cholesterol to steroids in trophoblasts, thus playing a vital role in the development and physiological function of the human placenta (96). *BOLA2/2B*, involved in the immune response in the gut, also plays a role in the maturation of cytosolic iron-sulfur proteins. Its duplication in *Hs* alters iron homeostasis, potentially serving an adaptive function by protecting against iron deficiency, which is crucial during embryonic development (97). Further investigation of these genes could provide valuable insights into the interplay between diet and reproduction in the evolution of the *Homo* genus. Together, these genes suggest a complex interplay between diet and reproduction in the evolution of the *Homo* genus, warranting further investigation.

Additionally, the coagulation factor *F7* gene, which is produced in the liver and is vitamin-K dependent, encodes an enzyme of the serine protease class that plays a critical role in the blood coagulation system, facilitating the formation of blood clots. This gene is duplicated in all archaic individuals including *Hn*,

Den and *Hs*. The duplication of *F7* induces fast blood coagulation, which could have been advantageous for these populations by enabling rapid wound healing and reducing the risk of infection, thereby increasing their chances of survival (75). Additionally, *F7* is also associated with dietary fat intake and cholesterol levels (98), with studies documenting its activity being influenced by diet (99). These findings underscore the broader adaptive roles of these genes, linking nutrition, reproduction, and survival strategies in archaic human populations.

Conclusion. The different CNV trends observed for *Hn* and *Hs* (and *Den*) reflect that these hominins relied on different sources of energy in their diets. Our unsupervised analysis of the CNV of multiple genes adds direct lines of evidence that confirm the biological adaptation of *Hn* to a narrow-spectrum diet and highlight the genetic changes facilitating the liver lipid pathway in both *Hn* and *Hs* living in the same nutritional environment of the cold Eurasia. The increasing duplication of *AMY1A/B/C*, *SULT1A3/4* and *SLX1A/B* genes evident in *Hs* is consistent with broad-based foraging strategies grounded, on the behavioral side, by the intentional transformation of starch-rich storage organs through the use of grounding stones, and this occurred much before crops domestication (59), started with the Neolithic (61,62). Different nutritional strategies among the three human populations inhabiting the temperate and cold latitudes during the crucial timing of their overlapping could have contributed to *Hs* better fitness leading to a demographic advantage that allowed them to conquer Eurasia and then the world.

Materials and Methods

An unsupervised procedure for the identification of genes involved in digestion presenting a CNV trend within an ancient population. The methodology (**Figure 2**) is based on several distinct computational steps:

1. Selection of all available genomes that could allow for an estimation of the haploid gene copy number.
2. Mapping of the selected ancient genomes to the modern human reference genome GRCh38/hg38 (42).
3. Clustering of recent paralogs and estimation of the haploid gene copy number for genes of the genomes identified in 1.
4. Identification of genes whose number of haploid copies in *Hn*, *Hs* or *Den* is higher or lower than in the modern reference genome GRCh38/hg38 for more than the half of individuals in the population. This means that a selected gene should display a *differential CNV* for at least 5 individuals in *Hn* or at least 4 individuals in *Hs*. This step identified 781 genes exhibiting population-dependent variation in copy number.
5. Selection of a subset of genes based on Human Protein Atlas tissue expression data and information on functional annotation in UniProt.
6. Comparison of CNV data from 28 diverse modern human ethnic groups to evaluate differences with ancient populations.
7. Gene Ontology (GO) and literature-based analysis of the final set of genes for a functional characterisation.

Steps 1-3 focus on read mapping and data preparation for the 19,988 genes identified across 15 ancient genomes. In steps 4-6, the dataset is refined by filtering the 19,988 genes down to 50 genes and gene clusters. Finally, step 7 links the selected genes to their functional characterization. Full details of the procedure are provided in **Supplementary Text 2**.

Figure 5A illustrates the analysis of the AMY1A/B/C gene cluster across modern and ancient individuals (see **Figure S1** for an extended analysis). In modern populations, *AMY1A/B/C* is duplicated, with a median of three haploid copies. When considering diploid CNV estimations, the distribution of pair and unpair diploid copy number aligns with prior analyses (54) of modern genomes, as illustrated in **Figure 5B**. CNV estimation of AMY1A/B/C of each genome is derived by summing the average read coverage over each AMY1 gene copy in

hg38. **Figure 5C** displays this read coverage across each copy for the *Hs* genome of Ust'Ishim and the *Hn* genomes of Vindija 33.19 and Mezmaiskaya 2 (see **Figure S8** for mapping on all ancient genomes). Ust'Ishim and Vindija 33.19 are high-coverage genomes, showing less sensitivity to mapping fluctuations compared to the lower-coverage Mezmaiskaya 2 genome.

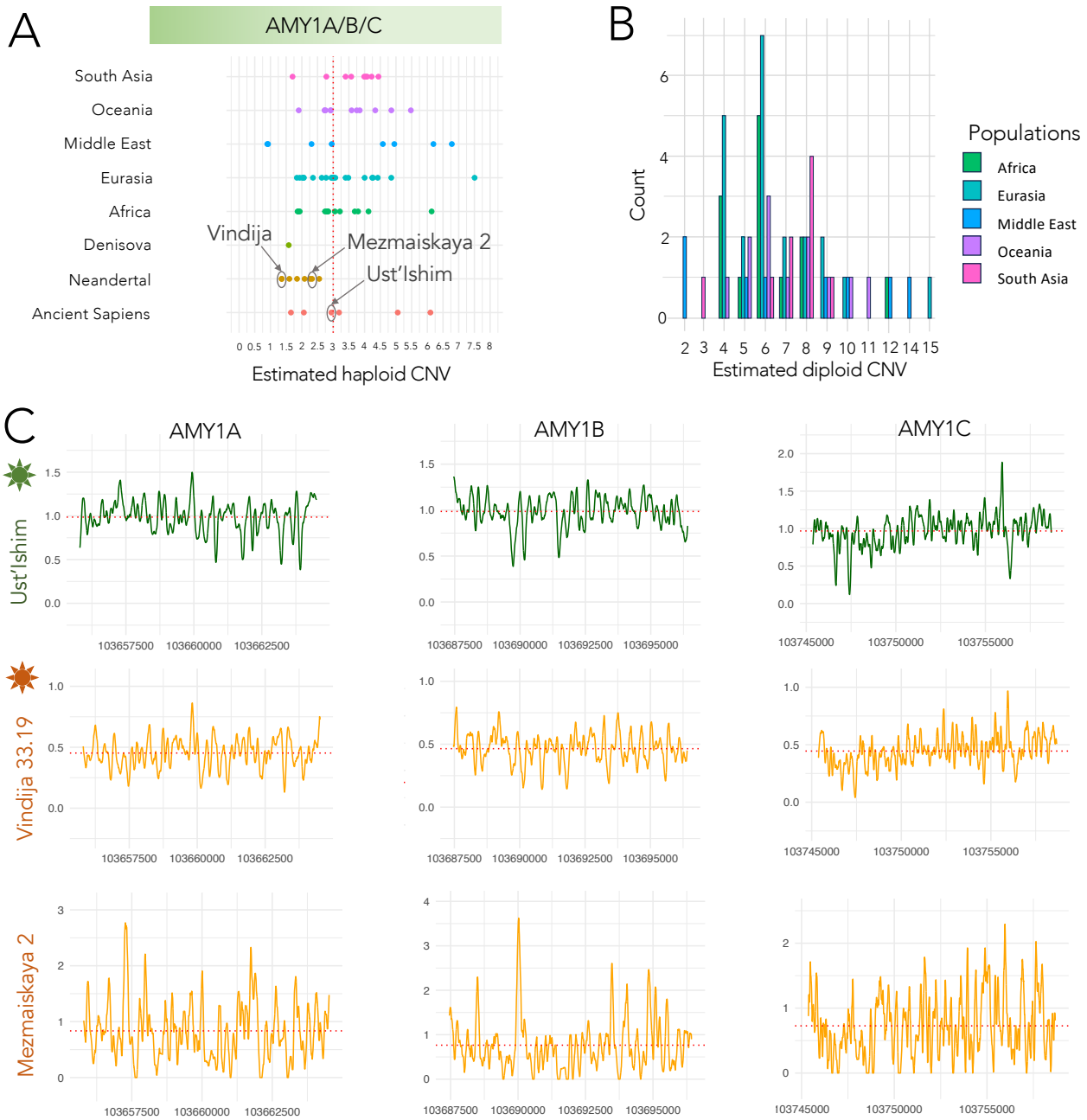


Figure 5. Read mapping analysis for the AMY1A/B/C gene cluster. A. Comparison of the estimated haploid CNV of the AMY1A/B/C gene cluster across all ancient and modern individuals in this study. See **Figure S1** for an extended analysis including genes AMY2A and AMY2B. **B.** Estimated diploid CNV for the modern populations considered in A. See also **Figure S1**. **C.** Read mapping for genes AMY1A (left column), AMY1B (center column) and AMY1C (right column) for Ust'Ishim (top row), Vindija 33.19 (middle row) and Mezmaiskaya 2 (bottom row). Stars for Ust'Ishim and Vindija 33.19 indicate high-coverage sequencing. For further analysis on all ancient genomes, see **Figure S8**.

After the first four steps identifying the genes with differential CNV, we used the notion of *CNV trend* to analyze them within more than one population. Formally, a gene has a positive/negative *CNV trend* in a population if its gene copy number is either greater/smaller than that of the modern reference genome for more than the half of the individuals in the population (5 of the 8 *Hn* genomes, 4 of the *Hs* 6 genomes). This criterion is aimed to highlight the tendency for a gene to have more or fewer CNVs than the modern reference genome GRCh38/hg38 within a population. The gene MST1, for instance, displays positive CNV trends for both *Hn* and *Hs*. The gene cluster AMY1A/B/C was selected because of its negative differential CNV in *Hn*, but in the *Hs* population there is no trend that is picked up by the CNV trend criterion. Also, the gene ZNG1E/F/C displays a positive trend in *Hn* and no specific trend for *Hs*. **Table 1**, **Table 2**, and **Table S6** show the aforementioned genes, their estimated copy number (rounded and not, respectively) for the 15 ancient genomes and the CNV trends for the two populations. The analysis on the Denisova genome is carried out for comparison.

Selection of ancient genomes for genome-wide analysis. Among all publicly available human ancient genomes, we identified all genomes presenting a minimum number of features that allowed us to properly perform a CNV analysis (see **Table S1**). Specifically, we considered genomes that:

1. present a minimum of 1x coverage of the human reference sequence, with available alignment;
2. have been affected less than 5% by contamination from modern human sequences;
3. retained sequences aligned with repetitive regions (indicated by a MAPQ value of 0).

The three conditions allowed 15 archaic human genomes to be considered. They were retrieved from accession codes published in manuscripts before December 2020, to which we added the more recent genome of Peştera Muierii (July 2021); see **Table S1** for downloading information.

Reads from these 15 genomes were remapped to GRCh38/hg38. We studied 18,043 genes/gene clusters (see below) which are covered by the 15 genomes and are shared with the modern reference human genomes hg38.

It is worth mentioning that some very interesting genomes, although characterized by a sufficiently high coverage, were discarded for various reasons: the available alignment files for Anzick-1, Kostenki 14, Sunghir 1--4, and Sunghir 6 do not comprise repeated regions; LaBraña lacks data for chromosome 22; Den D2, Hohlenstein Stadel, Scladina, San Teodoro 3, San Teodoro 5, and Zlatý kůň have low coverage and/or high contamination.

Selection of modern genomes for genome-wide analysis. To compare ancient and modern genomes, we selected 64 individuals from The International Genome Sample Resource (IGSR) database (40,41,100) of The 1000 Genomes Project (101), representing ethnic groups from Africa (13 individuals), Eurasia (22), the Middle East (8), South Asia (9), and Oceania (12) (refer to **Figure 1** and **Table S3)**. The sequence reads from these genomes were previously aligned to the hg38 reference genome. CNV analysis for these modern genomes was performed using the same pipeline as for the ancient genomes (see **Supplementary Text 2**).

Technical warning. The quality of sequencing data plays a crucial role in read depth analysis and in the copy number determination that can differ in precision (53,102,103). Indeed, human genome assembly might not fully map highly complex genomic regions raising potential doubts on the correct estimation of gene copy numbers in human populations. By choosing GRCh38 genome as the current human reference, we guarantee the latest accessible estimation to gene copy numbers in the reference genome.

Acknowledgments: We thank Alissa Mittnik (Johannes Krause lab) for sharing her estimate of contamination of the Motala 12 nuclear genome.

Funding: This work has been partially supported by:
- ANR grant STARCH4SAPIENS – grant ANR-20-CE03-0002 (AC, SC, LL);
- LabEx CALSIMLAB grant ANR-11-LABX-0037-01 - constituting part of the "Investissements d'Avenir" program grant ANR-11-IDEX-0004-02 (AC, RV);
- SUG grant M4081669.090 (NTU, Singapore) (LL);

- PHC Merlion, 2018, Singapore-France bilateral agreement (NTU, Singapore) (LL, SC).

Author contributions:
Conceptualization: AC, LL, SC
Methodology: AC, RV
Investigation: AC, LL, SC, RV, LP, NR
Visualization: AC, RV, LP, NR
Funding acquisition: AC, LL, SC
Supervision: AC
Writing – original draft: AC, LL
Writing – review & editing: AC, LL, SC
Writing – SOM: AC, LL, SC, RV, NR, LP
All the authors discussed the results and agree on the final version.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or the supplementary materials.

Supplementary Online Material

Dietary adaptation in Neandertal, Denisovan and Sapiens revealed by gene copy number variation

R. Vicedomini, N. Righetti, L. Polit, S. Condemi, L. Longo, A. Carbone

Supplementary Text 1 - A general presentation of the 15 ancient genomes considered in this study, pp 25

Supplementary Text 2 - Methods, pp 34

Supplementary Text 3 - A detailed description of the 50 diet-related genes, pp 42

Supplementary Tables - pp 66

Supplementary Figures - pp 85

Supplementary Text 1

S1. Presentation of "archaic humans"

Neandertals (*Hn)* are a human hunter-gatherer fossil population that lived in Eurasia between 300 kya and 40k ybp. They were totally replaced throughout their territory by us, *Homo sapiens*. Paleoanthropologists have shown that the evolution of this fossil population could be originated from *Homo heidelbergensis* in Europe, back at least from 450k ybp, through the identification and progressive accumulation over time of typical *Hn* morphological features (104).

In Africa H*omo heidelbergensis* gave rise to *Homo sapiens* (*Hs*) whose oldest representatives, from fossil remains, are identified around 300 Ka (105). Thus, while *Hn* was differentiating in Europe, *Hs* was differentiating in Africa. These data are consistent with the genetic ones (106). By analysing the genomes of *Hn* and of modern humans (*Hs*) it was estimated that the two ancestral populations separated between 220 000 and 470 000 years ago (107). This splitting time is highly relevant for the emergence of the *AMY1A/B/C* gene cluster (9) and it is relevant for the reasoning presented in this paper. According to some studies, this timeframe would seem to match that of the extinct species *Homo heidelbergensis*, which has been found in Africa, Europe, and possibly Asia (108). The fact that only present-day Europeans and Asians possess this *Hn* genetic capital indicates that interbreeding must have taken place just after the first *Hs* left Africa to conquer Eurasia, which the researchers estimate to occur in several waves starting approximately 100,000 years ago (47,109). This hybridation has been since confirmed by several analyses (the last being reported in (35)), which succeeded in sequencing the genomes of five *Hn*, including four late *Hn* from Western Europe - one from France, two from Belgium and one from Croatia - and an older one from the Russian Caucasus. They also show that the *Hn* genes of present-day Europeans are not close to those of late European *Hn*. Both this genetic distance and the enigmatic absence of *Hs* genes in late *Hn* can then be explained if one's makes - and validate - two hypotheses: on the one hand, that interbreeding mainly induced gene flows in the direction of *Hn* ➞ *Hs*; on the other hand, that the main hybridization event occurred outside Europe well before the arrival of our *Hs* ancestors on this continent (106,107,109).

The European *Hn*, for which extensive morphometric data are available, underwent four glacial events and thus progressively adapted to colder environments (13,110). It is therefore under particular climatic and environmental conditions, with alternating glacial and interglacial phases, that the evolution of *Hn* took place in Eurasia. The climate had important consequences for their adaptability, density and therefore diversity, as well as for their possibilities of expansion and contact with their neighbouring population (111). The *Hn* populations have experienced few bottlenecks, therefore they have never been large in size, structured in small interconnected groups (about 20 individuals) that never outnumbered 70,000 individuals at the time of their "golden age" (during the "Eemian" interglacial at MIS 5, around 120k ybp) (112). The small size of the population has to be correlated with the partial geographic isolation of *Hn* caused by European climatic fluctuations during Pleistocene.

It was around this time (circa 100k ybp) that the European *Hn*, well identified by singular morphological features, spread eastwards carrying the same lithic assemblages and technology as far east as the Altai mountains (113) where they encountered the Denisova (*Den)*. Genetic data has also pointed out low genetic diversity in *Hn*, with a demographic depression peak in the Altai, where a very consanguineous individual has been found (31), and a genetic continuity across Europe from 120k ybp until the disappearance of the population around 40k ybp (114). This low variability is also visible in the morphology of the *Hn*, which remained the same during the last 100k of their existence throughout their broad territory from the Atlantic to the Altai (104) (see SOM p.15 in (113)).

The expansion of *Hs* and the demise of *Hn*

After the great dispersion of *Hs* from Africa into Eurasia about 60,000 years ago (1), the other archaic humans (EAHs), notably the *Hn* and *Den*, who populated this geographical region, disappeared. The reasons for the success of *Hs* are debated among specialists. The most commonly accepted hypothesis is that EAHs would have competed with *Hs* for food resources during the climatic changes that affected Eurasia (104,115).
The replacement of EAHs, *Hn* in particular, would have been favoured by *Hs*' greater technical skills (116), their greater cognitive abilities (116–120) and lower social capacities and network (116,121,122).

However, all these hypotheses have been challenged (123,124) as also those suggesting that the disappearance of EAHs was brought about by violent confrontations between the two populations (125) and by the exposure to new infectious agents (126,127).

The great deal of attention paid to behaviourally and culturally driven reasoning somehow overwhealmed the reasoning on the biological complexity of physiological adaptation (110,128) and comparatively little attention was paid to the metabobolic and physiological processes ruling the bioenergetic requirements (8,129–131), the role of microbiomes (128,132) and how to transform different foods into nutrients (133,133). Our analysis add fresh data to the genomic architecture across EAHs suggesting a selective pressure on Neandertal, and possibly on *Den*, towards lipid-rich food to respond to the cold climatic regimes at the boreal latitudes. Conversely, our results support an adaptation to a diet rich in glycemic carbohydrates for *Hs*, evidenced by the old duplication of *AMY1A/B/C* gene cluster (9).

For our CNV analysis, 8 *Hn* genomes are available: 2 from Asia (in Altai: Altai Neandertal D5 and Chagyrskaya 8) and 6 from Europe (in Russia: Mezmaiskaya 1, Mezmaiskaya 2; in Croatia: Vindija 33.19; in Belgium: Spy 94a, Les Cottés Z4 1514, Goyet Q56.1).

Denisovan, *Den*, on the other hand, are also an extinct human population but their "discovery" in 2010 is mainly due to mitochondrial (mt) and nuclear (n) DNA studies (47,134–136) extracted from the collagen of bones so fragmentary that they did not allow paleoanthropologists to identify their morphological and anatomical characteristics. *Den* are considered an Asian sister group of *Hn* that have evolved in Asia from *Homo heidelbergensis* after the expansion from Africa and the split with the African ancestors. This African origin has been confirmed and reinforced by genetic studies (from 2010 up today), that have also contributed to date the time of divergence between the African lineage, that will evolve into *Hs,* and the Eurasians' one that will give rise to *Hn* in Europe and to *Den* in Asia. The analysis of the nuclear genome suggests a recent common ancestor between European *Hn* (notably with Vindija, Vi 33-19) and *Den* dating between 440 to 390 ka ago (34) and thus, situated the *Den* as a sister group to *Hn* (30,34,137) and *Hs*.

The exact geographic repartition of *Den* is not yet clear but available data supports that the Altai is the western edge of their territory that could likely expand to North and South East Asia, until Papua and possibly the Sonda archipelago. *Den,* originally known only from the Denisova cave in the Siberian Altai, has been recognized even in the incomplete hemi-mandible from Baishiya cave (Tibet), on the basis of proteomic study (138) and from DNA extracted from the sediments (139), providing evidence for their occupation of the cave perhaps until 45 ka, when early dispersal of *Hs* was already roaming into Asia.

Probably, *Den* occupied a large part of continental and southern regions of Asia (93,140–143). Indeed, it is in modern Asian populations that the highest percentage of *Den* genes are found (144,145). *Hn* and *Den* have interacted in the Altai area and they occupied successively and perhaps even together Denisova Cave, as witnessed by the Den D11 individual (118.1-79.3k ybp) that had a *Den* father and an *Hn* mother (134,146). The characteristics of the Den D3 genome enable our CNV analysis.

S1.1 <u>The *Hs* peopling of Europe seen from paleoanthropological and genomic analyses</u>

From archeological, paleoanthropological and genetic studies the European *Hs* peopling show that the ancestry of modern Europeans is quite complex (147). In western Eurasia five dispersals of *Hs* arrival can be identified:

1) An "Aurignacian dispersal" of *Hs*. This arrival is supported by the presence in Eastern and Central Europe of archeological evidence for "Aurignacian flaked industries " resembling those of the Levant, e.g. Ksar Akil, Lebanon, dating about 43k ybp (148) (see also the recent debate in (149)). These cultural assemblages, that lasted in western Europe until around 31k ybp are always associated to *Hs*, although at the time of modern humans' arrival *Hn* was still present in Europe (150) making it possible that hybridation have occurred (151).

Although for this early phase of dispersal no genomes are available that belong to it with certainty, in this study we analyzed the Peștera Muierii genome. Other genomes from early *Hs* individuals like Kostenki 14, dating to 38k ybp (152,153), Bacho Kiro, dating 45-58k ybp (151) and Zlatý kůň, 45-43k ybp (154) could not be included in our analysis due to genome sequencing characteristics not allowing for CNV analysis (see **Methods**). For this earliest dispersal we analyzed the Ust-Ishim genome (> 45k ybb).

2) *Hs* belonging to the Gravettian assemblages. The Gravettian population seems to have formed a homogeneous biological group despite their chronological and geographical extension that covers all Europe over about 10.000 years. The oldest Gravettian occurrences are dating from around 31k ybp, when *Hn* had already disappeared, until 23k ybp, hence Peștera Muierii 1 and Mal'ta child genomes fall in this group. These fossil populations of hunter-gatherers developed a meta-culture famous for its flourishing art, as well as some emblematic archaeological sites including the Cro-Magnon shelter discovered in 1868 (Dordogne, France) that gave the name with which modern humans are aknowledged.

Unfortunately, all the genomes available from this span time (notably Sungir, Kostenki I, Dolni Vestonice, Pavlov, Krems-Wachtberg) do not have the genome sequencing characteristics allowing for CNV analysis (see **Methods**).

3) A new arrival of an unidentified *Hs* population of hunter-gatherers around 14k ybp, probably coming from Anatolia (155,156). This new arrival expanded in Europe during the last Ice Age *Hs* experienced an important demographic bottleneck. Therefore, there is genetic evidence for discontinuities in the European peopling before and after last glacial maximum. For this span time no available *Hs* genomes are worth to match the criteria set for our CNV analysis.

4) Anatolian *Hs* farmers, who spread into Europe about 8,500 years ago, via the Mediterranean coast and the Danube valley (38). In our study, we analyzed the CNV of genomes from late Mesolithic *Hs*, namely Loschbour (Belgium) and Motala 12 (Sweeden) and from an early farmer from Stuttgart, all living in Europe after the great bottleneck that followed the last glacial event.

5) A nomad population from the steppes of southern Russia, who migrated to Europe and East and South Asia about 5,000 years ago (39,157) which is beyond the scope of the present analysis.

S2 Presentation of the fifteen re-analyzed probands

In recent years, the availability of good quality nuclear DNA from various human fossils has increased the potential of reconstructing the hominins phylogeny and their interbreeding (158). Also, the availability of hundreds of thousands of complete human genomes from different modern populations allows the reconstruction of maps of single nucleotide polymorphisms and structural variants. Among all the ancient DNA genome sequences from *Hn* and *Den* publically available to date for our unsupervised analysis (this study was finalised in 2024), we selected only those with high-quality genomes (see **Methods**) namely one *Den* (D3) and eight *Hn (*Altai Neandertal D5, Chagyrskaya 8, Mezmaiskaya 1, Spy 94a, Vindija 33.19, Mezmaiskaya 2, Les Cottés Z4 1514, Goyet Q56.1). These probands are representative of the two archaic human (*Hn* and *Den*) that spanned over 50,000 years of the Late Pleistocene and approximately 8,000 km across Eurasia. We also analysed three *early Hs* genomes (Ust'Ishim, Peștera Muierii 1, Mal'ta), two late hunter-gatherers and one early farmer from Europe *Hs* genomes (Loschbour, Motala 12, Stuttgart). In all, 15 genomes were selected for the CNV study (**Table S1**).

S2.1 Denisovan (*Den*)

Den D3 (Denisova cave, Altai Mountain, Central Siberia)
Den D3 ("pinky"), is a distal phalanx of the fifth finger belonging to an adolescent female individual. This is the first fossil attributed to a new fossil population on genetic ground (30,47,134,135). It was unearthed in layer 11.2 square D2 of the East Gallery at Denisova cave (51.40N, 84.68E) during 2008 field season. It has recently been

dated to 69–48k ybp from optical dating of the associated sediments and to 76.2–51.6 kybp using a Bayesian modelling approach that combines chronometric (radiocarbon, uranium-series and optical ages), stratigraphic and genetic information to estimate ages for the hominin fossils at the site (159). Den D3 lived approximately at the same time as *Hn* Chagyrskaya 8 (32) or slightly before, thus sharing similar environmental conditions.

S2.2 Neandertals (*Hn*)

Altai Neandertal, also known as **D5 (Denisova cave, Altai Mountain, Central Siberia)**
Denisova is the name of a cave which is located at a height of 700 m asl dominating a green valley crossed by the river Sibiryachikha. The cave is located at the foot of the Altai Mountains in Siberia, near the border between Russia, Mongolia and Kazakhstan. This is not the only cave or prehistoric site in the region as several *Hn* skeletal remains have been retrieved in the nearby Okladnikov cave and at Chagyrskaya cave, about 100 km away (32).

The Altai *Hn* occupied Denisova cave roughly 110 ka ago (159,160). The Altai Neandertal is represented by a proximal pedal phalanx of a female that furnished a 52-fold genome coverage (31). Several very incomplete fossils from layers 12.3 and 11.4 have also been assigned to *Hn* based on their mitochondrial (161) and nuclear DNA (31). Additional evidence for the presence of *Hn* in the cave comes from analyses of sedimentary DNA from at least 2 layers (11 and 14), which indicate several episodes of *Hn* occupation (146).

Chagyrskaya 8 (Chagyrskaya cave, Altai Mountains, Central Siberia)
Chagyrskaya is a karst cave discovered in 2007 and it is located in southern Western Siberia, near the Charysh River in the foothills of the Altai Mountains. Since 2008, 74 *Hn* remains, along with fauna and a rich lithic industry, have been unearthed, providing a large and mostly well-preserved collection of morphologically diagnostic *Hn* remains from the Altai mountains. The faunal and pollen remains suggest a milder climate in the Altai foothills that was likely a refugium for *Hn* populations during the late Pleistocene (162–166).
Chagyrskaya 8 is a distal manual phalanx from a *Hn* individual whose genomes have been sequenced to 27-fold genomic coverage (32).
Although the age of Chagyrskaya 8 is constrained by a DNA-based estimate of ~80 ybp (28) and an optical age for the associated deposits of 59–49 ybp, Chagyrskaya 8 may have lived at around the same time than Den D3 or up to several millennia later. The genome from Chagyrskaya 8 shows that this *Hn* is related to *Hn* from western Eurasia (notably with Vindija 33-19) more than to the Altai *Hn* who lived earlier in Denisova Cave (31,32,34,35). It is possible that distinct groups of Neandertals repeatedly migrated between Western Eurasia and Siberia (113), but it is also possible that Siberian *Hn* lived in relatively isolated populations of less than 60 individuals on the Eastern edge of the *Hn* territory. The long-distance movements from western Europe to Altai of *Hn* is further confirmed by the studies of lithic assemblages – more than 90,000 pieces – resembling to the so-called "Micoquian-like assemblages" observed in Central and Eastern Europe (113). These connections are also supported by a recent study on *Hn* blood system (104). These data make it possible to explain the great homogeneity that is observed in the anatomical characteristics of the *Hn*, which is found throughout their territory from west to east and over the course of time from the earliest to the most recent.

Mezmaiskaya 1, **Mezmaiskaya 2 (Mezmaiskaya cave, Krasnodar region, Caucasus, Southern Russia)**
Mezmaiskaya Cave is located near the right bank of the Sukhoi Kurdzhips river, in the north-western foothills of the North Caucasus, southern Russia. Among several *Hn* remains retrieved in the cave, the genome extracted from a rib of a still breast-feeding neonate (Mez 1, who died about 2 weeks after birth) dating back to ~70k ybp was the second *Hn* genome sequenced after the Feldhofer, retrieved in the eponym site, in the early 2000s (167,168). In 2016, a successful extraction from a skull fragment of another infant (Mez 2, 1-2 years old, so putatively still breast-feeding) dated to 44,6–42,96k ybp, revealed an optimal collagen preservation that permitted to sequence the genome (35).

The still ongoing excavation, by yielding human skeletal remains associated with middle Paleolithic assemblages of local facies called "Micoquian tradition" (169), has revealed the key role played by the Caucasus corridor not only for *Hn* but also as one of the gateways for *Hs* into the boreal latitudes.

Vindija 33.19 (Vindija cave, Northern Croatia, Southern Central Europe)

The Vindija Cave, in the Dinarides Alps, discovered in 1974, is a highly recognized site for the richness in human skeletal remains of both *Hn* and a few *Hs.* Three fragmentary bones of *Hn* were analyzed for the first draft sequence of the Neanderthal genome project (109). In our analysis we consider Vindija 33.19, one of the five *Hn* osseous remains that were recently genotyped (34,35). It belongs to a *Hn* female that lived 48-65k ybp (170) and furnished a 30-fold genome (34). The comparison with the previous *Hn* genomes (109) allowed inferences on the similarity the Neandertal population from Vindija with late *Hn* from western Europe and from the Gibraltar area (114). However, the recently obtained Vi 33-19 genome was compared with those of Chagyrskaya 8 and of the other *Hn* individuals retrieved in that cave, allowing inferences on the western origins of these Altai *Hn*, even considering the technological skills observed in the lithic assemblages (32,113). Several individuals from Croatian site have been also deeply investigated for dental microwear texture analysis and the results suggest that some individuals among the Vindija *Hn* group where consuming soft food like meat at high level (171) and this data are fully in agreement with our results and with the isotopic analysis (172) although anisotropy on teeth wear may suggest some raw fibrous plant food, features that can be also compatible with the use of teeth as a third hand involved in fibres processing.

Spy I/94a (cave in the Namur province, Belgium)

The site of Spy is well known since the end of 19th century and yielded *Hn* several individuals, highly fragmented. The first two skeletons were discovered in 1886 in Jemeppe-sur-Sambre, province of Namur, Belgium (173,174) and represent the second official discovery of Neandertal fossils ever reported, thirty years after the one in the Neander Valley in Germany (Feldhofer). Since 2010, the re-examination of all the human remains have permitted to identify at least 6 individuals, one juvenile and five adults.

The genome considered in our study, was obtained from Spy 94a upper right molar (M_3) from a male individual which maxillary fragment including the M_3 *in situ* was radiocarbon dated to 35,810 + 260, –224 years BP (173,175). The dental calculus analysis of Spy I lower right molar (M_1) revealed some starch grains, most probably to be attributed to USOs (under storage organs such as water lily (15) which rhizomes are poor in proteins, (175)) and also to grasses. On the basis of these occurrences have been inferred that Neandertals from Spy were including plant food in their diet (15), which is in accordance with the tooth wear-traces observed by (176). Also, tooth wear from Spy I resemble those observed in different individuals from Vindija (but not from Vi 33.19, the specimen in our analysis), Krapina, Hortus and from Kulna 1 (these three latter individuals are not genotyped), and the wear is interpreted as coherent with some plant food consumption (175). By showing two copies of the AMY1 A/B/C gene cluster, our data on Spy I/94a lend poor support to the efficient digestion of starchy food.

However, at present, it is not possible to directly compare our genes CNV trends with the data from other *Hn*/Early *Hs* specimen yielding starch entrapped in dental calculus as reported in recent reviews (175,177). Furthermore, other studies caution about the attribution of plant remains retrieved in dental calculus to the exclusive connection with food consumption (177) and bring about insights on the use of teeth as the “third hand” therefore starch, fibres, tissues residues entrapped in the calculus might also be referred to alternative plants processing (e.g. for technological processing) or for medical purposes (178,179). This economic activity can also serve as an alternative interpretation for the tooth wear recognized on Spy 1(M^2) previously interpreted as due to course dietary resources (175).

Goyet Q56–1 ("third cavern" of Goyet cave, Belgium, Central Europe)
The so-called "third cavern" cave, excavated during the 19th century and then re-opened in the 1990s, belongs to a large karstic complex in the Mosan basin, Belgium. It yielded almost 100 human skeletal remains and a rich Middle and Upper Paleolithic industries difficult to disentangle due to the methodology applied by Dupont, one of the pioneers of the studies in Belgian Prehistory. Among the many remains attributed to *Hn*, for the CNV analysis we consider the Goyet Q56-1 genome, obtained from the fragment of the right femur (35) and dated to 40.5–45.5k ybp. The isotopic analysis has been carried out on several fragments of other *Hn* genotyped fragments (Q119-2, Q305-4, Q305-7, according to mtDNA) and the data closely resemble those from other Western and Central Europe hominins like Feldhofer, El Sidron and Vindija (180). The modest genetic variation shown by these genomes is supportive of the low genetic variability typically explained by the limited number of individuals of late *Hn* clades. It has to be mentioned that dental calculus analysis was carried out on Goyet VI *Hn* individual highlighting that this individual was adding some plant food on his prevailing carnivore diet (16). According to our data Goyet Q56-1 – together with Spy I/94a and Les Cottés Z4–1514 (see below) – shows the duplication of the *AMY1 A/B/C* gene cluster lending credence to the digestion of dietary carbohydrates for these individuals. However, carbon isotopic data suggest that other *Hn* individuals living at Goyet cave were heavily relying on mammoth, as it occurred for Spy and Scladina Neandertals (181), which is also a dietary condition supported by our data (high duplication of the genes involved in lipid and protein metabolism).

Les Cottés Z4–1514 (Les Cottés cave, New Aquitaine, France, Western Europe)
The Cottés cave or Cottets cave is located in Saint-Pierre-de-Maillé, in France. The first excavations and discoveries of Mousterian and Aurignacian occupation levels date back to 1880. Since 2006, new excavations have been carried out, as this cave has the particularity of presenting a continuous archaeological sequence from the end of the Middle Paleolithic to the beginning of the Upper Paleolithic, thus including the Mousterian, the Chatelperronian and the first phases of the Aurignacian. Les Cottés Z4–1514, a *Hn* tooth dated to 43,740–42,720 ybp, has been recently genotyped (35) allowing the CNV analysis. As for the two mentioned Belgian *Hn* individuals, even for Le Cottés Z4–1514 it would be possible to suggest the access to a more mixed diet.

S 2.3 Homo sapiens (*Hs*)

In the frame of this article, we analysed the CNV of an Early *Hs* Palaeolithic hunter-gatherer from Europe, Peștera de Muierii, and two from Asia, Ust'Ishim (north-western Siberia), at present the oldest *Hs* at the boreal latitudes who lived prior to the split between Western and Eastern Eurasian, and the Mal'ta child, that lived in eastern Siberia around 24k ybp.

A- *Hs* Palaeolithic hunter-gatherers

Peștera Muierii 1 ("Women's cave", Romania, Eastern Europe)
Peștera Muierii, or the "Women's Cave", opens in a complex karst system in the Carpathian Mountains. The cave yielded numerous cave bear remains as well as a *Hs* human skull.
Together with the human remains found in 1952 – a skull and some post-cranial remains (Peștera Muierii 1 or PM1) a temporal bone (Peștera Muierii 2) that cannot be articulated with the skull, and an isolated fibular diaphysis (Peștera Muierii 3) – lithic artifacts attributed to both the Mousterian technology (usually referred to Neandertals) and to the Aurignacian (associated to early modern humans) were retrieved. The *Hs* remains are dated to about 34 ybp (182), thus, they are somewhat younger than those from the nearby cave of Peștera cu Oase, dated to about 40k ybp (183).
The PM1 skull displays modern human features, including a high forehead, a small maxilla and small eyebrow arches. The large cranial vault and intact facial bones indicate a woman with 'robust features'.

The mitogenome of the skull (182) obtained from two of its teeth suggest that the female from Peștera Muierii was part of the first population of *Hs* after its expansion into Eurasia.
The genome was recently analysed (36) and shows some level of admixture (~3.1%) and also a quite high level of genetic diversity, a common trait within the earliest European. The human remains of Peștera Muierii represent a group that was a side branch to the ancestor of modern-day Europeans that underwent a great bottleneck during and after the most recent Ice Age.

Ust'-Ishim (cave on the bank of the Irtysh river, Central Asia, Northern Siberia)
Ust'-Ishim Man is the name given to the fossil of a *Hs* femur, fortuitously discovered in 2008 on the bank of the Irtysh River by a Russian sculptor near Ust'-Ishim, in the Omsk province of north-western Siberia. This fossil is notable for the preservation of its DNA, which allowed the complete sequencing of its genome, thus becoming the earliest – > 45k ybp – modern human genome sequenced to date (33,184).

The examination genome shows that Ust'-Ishim individual is closely related to the 24 ybp Mal'ta boy (see below) from central Siberia, but also to the much later individual from La Braña (Spain) who lived about 8,000 ybp. This genetic relationship show that Ust'-Ishim femur belong to a population that predated the separation between western Eurasians (European population) and the eastern Eurasians (Asians populations) (185). Some authors consider Ust'-Ishim likely belonged to a population that predated this split or represented another migration to the Asian continent that would not have left descendants among the present human populations (151). The same authors observed that Neanderthal DNA in Ust'-Ishim appears in longer sequences, indicating that interbreeding between modern humans and *Hn* took place shortly before the life time of this fossil. Calibration of the genetic clock, following the sequencing of the Ust'-Ishim genome, permits to estimate the date of interbreeding between *Hs* and *Hn* at between 52,000 and 58,000 ybp. This age corresponds to the estimated age of the last exit of *Hs* from Africa (148). No hybridisation between Denisovans and Ust'-Ishim has been demonstrated, although Denisovans are thought to have lived at the same time in East Asia as suggested by the DNA retrieved in the sediments of Denisova (146,151).

Mal'ta (open air site, Eastern Siberia, Central Asia, Russia)
The four-year-old **Mal'ta** boy lived along the Bolshaya Belaya River near Irkutsk, in central Siberia. This individual is dated 24,300 ybp and belonged to a population settled between the Lake Baikal and the Yanisey River under very rigid climatic conditions with few plants and animals to live on. Nonetheless, they survived in a large community – much larger than the Neandertal's ones – and built hypogeum construction made of mammoths' bones and reindeer antler likely covered with leather, and they were capable of fine carving amulets using the ivory and the antler of that animals (186). The Mal'ta eastern Paleo-Siberian group became genetically isolated from its original population being different from the other populations of the eastern Asia (187). The Mal'ta population is ancestral to Native Americans that received around the 40% of their ancestry (37) and to the present-day Kets and the Selkups, known as highly carnivores (188). Our data clearly show that this same adaptation is evident in Neandertal – Spy I/94a and D5 Altai Neandertal – and also Denisovan – D3. Mal'ta represents the closest link to Motala 12, since the Swedish late hunter-gatherer shows about 22% of ancient North Eurasian genetic capital (157).

B. *Hs* from Europe prior the Neolithic: Mesolithic (Motala 12, Loschbour) and early farmer (Stuttgart).

Motala 12 (necropolis – Kanaljorden, Sweden, Northern Europe)
Motala 12 is a maxilla belonging to a male individual from the Scandinavian site of Kanaljorden, excavated in the town of Motala between 2009-2013. In this site, attributed to late Mesolithic hunter-gatherers, several individuals - an infant and 10 adult burials – were retrieved on the shore of a small lake. Motala 12 maxilla was

directly dated to 7,212 ± 109 BP (SOM p. 4 in (38)). The sequenced genome (38,39) informs that this fossil shares with its cohort large teeth and thick dark hairs and pigmented skin, as it occurred in present days East-Asians (157) and Native Americans. It must be noted that our results finely match with previously available ones for the late northern hunter-gatherers regarding the efficient digestion of dietary carbohydrates and the perception of bitter taste (37).

Loschbour LB 1-3 (rock shelter, Luxembourg, Central Europe)

The "Loschbour man" (also Loschbur man) is a skeleton discovered in 1935 in rock shelter in Mullerthal (Luxembourg) by Nicolas Thill (SOM p. 2 in (38)). The skeleton now belongs to the collections of the National Museum of Natural History in Luxembourg City. Loschbour man, directly dated in 1998 (AMS) to 7,205 ± 50 ybp (SOM p. 2 in (38)), was a late Mesolithic hunter-gatherer aged 34-47 years old, circa 1.6 m tall who weighed between 58 and 62 kg (189). The Loschbourg population was supplanted by new populations more likely used to herd rather than hunting. According to DNA analysis made from a tooth (LB1-3) by (38), Loschbour man was "a light skinned (white) individual" (90%), with brown or black hair (98%), and likely blue eyes (56%). Loschbour's late hunter-gatherer was effective in digesting starchy food, displaying multiple copies of *AMY1A/B/C*, as the later agro-pastoralist populations (19) that lived in the same area. Also, this individual shows the PAV haplotype for taste receptor *TAS2R38* as Motala 12 individual. Loschbour shares with ancient North Siberians the duplication of those genes involved in the amino acid catabolic process, with some genes also involved in the urea cycle and in tyrosine and phenylalanine catabolism (188).

Stuttgart (LBK 1) (Viesenhäuser Hof, Bavaria, Germany)

A tooth – M_2 listed as LBK1– was genotyped within one of the largest studies dedicated to the sequencing of the Holocene ancestral population for present-day Europeans (38). This is an interesting site where both local population and migrant from Anatolia was buried. A female – 25-to 35 years old – affected by primary hyperparathyroidism, lived all her life in the surrounding as the strontium isotope analysis suggests (190) and was buried in this Early Neolithic site as showed by the associated Linear Band Keramik artifacts dating back to 5100 to 4800 ybp (191).

Supplementary Text 2

Methods

Detailed steps of the unsupervised procedure identifying genes with differential copy number involved in digestion

The following seven steps of the pipeline will be described in detail:
1. Selection of all available ancient genomes suitable for estimating haploid gene copy number.
2. Mapping of the selected ancient genomes on the modern human reference genome GRCh38/hg38.
3. Clustering of recent paralogs and estimation of haploid gene copy numbers for the genomes identified in 1.
4. Identification of genes with population-dependent variation in copy number.
5. Selection of a subset of genes based on tissue-specific expression and functional relevance.
6. Comparison with CNV data of modern human populations.
7. Gene Ontology (GO) and literature-based analysis for functional characterisation.

Step 1. Genome selection for genome-wide analysis

From publicly available human ancient genomes, we identified those that met specific criteria necessary for accurate CNV analysis (see **Table S1**). Specifically, we included genomes that:

1. had a minimum of 1x coverage of the human reference sequence GRCh37/hg19, with available alignment;
2. exhibited less than 5% by contamination from modern human sequences;
3. retained sequences aligned with repetitive regions (indicated by a MAPQ value of 0).

The three conditions allowed 15 archaic human genomes to be considered. It is worth mentioning that some very interesting genomes, although characterized by a sufficiently high coverage, were discarded for various reasons: the available alignment files for Anzick-1, Kostenki 14, Sunghir 1--4, and Sunghir 6 do not comprise repeated regions; LaBraña lacks data for chromosome 22; Den D2, Hohlenstein Stadel, Scladina, San Teodoro 3, San Teodoro 5, and Zlatý kůň have low coverage and/or high contamination.

Reads from these 15 genomes were remapped to the GRCh38/hg38 (42), and CNV estimation was directly based on hg38. Starting with 19,988 human genes, we organized them into 18,043 genes/gene clusters (see below), which are covered by the 15 genomes and shared with the modern reference human genomes hg38.

Data download and general settings. Reads for the ancient genomes were downloaded either from ENA database or cdna.eva.mpg.de or NCBI with accession IDs in **Table S1**. For genomes available in the ENA browser, all reads in fastq files have been mapped into the GRCh38/hg38 refrence genome. Vindija 33.19 genome is an exception since it displays more than 30B reads and we decided to unmap the BAM files for hg19 available at cdna.eva.mpg.de, and re-map approximately 1.7B reads on hg38. For the genome of Chagyrskaya 8, not available on ENA, we unmapped the BAM files for hg19 available at cdna.eva.mpg.de, and re-mapped approximately 1.2B reads on hg38. All the 79M reads of the Mal'ta genome which were available on NCBI, have been mapped on hg38. Each genome was processed independently up to the final CNV analysis (see below), with the following settings:

- `genome =` "GenomeID"
- `fr_ln` = "Fragment length used for sequencing (might differ from genome to genome)"
- `autosome_sum_length` = "Sum of the lengths of all autosomes in the reference genome (GRC37/hg38)"

Scripts for the different steps of CNV analysis. We developed four main scripts for data preparation towards CNV analysis:

- **read2CNV.slurm**: handles *de novo* mapping, data preparation, and CNV estimation for individual genomes (**Figure 2AB**);
- **CNV_Analyses.slurm**: merges single genome estimations and clusters recent paralogs (**Figure 2C**);
- **genes_with_CNV_at_pop_lvl_extracter.R**: selects genes with differential copy number at the population level (**Figure 2C**);
- **HPA_Info_Extracter.R**: performs gene expression analysis using data from the Human Protein Atlas (**Figure 2D**).

All command lines and programs will be available with the publication. Note that read2CNV.slurm is designed for processing single genomes, while CNV_Analyses.slurm is designed to combine the results from individual genomes.

Step 2. *De novo* mapping, data preparation and CNV estimates

This step is performed using the slurm file **read2CNV.slurm**. It takes as input the reads of a single ancient genome, maps them to the GRCh38/hg38 reference genome and, following the filtering process detailed below, returns the copy number estimate for each gene.

De novo mapping

The reads in .fastq format from each ancient genome were mapped to the human reference GRC37/hgr38 as follows. The output is then processed with a filter on PCR duplicates, on the percentage of identity, and on read length. All parameters were set in accordance with the guidelines given in (192) and (193), and taken up by (53) specifically for the study of ancient genomes.

a - Alignment to the reference genome
Reads were mapped to the reference through `bwa aln` (v. 0.7.17-r1188) (194) with parameters `-l 16500, -n 0.01, -o 2, and -t 48`. Note that it has been shown that bwa mem works worse when mapping reads from ancient genomes (193). Indexed files from hg38 were retrieved from Zenodo database (https://zenodo.org/records/5146236/files/bwa.tar.gz?download=1) and were stored in a single directory named `Data/References/hg38/bwa_2021/hg38`, while fasta sequences were obtained from https://hgdownload2.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.gz.

b - Format conversions
Mapped reads were converted to SAM format through bwa samse with default parameters. Aligned reads were then filtered to exclude unmapped reads and converted to BAM format using Samtools, `samtools view -F 4 -h -Su` (v. 1.6) (195,196). The resulting BAM file was sorted using Samtools sort and saved to `Analyses/mapping/$genome/${genome}.bam`.

Filter out of PCR duplicates. To remove PCR duplicates, the BAM file was processed using Picard's MarkDuplicates tool with `REMOVE_DUPLICATES` and `REMOVE_SEQUENCING_DUPLICATES` set to *true* (http://broadinstitute.github.io/picard/). This process generated a new BAM file without PCR duplicates and a metrics file recording the details of the duplicates removed.

Filter reads with percentage of identity < than 90% and shorter than 30bp. In order to remove reads with a percentage of identity lower than 90% and shorter than 30bp, the BAM file was first converted to BED through bedtools bamtobed (v2.30.0) (197). Read lengths and percentage of mismatches were then retrieved through custom bash scripts. Finally, reads were filtered out with Picard's FilterSamReads and the output was saved in `Analyses/mapping/$genome/${genome}_nopcr_90perc_30bp.bam`. The 90% threshold was chosen based on standard practices in read mapping for both ancient and modern genomes, while the 30bp threshold was selected to capture a larger number of reads. The validity of these parameters was confirmed using the Mezmaskaia 1 genome.

Data preparation

D*ata preparation* is carried out in several steps, creating distinct directories and subdirectories:

1) The BAM file is separated in 24 separate files, one for each chromosome, with numbers from 1 to 22 or letter X/Y. From now on, all the analyses of this section will be performed on each chromosome.
2) The GC-bias for each chromosome is computed and each alignment is corrected based on the results.
3) The per-base depth of coverage for each chromosome is computed.
4) The average depth of coverage for each chromosome is computed and the results are concatenated in a single file.
5) The average depth of coverage is calculated excluding the blacklisted regions from step 3.
6) A .txt file containing the average depth of coverage of the genome, taking into account all autosomes, is generated.
7) A gzipped file containing the depth of coverage of each base of protein coding genes defined in the .bed reference is obtained. This will be the main input for the CNV estimation.

Next-generation sequencing (NGS) technologies can introduce, to varying degrees, uneven coverage of reads resulting from GC bias (198). As our analysis of copy number variation is based on aligned reads and depth of coverage of specific regions, to correct the datasets for such biases is of utmost importance. We therefore used deepTools2 (v. 3.5.4) (199) to compute and correct GC-bias using Benjamini's method (200). More precisely, we used the utility computeGCBias with parameters `–effectiveGenomeSize chr_size` and `-bl blacklist.bed` to compute the GC-bias, where chr_size is the ungapped length of placed scaffolds of the chromosome under consideration in GRCh38/hg38 reference genome (lengths are available in column "Placed scaffolds" of section "Ungapped lengths" at https://www.ncbi.nlm.nih.gov/grc/human/data) and blacklist.bed refers to the ENCODE DAC Blacklisted Regions (201,202) (i.e. a comprehensive set of regions in the human genome that have anomalous, unstructured, high signal/read counts in NGS experiments). Finally, alignment files were corrected for GC-bias using the utility `correctGCBias` with parameter `--effectiveGenomeSize chr_size`. Given a genome G, the depth of coverage (average or per-base) was computed using mosdepth (v. 0.3.3) (203) with default parameters on the GC-bias corrected alignment files. More precisely, given a chromosome chri where *i ∈{1,…,22}*, its average depth of coverage C^i was computed as

$$C^i = \frac{\sum_{j \in chr_i} doc_j^i}{L_{hg38}^i}$$

where doc_j^i is the depth of coverage computed by `mosdepth` at position *j* of chromosome *i,* and L_{hg38}^i is the ungapped length of chromosome *i* in the GRCh38/hg38 reference genome. As analyzed datasets involve male and female individuals, we decided to restrict the copy number variation analysis to autosomes.

A set of protein coding genes G for GRCh38/hg38 was retrieved from GENCODE (204) and genes overlapping any of the ENCODE DAC Blacklisted Regions (201,202) were excluded from the analysis, regardless of the length of the overlap.
The coverage C_g^i of a gene *g* annotated on chromosome *i* is computed as

$$C_g^i = \frac{\sum_{j \in g} doc_j^i}{L_g}$$

where the depth of coverage doc_j^i is computed exclusively at positions *j* of the gene *g* of length L_g . Hence, we estimate the *haploid* copy number of a gene *g* as C_g^i/C^i.

Step 3. CNV estimation

For each genome, CNVs were estimated with a custom python script (computeGeneCopy.py) which is called in the main slurm file. The script uses two intermediate input files:

- `chrom_doc="Analyses/Data_preparation/$genome/avgdepth/nobl/$genome.hg38.nobl.avgdepth.txt"`, which provides the average depth of coverage, calculated by excluding blacklisted regions, with each line corresponding to a chromosome.
- `gene_doc="Analyses/Data_preparation/$genome/$genome.fixed.hg38.genes.protein_coding.coverage.gz"`, which contains the depth of coverage for each base of protein coding genes.

The output, Analyses/CNV_estimates/$CNV_estimates_dir/$genome.nobl.cnv.tsv, contains the CNV estimates for each gene in the genome.

Merging single genome estimates and clustering of recent paralogs

This step is performed using the slurm file **CNV_Analyses.slurm**, which takes as input the gene CNV estimates from each ancient genome ($genome.nobl.cnv.tsv). The process generates two output tables:

- The first table, AllGenomes.type.nobl.cnv.matrix.ALL.tsv, groups genes into clusters of recent paralogs, with each cluster represented as a single line, replacing individual paralog entries. Each column corresponds to an ancient genome, and an additional column provides CNV estimates for genes/clusters of genes in the human reference genome hg38. A total of 18,043 genes/clusters of genes are included.
- The second table, AllGenomes.type.nobl.cnv.matrix.tsv, selects only those genes/clusters of genes that show a different number of copies in at least one genome when compared to hg38. This results in a subset of 2,217 genes/clusters of genes.

The CNVs estimates for each genome were unified in a single file through bash commands. Then, human paralog genes, defined as duplicated copies at the *Homo sapiens* level in ENSEMBL database (column "Ancestral Taxonomy" in ENSEMBL), were clustered together if they had over 80% sequence identity. For exemple, the three copies of the AMY1 gene appeared at the *Homo sapiens* level, remain highly conserved and are merged into a cluster AMY1A/B/C. After clustering, the estimated copy numbers of the "clustered genes" is defined by adding together the copy number estimation of each of the paralogs. Note that, retrieval of human paralogs from the ENSEMBL database was done with a custom R scripts, using the package BiomaRt (205,206).

Definition of population-dependent CNV

This step is performed using the R script **genes_with_CNV_at_pop_lvl_extracter.R**. It takes the input file AllGenomes.type.nobl.cnv.matrix.tsv and generates the output file AllGenomes.type.nobl.cnv.matrix_with_CNV_pop_lvl.tsv, which includes two additional columns, "gain" and "loss", reflecting differences in CNV between *Hn* and/or *Hs* populations. Genes without such differences are discarded, resulting in a final table of 781 genes.

To define population-dependent CNVs, we focused on genes showing an increase or decrease in copy number in at least 5 Neandertals (>50% of the population, where pop=8) and/or in at least 4 ancient *Sapiens* (>50% of the population, where pop=6), compared to the reference human genome hg38. Since only one *Denisova* specimen (Den *3*) was available, it was treated as a point of comparison rather than a population.

Step 5. Analysis of tissue-specific expression and functional relevance

This step is performed using the R script **HPA_Info_Extracter.R**. It uses two downloadable datasets from the Human Protein Atlas (HPA) (44,207–209) as input:

https://www.proteinatlas.org/download/rna_tissue_consensus.tsv.zip for RNA levels, and https://www.proteinatlas.org/download/normal_tissue.tsv.zip for protein expression. Two output files are generated. The first file (all_genes_3organs_protein.tsv) lists the three organs with the highest protein expression for each human gene. The second file (all_genes_3organs_rna.tsv) lists the three organs with the highest RNA expression for each human gene.

All human genes were analyzed using data from HPA. For each gene, the three organs with the highest RNA levels (given as a continous value) and the three organs with the highest protein expression levels (three values are given, low, medium and high) were identified. In cases where multiple organs showed equal expression levels, additional organs were included. The subset of 781 genes with differential CNV at the population level (see above) was combined with the RNA/protein expression analysis. For each gene, we considered the list of organs with highest RNA/protein expression and selected only those associated to at least one digestive organ—the gastrointestinal tract, pancreas, proximal digestive tract, liver and gallbladder. These 489 selected genes were classified as "digestion-related genes".

Filtering for strictly diet-related and accessory diet-related genes using HPA

Each gene was manually reviewed using the HPA database to assess whether it was expressed in a wide range of organs beyond digestive ones. Additionally, “protein function” data from UniProt, accessed through HPA, or "gene description" from Ensembl (when UniProt data was unavailable in HPA) was used to exclude genes with overly generic functions associated with multiple organs (**Figure 2E**). This manual screening resulted in a final set of 55 genes/clusters of genes.

Step 6. Filtering by comparison with modern human CNVs

CNV estimates for the subset of 55 genes in ancient genomes were compared with those from modern populations (**Figure 2F**). This step aimed to ensure that the analysis captured adaptive variations specific to the two ancient populations, excluding genes with common variation in modern populations. An example is provided in the main text for the *PGA3/4/5* gene, and **Figure S3** shows all genes filtered out by this process. This step refined the final set to 50 genes.

Step 7. GO-terms analysis and diet-related literature

All genes chosen based on the Tissue expression and passing the modern CNV estimation filter have been analysed with a GO-term analysis. The 50 selected genes were given as an input to pantherdb.org (210) to compute a first functional classification. Subsequently, the same has been done with DAVID online server (45,211) and the resulting GO-terms were grouped in slim categories with QuickGO online server (212). Subsequentely, each gene/cluster of genes was analysed further based on literature search to decide how to classify them in six functional categories: dietary carbohydrate metabolism, lipid metabolism, liver lipid/brown fat metabolism, immunity/antibiotic action in the digestive tract, environmental related pathways, obesity/diabetes. See **Figure 2GH.**

Estimation of the “absence” of a gene

We estimated that some genes (*UGT2B28, ACOT1, CD24, LYPD8, PECAM1, TCAF2, TCAF2C*, and *TMEM86B*) are not present in some ancient population or individual. For instance, the *ACOT1* gene is not present in *Hn,* Den D3 and Loschbour genomes. In this respect, notice that in the 8 *Hn* genomes, *ACOT1* coverage is estimated to be

at most <0.42 (**Table S5**), which can be explained by the presence of the paralogous gene *ACOT2*, duplicated at the Homininae level, and of many other copies which preceeded the *ACOT2* duplication, such as *ACOT4* with 70% sequence identity from *ACOT1. ACOT1* and *ACOT2* are known to have the same function, *ACOT1* acting in the cytosol and *ACOT2* in mitochondria (213). *ACOT2* codes for a protein of 483 amino acids that contains a 62 amino acid leader sequence at the N-terminal that targets the protein to mitochondria. The *ACOT1* protein is shorter than *ACOT2*, with its 421 amino acids showing 98.81% sequence identity to *ACOT2*, differing by only 5 amino acids. Read mapping to *ACOT2* and *ACOT4* is frequently estimated below 1, which supports the absence of the *ACOT1* gene copy in the genome and suggests incorrect read mapping to *ACOT1*. See (214) for a general description of type 1 and 2 *ACOT* genes. For the other genes: *UGT2B28* has a paralog *UGT2B11* of 94.71% sequence identity, duplicated at the Hominoidea level; *TCAF2* has a paralog *TCAF2C* at 99.58% sequence identity, duplicated at the Apes and Old World Monkey/Catarrhini level. *CD24, LYPD8, PECAM1* and *TMEM86A* do not have recent paralogs within *Homo sapiens* nor sufficiently high sequence identity, leaving their absence unexplained. For example, *TMEM86B* and *TMEM86A* originated from a duplication event at the *Vertebrata* level and exhibit 36.28% and 34.17% sequence identity, respectively, when *TMEM86B* is used as the query and *TMEM86A* as the target. This low sequence identity between these paralogs in the reference human genome is insufficient to support the possibility of incorrect read mapping.

Construction of the multidimensional space defined by CNV

Individuals plotted in the CNV multidimensional space. We represented individuals in a multidimensional space by using gene CNVs as real-valued features.

Principal Components Analysis (PCA). PCA is a well-known computational method used in exploratory data analysis and for making predictive models (50). It is commonly used for dimensionality reduction when the number of features (or dimensions) in a given dataset is too high. We used PCA to project each data point onto only the first few principal components to obtain lower-dimensional data while preserving as much of the variation as possible. The first principal component can be defined as a direction that maximizes the variance of the projected data. The 2nd principal component can be taken as a direction orthogonal to the first principal component that maximizes the variance of the projected data, and so on. Thus, a new dimensional system, where each dimension relies on a combination of the initial dimensions of the data, is generated.

CNVs of the final set of genes were studied from a statistical point of view in two subsets: strictly diet-related (23 genes) and accessories to diet (35 genes, different from the 23 genes above). In both cases, a PCA has been performed to reduce the 23 (or 35) dimensions of the high-dimensional space where *Hn* and *Hs* individuals are represented by their gene CNVs. In **Figure 4**, the position of the reference genome GRGh38/hg38 was plotted for comparison. PCA was performed with the program R (215) with `prcomp` function and `scale` set to *TRUE*. The contribution of each variable to the first three components has been computed through the function *`fviz_contrib`* from the `Factoextra` package (216), and a 2D plot of the two most representative PCs was otained using the `plotly` package (v4.9.3) (217).

The dotted red line is the expected average contribution. A variable with a higher contribution than this is considered important, and was highlighted in **Figure 4CD**. For the visualization, we used the function *`plot_ly`* of the R package `plotly`. With this function the main dimensions of the PCA to be represented cin the n-dimensional space can be chosen. We chose to represent the first three principal components.

The fossils, geographical regions and time period

The fossils/genomes (**Table S2**) of the 15 individuals have been associated to their geographical region and time period (**Figure S5** and **Figure S6**). For this, we defined six different regions of the Asian European belt, from the

most eastern to the most western: Central Asia, Eastern Europe, Central Europe, Northern Europe, Southern Europe, Western Europe. These geographic areas were identified according to (218,219), modelling eco-geographical and morphological data on the Neandertalian territory and variability of mtDNA. To take ecological diversity into account, we added Central Asia and Central Europe to the geographic area already defined in these two publications. Indeed, since the latter studies, new evidence for the great *Hn* extension in Central Asia has been collected (from Altai Neandertal and Chagyrskaya caves). Also, we reduced the "Northern Europe" area by splitting it into two regions and adding an eastern one, named "Central Europe", where the climate was different. This geographical division is used here for *Hn* and *Hs*. We considered seven time periods: 120k-90k ybp, 90k-50k ybp, 50k-45k ybp, 45k-40k ybp, 40k-35k ybp, 30k-20k ybp, and 10k-5k ybp.

Supplementary Text 3

A description of the 19 strictly diet-related genes

In the following section, refer to **Table 1** for the list of diet-related genes identified in our study, and to **Figure 3** for a direct CNV comparison between ancient individuals and those from modern human populations.

Duplication of three gene clusters and one gene adapted to carbohydrate digestion

AMY1 is highly duplicated on the human chromosome 1 and we refer to it as a gene cluster, ***AMY1A/B/C***. *AMY1A/B/C* is expressed in salivary glands (**Figure 2E**) and protein expression of *AMY1C* has been also observed in the pancreas (together with the pancreatic amylase gene *AMY2*). Amylases **efficiently break down starch into maltose** in the oral cavity (7,9). The CNV of *AMY1* has an observed range of 2–18 copies per person (5,8–10) and an average of 6 copies per person (25). *AMY1A/B/C* duplication has been estimated to have emerged after the divergence between human-specific lineages and archaic hominins either approximately 550-590k ybp (5,9) or around 400-275k ybp (9,10,19,107), a reconstruction that agrees with recent analyses of genomes from Cameroon (220). Across the 15 ancient genomes studied here, our data confirm that *AMY1* is highly duplicated in *Hs,* whereas two copies are present in most *Hn* and *Den* (**Table 1AB** and **Figure 3**). More precisely, the trend in *Hs* since Ust'Ishim dated at 45k ybp, including present-day humans, is of copy numbers greater than 3. This particularly high copy number is additional evidence that *AMY1A/B/C* was duplicated after the ancestral lineages split. Interestingly, the late Palaeolithic hunter-gatherer from Motala 12 (Sweden) has two copies of the cluster, although Mal'ta, the easternmost Upper Palaeolithic modern human who lived 24k ybp in severe climatic conditions in Central Asia, has only two copies of the *AMY1* gene. On the other hand, a late *Hn* in our sample, Les Cottés ZA-1514 has 3 copies.

SULT1A3 and ***SULT1A4***, collectively called *SULT1A3/4*, encode identical protein products (221). Previous genomic studies revealed the duplication of these genes during the evolutionary process (222,223). From our analysis, all the EAHs have only one copy of *SULT1A3/4* whereas all the *Hs*, from the oldest to present-day humans, have 2 or 3 copies of this gene cluster. *SULT1A3/4* is a **sulfotransferase enzyme** catalyzing the sulfate conjugation of many hormones, neurotransmitters, drugs, and xenobiotic compounds, and so its function depends on its tissue location (224). Highly expressed in the gastrointestinal tract of humans and closely-related primates, *SULT1A3/4* indirectly **influences carbohydrate metabolism** (225) through dopamine sulfonation, a reaction involved in the regulation and biotransformation of catecholamines (226). The gene cluster may therefore detoxify potentially lethal dietary monoamines in the intestine and eliminate toxic compounds from the body (227). A diet that includes a large intake of starch, nuts and seeds, is rich in flavonoids, which boost the levels of serotonin and dopamine, two key neurotransmitters regulating many physiological processes under hormonal control.

SLX1A/B encodes proteins that play a crucial role in **maintaining genome stability**, DNA repair, homologous recombination, and telomere maintenance (228). These proteins form the catalytic subunit of the *SLX1-SLX4* structure-specific endonuclease complex, which is essential for resolving DNA secondary structures that arise during repair and recombination processes. They exhibit read-through transcription with the adjacent carbohydrates-related *SULT1A3* gene. Note that the CNV pattern observed in *SLX1A/B* is the same as for *SULT1A3/4*, possibly suggesting a co-evolution between gene clusters.

UGT2B28 encodes a protein that belongs to the uridine diphosphoglucuronosyltransferase family. This enzyme facilitates the **transfer of glucuronic acid** from uridine diphosphoglucuronic acid to various substrates, including steroid hormones and lipid-soluble drugs, through the process of glucuronidation, an intermediate step in steroid metabolism. Dietary intake can influence this metabolic pathway, as certain foods, particularly those **rich in flavonoids**, can affect glucuronidation activity. Consuming a diet high in fruits, vegetables, and other flavonoid-rich foods may enhance the body's ability to metabolize and eliminate various substances, potentially impacting the activity of *UGT2B28*. **Flavonoids have been found to interfere with the digestion, absorption, and metabolism of carbohydrates** (229). Also, significant amounts of flavonoids can be found in leafy vegetables, onions, apples, berries, cherries, soybeans, and citrus fruits (230), tea and wine.

Gene copy number for adaptation to fatty food of animal origin. A large number of genes directly or indirectly related to high intake of **animal fat** present a differentiated CNV in EAHs (**Table 1A**). A few genes are differentiated either in *Hn* (*LPA, ACOT1, ZNG1E/F/C*) or in *Hs* (*ACOX3, HAND1*), but the vast majority have a higher CNV for both *Hn* and *Hs* than the modern reference genome.

LPA, CLPS and ***CLPSL1*** all **encode enzymes involved in lipid hydrolysis and oxidation of long-chain FA** (i.e., arachidonic acid metabolism). More FA are consumed in diets incorporating more animal food and are critical for brain development and metabolism and for inflammatory responses (91). *CLPS*, *CLPSL1* and *LPA* genes are **involved in lipid digestion and lipid transport** (231). Colipases (*CLPS* and its paralog *CLPSL1*, Colipase Like 1) are cofactors needed by pancreatic lipase for efficient dietary lipid hydrolysis as they allow the lipase to anchor itself to the lipid-water interface and avoid being washed off by bile salts (232). Lipases (*LPA*) are mainly expressed in the liver and perform essential roles in digestion, transport and processing of dietary lipids. There is huge inter-individual variation in *LPA* expression levels and a high heritability of the traits, and while the physiological function of *LPA* is not fully elucidated, it is acknowledged to be linked to the risk of cardiovascular disease (233). All *Hn*, almost all *Hs* and *Den* individuals display a positive CNV trend for *CLPS* and *CLPSL1*. Most *Hn* individuals and the early *Hs*, Peştera Muierii 1 and Malta, have multiple copies of *LPA*. The oldest *Hs*, Ust'Ishim, along with later *Hs*, Loschbour, Motala 12 and Stuttgart, exhibit the same pattern as the modern reference genome—a single copy. Notably, most modern individuals in our study show duplication of *LPA*.

ACOT genes are involved in **lipid metabolism** as they encode acyl-CoA thioesterase that regulates the cellular balance between free FA and acyl-CoA, feeding into key pathways for energy expenditure and neuronal function (49). In contrast with *Hs*, none of the *Hn* individuals has the *ACOT1* gene whereas copies of all other type I *ACOT* genes (214), whether peroxisomal or mitochondrial, are present in *Hn*, and most *Hs*. *ACOT2* is 98.6% identical (213) to *ACOT1* and is present as one copy in all *Hs*, *Den* and *Hn* individuals. The *ACOT2* protein localizes to the mitochondrial matrix and shares the same function as *ACOT1* by hydrolysing long-chain fatty acyl-CoA into free fatty acid (FA) and CoASH. It has been demonstrated that *ACOT1* regulates fasting hepatic FA metabolism by balancing oxidative flux and capacity (84). Hence the absence of *ACOT1* could have been lethal for EAHs in periods when FA intake was scarce, such as during the colder stages of marine isotope stage (MIS) 3 (chiefly Heinrich Event 4) when a strong reduction in herds of large herbivores deprived EAHs of their basic food (234,235). Conversely, *Hs* would have had broader access to different kinds of food. An exception is the individual from Loschbour who lacked this gene.

RFPL4A/L1 produces a protein **metal-binding zinc**, an essential mineral, and the ***ZNG1E/F/C*** cluster, crucial for **zinc transfers**, exhibits a notable trend of increased CNV in *Hn* and certain *Hs*. Notably, the richest sources of zinc are animal products and oysters (71) compared to plant foods. However, pinenuts, seeds such as hemp, sesame, or chickpeas and lentils among legumes are also rich in zinc and were available in the ecological realm of both *Hn* and *Hs*.

ACOX3 (Acyl-Coenzyme A oxidase 3, also known as pristanoyl-CoA oxidase) possibly facilitated **strong intake of animal fat**. Namely, the expression of *ACOX3* has been documented to increase in response to a high-fat diet (236) and to be related to fat metabolism (237). Both *Hs* and a large number of *Hn* were equipped with multiple copies of this gene.

HAND1, especially duplicated in *Hs* and in the latest *Hn* (Mezmaiskaya 2, Les Cottés et Goyet), has been shown to inhibit the conversion of cholesterol to steroids in trophoblasts, playing a crucial role in the development and physiological function of the human placenta (96). Further analysis of this gene could provide valuable insights into the relationship between **diet and reproduction** in the evolution of the Homo genus.

PDX1 is a transcription factor active in the **pancreas,** where the protein produced by this gene plays as transcriptional activator of several genes, including insulin, somatostatin, glucokinase, islet amyloid polypeptide, and glucose transporter type 2. It is essential for pancreatic β-cell development and adult β-cell function (238). It has been documented to **mediate lipid metabolism** gene networks in intestinal cells (239). In high-fat diets, **saturated fatty acids** entrap *PDX1* in stress granules and impede islet beta cell function (240). An increased production of *PDX1* due to the higher number of copies of the gene could facilitate the activity of *PDX1* in enhancing the efficiency of this enzyme activity in *Den*, *Hn* and *Hs*.

WFS1 encodes for a protein called wolframin that is thought to **regulate the amount of calcium** in cells. It is expressed in all the endoplasmic reticulum of all the cells but shows its highest levels in brain, pancreas, heart, and insulinoma beta-cell lines tissues. Due to the varied and fundamental functions played by calcium in cellular regulation, it plays a fundamental role in cell-to-cell communication, muscles contraction, triggering the myosin and actin filaments to slide into each other, and protein processing. *WFS1*'s association with diet has been elucidated in (72), where it was demonstrated that heterozygous *WFS1* mice, possessing only one functional gene copy, exhibited metabolic impairments when subjected to a high-fat diet. The presence of multiple copies could facilitate individuals following high-fat regimes such as *Hn*, *Hs*, and *Den*.

H2AC11 is a member of the histone deacetylase family and it is involved in the biological functions of almost every system of the human body and its duplication may play a role in tumorigenesis (241). **Deficiency in *H2AC11* prevents the onset of obesity and metabolic syndrome induced by a high-fat diet** (242). Surprisingly, we observe a duplication in all *Hs* and *Den*, while we see a heterogeneous pattern with regard to *Hn*, suggesting that the role of this gene in diet needs to be further investigated.

F7 (coagulation factor *F7*) encodes for an enzyme of the serine protease class. It is part of a group of proteins produced in the liver and it is vitamin K-dependent. They are involved in the **coagulation system**, a series of chemical reactions that form blood clots. *F7* is the most common of the hemophilia-like bleeding disorders. **Quick time for coagulation**, induced by the duplication of the gene, may have provided a leap to both *Hn* and *Hs*, who might have inherited these traits while breeding with *Hn* (75).

Under possible harming conditions that *Hn*, *Den* and *Hs* might have encountered, the capacity of quickly healing the wound so that being less exposed to infection, ending with a higher probability to survive and thrive. Yet, under far less dangerous settings, hypercoagulation might have turned into a higher risk for stroke, embolism and troublesome pregnancies (243). *F7* has also been documented to be **linked to fat intake and cholesterol levels** in mice (98). The impact of diet on *F7* activity is reported in (99).

Adaptation to cold stress metabolism related to eating food of animal origin. Cold conditions are expected to be better tolerated by human populations that react better to cold stress. We found that both *Hn* and *Hs* populations show a positive CNV trend for a gene involved in this metabolic process – ***MST1*** - compared to the modern human genome. We noted that the trend was due to the positive differential CNV of this gene in the Holocene *Hs* individuals, while the three early *Hs*, Ust'Ishim, Peştera Muierii 1 and Mal'ta, belonging to the earliest migration wave out of Africa about 60ka, have single copies of each gene. One more gene shows higher CNV in *Hs*, ***H1.2***, and one lower, ***CD24***.

MST1 (Macrophage stimulating 1) gene is known to be important in **liver lipid metabolism**, limiting liver damage that might be induced by a high-fat diet or during fasting (244). This gene also plays a role in **cellular responses to hypoxia**, **oxygen supply** and **thermogenic metabolism**. Other genes, *EPAS1* (endothelial PAS domain containing protein 1) and *EGLN1* (Egl-9 family hypoxia inducible factor 2) have previously been highlighted as being involved in the physiological response to oxygen depletion in ancient populations (69). Indeed, tolerance of lower oxygen availability was observed in modern Andean and Tibetan people who have hemoglobin with adapted oxygen affinities (245,246). When the *EPAS1* gene was identified in *Den* DNA it was proposed that the capacity of Tibetan people to thrive in low-oxygen or high-altitude conditions was possibly inherited through introgression of Den-like DNA (69). *MST1* is duplicated in *Den*, all *Hn* and Mesolithic *Hs* individuals.

***H1.2*, highly expressed in adipocytes,** regulates thermogenic genes in white adipose tissue (WAT) and affects energy expenditure. Its deletion promotes browning of WAT (brown adipose tissue - BAT) and **improves cold tolerance**, while its overexpression has the opposite effect (66). In cold environments, it has been observed a higher expression of *H1.2*, sensing thermogenic stimuli in WAT and BAT (66). Notably, we observe an increase in the copy number of this gene in the most part of *Hs* while it remains at a single copy in half of *Hn*, indicating a plausible adaptation to different environments.

CD24 gene loss, observed in the older *Hn* (Altai Neandertal D5, Chagyrskaia 8) and Vindija 33.19, and most *Hs*, has been linked to a **reduction of white adypocyte tissue** (67).

A description of the 31 accessory to diet genes

High fat diets are known to be linked to several diseases including obesity, diabetes, fatty liver, inflammatory bowel disease and colon cancer (247). The set of 31 accessory genes is linked to these diseases. In the following section, refer to **Table 2** for the list of accessory genes identified in our study, and to **Figure S2** for a direct CNV comparison between ancient individuals and those from modern human populations.

Inflammation/Immune response

BOLA2/2B is involved in the immune response in the gut and may be associated with an increased risk of digestive disorders such as **inflammatory** bowel disease. It is known to be involved also in the maturation of cytosolic iron-sulfur proteins. Its duplication modifies iron homeostasis and its deletion is associated with anemia as it is active in the iron regulation. It is conjectured that its emergence is unique to *H. sapiens* and dated back to 282 kyr (24) and its expansion in modern populations, has a potential adaptive role in protecting against iron deficiency crucial during embryonic development (97). No *Hn* nor *Den* show duplication of this gene, which is duplicated in all *Hs*, including modern individuals and the reference human genome. Notably, copy number variation is highly diverse within modern human populations.

ORM1 encodes a key acute phase plasma protein, classified as an acute-phase reactant due to its increase during acute inflammation, and it may be involved in aspects of **immunosuppression**. ***ORM2*** (ancient paralogue of *ORM1*) encodes an acute-phase plasma protein with a role in **immunosuppression during acute inflammation**. *ORM1* and *ORM2* **regulate lipid homeostasis** and protein quality control (248).

TPSAB1/B2 encodes alpha-tryptase, the major neutral protease in mast cells, playing a role in **innate immunity**. Mast cell tryptases have proposed roles in allergic inflammation and host defense against infection (249). Duplication of this gene can cause severe gastrointestinal disorders and stomach disorders (250). Hereditary alpha tryptasemia syndrome (HαT) is an autosomal dominant genetic disorder caused by an increased number of copies of the *TPSAB1* gene. This condition is characterized by elevated basal serum tryptase levels. The syndrome presents with symptoms that can affect multiple organ systems (250). Nearly all *Hn*, *Hs*, and *Den* have a single copy of this gene, in contrast to the modern human reference genome and all modern human populations, which dispay two copies.

GRTP1 is associated with poor prognosis in stomach and colon cancer when expressed at low levels (251). Most expressed in liver and intestine, poorly studied and identified as participating in the multigene signature for **non-alcoholic fatty liver disease** (NAFLD) (252). It is duplicated in *Den*, *Hn* (with the exception of Spy 94a), and all *Hs*.

CLPTM1L (Cleft lip and palate transmembrane protein 1-like protein) overexpression can induce apoptosis, and polymorphisms in this gene increase susceptibility to various cancers, including lung, pancreatic, and breast cancers. It participates in the **biosynthesis of glycosylphosphatidylinositol and lipid transport** (253). Most expressed in liver and pancreas. Nearly all *Hn* individuals show duplication of this gene.

ZNF431 is a member of the Krueppel C2H2-type zinc-finger family of transcription factors, the largest transcription factor family in the mammalian genome. It may negatively regulate transcription of target genes and mutations may be linked to non-syndromic facial clefting. It has been observed that homozygous mutations in the *ZNF431* gene, inherited in an autosomal recessive manner, result in reduced STAT3 levels, producing a phenotype that closely resembles autosomal dominant hyper IgE syndrome (254). Hyper IgE syndromes (HIES) are rare forms of **primary immunodeficiencies** (PI) characterized by recurrent eczema, skin abscesses, lung infections, eosinophilia (high numbers of eosinophils in the blood), and high serum levels of immunoglobulin E (IgE). *Den*, most *Hn,* and *Hs* show duplications.

C4A/B encodes complement factors essential for the binding of the CR1 receptor with the preformed and nascent immune complexes (255,256), hence with roles in **immune complex solubilization and clearance**. Deficiency is linked to systemic lupus erythematosus. *Den*, most *Hn,* and *Hs* exhibit a high copy number (at least three copies), whereas all modern human populations consistently show two copies.

N4BP3 enhances **antiviral innate immune** signaling by promoting polyubiquitination of MAVS. It plays a crucial role in the RIG-I-like receptor (RLR)-mediated innate immune response by targeting MAVS, shedding light on the mechanisms underlying innate antiviral responses (257). Most *Hn* and *Hs* exhibit duplication of this gene.

LYPD8 **prevents invasion of Gram-negative bacteria** in the colon epithelium by binding to bacterial flagella (intestinal homeostasis) (258,259). *Hs* and the older *Hn* are depleated of this gene.

STAB1 plays a role in host defense mechanisms against infections and regulates **anti-inflammatory responses**, potentially influencing atherosclerosis (260,261). Most *Hn* and *Hs* exhibit duplication of this gene.

PECAM1 (platelet endothelial cell adhesion molecule-1) is a 130-kDa transmembrane glycoprotein expressed on blood and vascular cells. It acts as a negative regulator of **nonalcoholic fatty liver disease** progression, nonalcoholic steatohepatitis (NASH) (262). This study shows that **genetic deficiency** of *PECAM1* potentiates the development and progression of NASH. The NASH disease spectrum ranges from **accumulation of fat** in hepatocytes (steatosis) to the presence of an **inflammatory infiltrate and fibrosis**, and ultimately to progressive fibrosis and cirrhosis. *Hs* and the older *Hn* are depleated of this gene.

UQCRFS1 is involved in mitochondrial respiratory chain complex assembly and function, with deficiencies linked to mitochondrial disorders (263). A diet-dependent proteome-wide differences in adipose tissues, highlighted down regulation of oxidative phosphorylation genes such as *UQCRFS1* in response to **dietary cholesterol** (264). Multiple copies may benefit individuals with dietary cholesterol and are observed in *Den*, as well as in nearly all *Hn* and *Hs*.

IFNL2/3 are cytokines with **antiviral, antitumor, and immunomodulatory activities** that play a crucial role in antiviral host defense, particularly in epithelial tissues (265). They act as ligands for the heterodimeric receptor composed of *IL10RB* and *IFNLR1*, activating the JAK/STAT signaling pathway to induce IFN-stimulated genes (ISGs) and mediate the antiviral state. Their action is primarily restricted to epithelial cells due to the cell-specific expression of the *IFNLR1* receptor and they significantly contribute to antiviral defense in the intestinal epithelium and up-regulate MHC class I antigen expression. All modern populations dispay duplications of this gene, while all ancient individuals exhibit only one copy.

Environment and nutrition

TCAF2 and ***TCAF2C*** are *TRPM8* channel associated factors. They enable transmembrane transporter binding activity and are involved in the regulation of anion channel activity, cell migration, and protein targeting to membranes, with a localization in cell junctions and the plasma membrane. *TCAF2* and ***TCAF2C*** are ancestrally related within *Homo sapiens*, but their sequence similarity does not meet our

strict threshold. *TCAF2* is longer than *TCAF2C*, with the two genes showing ENSEMBL Target_id of 99.58 and a Query_id of 77.15. This suggests that *Hn* likely has only 1 copy of the two genes, although it is unclear which of the two, while *Hs* possesses both copies. Research on the **deletion** of the **cold thermoreceptor** TRPM8, which is related to *TCAF2* and *TCAF2C*, shows **increased heat loss and food intake, leading to reduced body temperature and obesity** in mice (266). It has been established that *TCAF* duplications originated ~1.7 million years ago but diversified only in *Hs* by recurrent structural mutations. Multiple evidence supported the hypothesis that *TCAF* diversification among hominins was likely due to positive selection, in **response to cold or dietary adaptations** (68). Both *TCAF2* and *TCAF2C* are absent in *Hn* and *Den*, yet present in *Hs*. Interestingly, modern human populations align with ancient individuals regarding *TCAF2*. For *TCAF2C,* it is noteworthy that ethnic groups from South Asia similarly lack this gene, as seen in *Den* and *Hn*.

TMEM86B encodes an enzyme that **degrades lysoplasmalogen**, a group of lipids, impacting cell membrane dynamics. Plasmalogens are found in various **foods** such as scallops, squid, mussels, octopus but also **big game cervids** (267–270). Decrease in plasmalogen is related to Alzheimer and Parkinson diseases. All ancient individuals lack this gene, which is present in all modern populations.

ATP4B is essential for **gastric acid secretion**, encoding the beta subunit of the gastric H+, K+-ATPase proton pump, which is vital for exchanging H+ and K+ ions across the plasma membrane (271). All ancient individuals exhibit a duplication of this gene.

HGFAC, a member of the peptidase S1 protein family, is **nutritionally regulated** and plays a critical role in **maintaining carbohydrate and lipid homeostasis** by activating hepatocyte growth factor (HGF) (272). All ancient individuals exhibit a duplication of this gene.

SLC9A3 is crucial for pH regulation and signal transduction, encoding an **epithelial Na/H exchanger** implicated in congenital secretory sodium diarrhea, bladder fibrosis, and dysfunction due to systemic homeostasis deregulation (273). Most ancient individuals exhibit a duplication of this gene.

Obesity/Diabete

FGFRL1 has a negative effect on cell proliferation and shows significant methylation level differences that **correlate with obesity** status (274). All ancient individuals dispay duplications of this gene, with the exception of Spy 94a.

DRD5, a dopamine receptor gene, has been **linked to obesity**, with studies showing altered dopamine levels and receptor expression in adipocytes of obese individuals (275). All ancient individuals dispay duplications of this gene, with the exception of Goyet Q56.1.

NKX6-1, a homeobox gene, is essential for the development and function of pancreatic beta cells, and its inactivation causes beta cell dysfunction and hypoinsulinaemia, affecting insulin synthesis and secretion (276). Many ancient individuals display duplications of this gene; however, the oldest *Hn*, Altai Neandertal and Chagyrskaya 8, and Den D3 do not exhibit duplications.

HAND2, a basic helix-loop-helix transcription factor, plays a crucial role in cardiac and limb development and has been identified as an **obesity-linked** adipocyte transcription factor (277). It is duplicated in most *Hs*.

H4C16 (H4 clustered Histone 5 and Histone 16) is involved in central **carbon metabolism regulation**, and its acetylation status influences **diet-induced obesity** in mice (278). Most *Hs* exhibit a very high duplication number, with at least three copies.

IDUA encodes an enzyme essential for the lysosomal degradation of glycosaminoglycans, with mutations leading to mucopolysaccharidosis type I, and has been shown to **prevent diet-induced obesity** in mice (279). All ancient individuals display duplications of this gene.

ATP5ME, part of mitochondrial ATP synthase, catalyzes ATP synthesis during oxidative phosphorylation and has been linked to attenuating type 1 GDM-induced fetal myocardial injury (280). Note that dysfunction of **mitochondria of adipose tissues** results in detrimental effects **on adipocyte differentiation**, lipid metabolism, insulin sensitivity, oxidative capacity, and thermogenesis, which consequently lead to metabolic diseases (281). All ancient individuals, except *Den*, display duplications of this gene.

SEMA3B is part of a gene family known to **regulate obesity** by influencing adipogenesis (282). Most *Hn* display duplications of this gene.

IFRD2 deficiency, along with *IFRD1*, results in severely reduced adipose tissue and resistance to **high-fat diet-induced obesity** in mice (283). Most *Hs* display duplications of this gene.

IER3 (***IEX-1***) deficiency induces the browning of white adipose tissue and provides **resistance to diet-induced obesity**, highlighting its role in expanding adipocyte precursor cells (284,285). Most *Hs* display duplications of this gene.

PLXND1 determines body fat distribution by regulating the collagen microenvironment in visceral adipose tissue, playing a significant role in **obesity and fat distribution** (286). Most *Hs* display duplications of this gene.

References

1. Rito T, Vieira D, Silva M, Conde-Sousa E, Pereira L, Mellars P, et al. A dispersal of Homo sapiens from southern to eastern Africa immediately preceded the out-of-Africa migration. Scientific reports. 2019;9(1):4728.
2. Mann J, Cummings JH, Englyst HN, Key T, Liu S, Riccardi G, et al. FAO/WHO scientific update on carbohydrates in human nutrition: conclusions. European journal of clinical nutrition. 2007;61(1):S132–7.
3. Speth JD. The Paleoanthropology and Archaeology of Big-Game Hunting [Internet]. New York, NY: Springer New York; 2010 [cited 2024 Sep 27]. (Eerkens J, editor. Interdisciplinary Contributions to Archaeology). Available from: http://link.springer.com/10.1007/978-1-4419-6733-6
4. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. Nature genetics. 2007;39(10):1256–60.
5. Perry GH, Kistler L, Kelaita MA, Sams AJ. Insights into hominin phenotypic and dietary evolution from ancient DNA sequence data. Journal of human evolution. 2015;79:55–63.
6. Buckley MT, Racimo F, Allentoft ME, Jensen MK, Jonsson A, Huang H, et al. Selection in Europeans on fatty acid desaturases associated with dietary changes. Molecular biology and evolution. 2017;34(6):1307–18.
7. Butterworth PJ, Warren FJ, Ellis PR. Human α-amylase and starch digestion: An interesting marriage. Starch Stärke. 2011 Jul;63(7):395–405.
8. Hardy K, Brand-Miller J, Brown KD, Thomas MG, Copeland L. The Importance of Dietary Carbohydrate in Human Evolution. The Quarterly Review of Biology. 2015 Sep;90(3):251–68.
9. Inchley CE, Larbey CD, Shwan NA, Pagani L, Saag L, Antão T, et al. Selective sweep on human amylase genes postdates the split with Neanderthals. Scientific reports. 2016;6(1):37198.
10. Carmody RN, Dannemann M, Briggs AW, Nickel B, Groopman EE, Wrangham RW, et al. Genetic evidence of human adaptation to a cooked diet. Genome Biology and Evolution. 2016;8(4):1091–103.
11. Richards MP, Trinkaus E. Isotopic evidence for the diets of European Neanderthals and early modern humans. Proc Natl Acad Sci USA. 2009 Sep 22;106(38):16034–9.
12. Bocherens H, Drucker DG, Madelaine S. Evidence for a 15N positive excursion in terrestrial foodwebs at the Middle to Upper Palaeolithic transition in south-western France: Implications for early modern human palaeodiet and palaeoenvironment. Journal of Human Evolution. 2014;69:31–43.
13. Churchill SE. Thin on the ground: Neandertal biology, archeology, and ecology [Internet]. John Wiley & Sons; 2014 [cited 2024 Sep 27]. Available from: https://books.google.com/books?hl=en&lr=&id=63TCBwAAQBAJ&oi=fnd&pg=PA179&dq=Thin+on+the+ground:+Neandertal+biology,+archeology,+and+ecology+&ots=bZr4uR-cRh&sig=BFE4flF9037qiEcO30EM6oEirrk
14. Drucker DG, Naito YI, Péan S, Prat S, Crépin L, Chikaraishi Y, et al. Isotopic analyses suggest mammoth and plant in the diet of the oldest anatomically modern humans from far southeast Europe. Scientific reports. 2017;7(1):6833.
15. Henry AG, Brooks AS, Piperno DR. Microfossils in calculus demonstrate consumption of plants and cooked foods in Neanderthal diets (Shanidar III, Iraq; Spy I and II, Belgium). Proc Natl Acad Sci USA. 2011 Jan 11;108(2):486–91.
16. Power RC, Salazar-García DC, Rubini M, Darlas A, Harvati K, Walker M, et al. Dental calculus indicates widespread plant use within the stable Neanderthal dietary niche. Journal of Human Evolution. 2018;119:27–41.
17. Hardy K, Bocherens H, Miller JB, Copeland L. Reconstructing Neanderthal diet: The case for carbohydrates. Journal of Human Evolution. 2022;162:103105.

18. Dodat PJ, Albalat E, Balter V, Couture-Veschambre C, Hardy M, Henrion J, et al. Diverse bone-calcium isotope compositions in Neandertals suggest different dietary strategies. Journal of Human Evolution. 2024;193:103566.
19. Mathieson S, Mathieson I. FADS1 and the timing of human adaptation to agriculture. Molecular biology and evolution. 2018;35(12):2957–70.
20. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010;464(7289):704–12.
21. Iskow RC, Gokcumen O, Lee C. Exploring the role of copy number variants in human adaptation. Trends in Genetics. 2012;28(6):245–57.
22. Saitou M, Masuda N, Gokcumen O. Similarity-Based Analysis of Allele Frequency Distribution among Multiple Populations Identifies Adaptive Genomic Structural Variants. Heyer E, editor. Molecular Biology and Evolution. 2022 Mar 2;39(3):msab313.
23. Sudmant PH, Mallick S, Nelson BJ, Hormozdiari F, Krumm N, Huddleston J, et al. Global diversity, population stratification, and selection of human copy-number variation. Science. 2015 Sep 11;349(6253):aab3761.
24. Nuttle X, Giannuzzi G, Duyzend MH, Schraiber JG, Narvaiza I, Sudmant PH, et al. Emergence of a Homo sapiens-specific gene family and chromosome 16p11. 2 CNV susceptibility. Nature. 2016;536(7615):205–9.
25. Carpenter D, Mitchell LM, Armour JAL. Copy number variation of human AMY1 is a minor contributor to variation in salivary amylase expression and activity. Hum Genomics. 2017 Dec;11(1):2.
26. Rice AM, McLysaght A. Dosage-sensitive genes in evolution and disease. BMC Biol. 2017 Dec;15(1):78.
27. Almal SH, Padh H. Implications of gene copy-number variation in health and diseases. Journal of human genetics. 2012;57(1):6–13.
28. Shao X, Lv N, Liao J, Long J, Xue R, Ai N, et al. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. BMC Med Genet. 2019 Dec;20(1):175.
29. Mallick S, Micco A, Mah M, Ringbauer H, Lazaridis I, Olalde I, et al. The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. Scientific Data. 2024;11(1):182.
30. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science. 2012 Oct 12;338(6104):222–6.
31. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505(7481):43–9.
32. Mafessoni F, Grote S, De Filippo C, Slon V, Kolobova KA, Viola B, et al. A high-coverage Neandertal genome from Chagyrskaya Cave. Proc Natl Acad Sci USA. 2020 Jun 30;117(26):15132–6.
33. Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature. 2014;514(7523):445–9.
34. Prüfer K, De Filippo C, Grote S, Mafessoni F, Korlević P, Hajdinjak M, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science. 2017 Nov 3;358(6363):655–8.
35. Hajdinjak M, Fu Q, Hübner A, Petr M, Mafessoni F, Grote S, et al. Reconstructing the genetic history of late Neanderthals. Nature. 2018;555(7698):652–6.
36. Svensson E, Günther T, Hoischen A, Hervella M, Munters AR, Ioana M, et al. Genome of Peştera Muierii skull shows high diversity and low mutational load in pre-glacial Europe. Current Biology. 2021;31(14):2973–83.
37. Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature. 2014;505(7481):87–91.
38. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014;513(7518):409–13.

39. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, et al. Genome-wide patterns of selection in 230 ancient Eurasians. Nature. 2015;528(7583):499–503.

40. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 2016 Oct 13;538(7624):201–6.

41. Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. Science. 2020 Mar 20;367(6484):eaay5012.

42. Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen HC, Kitts PA, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome research. 2017;27(5):849–64.

43. Consortium GO. The gene ontology resource: 20 years and still GOing strong. Nucleic acids research. 2019;47(D1):D330–8.

44. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015 Jan 23;347(6220):1260419.

45. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4(1):44–57.

46. UniProt: the Universal Protein knowledgebase in 2023. Nucleic Acids Research. 2023;51(D1):D523–31.

47. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010;468(7327):1053–60.

48. Bolognini D, Halgren A, Lou RN, Raveane A, Rocha JL, Guarracino A, et al. Recurrent evolution and selection shape structural diversity at the amylase locus. Nature [Internet]. 2024 Sep 4 [cited 2024 Oct 1]; Available from: https://www.nature.com/articles/s41586-024-07911-1

49. Hsieh P, Vollger MR, Dang V, Porubsky D, Baker C, Cantsilieris S, et al. Adaptive archaic introgression of copy number variants and the discovery of previously unknown human genes. Science. 2019 Oct 18;366(6463):eaax2083.

50. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Phil Trans R Soc A. 2016 Apr 13;374(2065):20150202.

51. Saitou M, Gokcumen O. An Evolutionary Perspective on the Impact of Genomic Copy Number Variation on Human Health. J Mol Evol. 2020 Jan;88(1):104–19.

52. Yan SM, Sherman RM, Taylor DJ, Nair DR, Bortvin AN, Schatz MC, et al. Local adaptation and archaic introgression shape global diversity at human structural variant loci. eLife. 2021 Sep 16;10:e67615.

53. Söylev A, Çokoglu SS, Koptekin D, Alkan C, Somel M. CONGA: Copy number variation genotyping in ancient genomes and low-coverage sequencing data. Robinson-Rechavi M, editor. PLoS Comput Biol. 2022 Dec 14;18(12):e1010788.

54. Usher CL, Handsaker RE, Esko T, Tuke MA, Weedon MN, Hastie AR, et al. Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. Nat Genet. 2015 Aug;47(8):921–5.

55. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006 Nov;444(7118):444–54.

56. Perry GH, Tchinda J, McGrath SD, Zhang J, Picker SR, Cáceres AM, et al. Hotspots for copy number variation in chimpanzees and humans. Proc Natl Acad Sci USA. 2006 May 23;103(21):8006–11.

57. Gokcumen O, Babb PL, Iskow RC, Zhu Q, Shi X, Mills RE, et al. Refinement of primate copy number variationhotspots identifies candidate genomic regions evolving under positive selection. Genome Biol. 2011 May;12(5):R52.

58. Wadley L, Backwell L, d'Errico F, Sievers C. Cooked starchy rhizomes in Africa 170 thousand years ago. Science. 2020 Jan 3;367(6473):87–91.

59. Revedin A, Aranguren B, Becattini R, Longo L, Marconi E, Lippi MM, et al. Thirty thousand-year-old evidence of plant food processing. Proc Natl Acad Sci USA. 2010 Nov 2;107(44):18815–9.

60. El Zaatari S, Grine FE, Ungar PS, Hublin JJ. Neandertal versus modern human dietary responses to climatic fluctuations. PloS one. 2016;11(4):e0153277.

61. Longo L, Altieri S, Birarda G, Cagnato C, Graziani V, Obada T, et al. A Multi-Dimensional Approach to Investigate Use-Related Biogenic Residues on Palaeolithic Ground Stone Tools. Environmental Archaeology. 2021 Sep 28;1–29.

62. Longo L, Skakun NN, Pantyukhina IE, Terekhina VV, Sorrentino G. Aurignacian grinding stone from Surein I (Crimea):"trace-ing" the roots of starch-based diet. Journal of Archaeological Science: Reports. 2021;38:102999.

63. Tsibulnikov S, Maslov L, Voronkov N, Oeltgen P. Thyroid hormones and the mechanisms of adaptation to cold. Hormones. 2020 Sep;19(3):329–39.

64. Pombo CM, Iglesias C, Sartages M, Zalvide JB. MST kinases and metabolism. Endocrinology. 2019;160(5):1111–8.

65. Cho YK, Son Y, Saha A, Kim D, Choi C, Kim M, et al. STK3/STK4 signalling in adipocytes regulates mitophagy and energy expenditure. Nature Metabolism. 2021;3(3):428–41.

66. Yuan Y, Fan Y, Zhou Y, Qiu R, Kang W, Liu Y, et al. Linker histone variant H1. 2 is a brake on white adipose tissue browning. Nature Communications. 2023;14(1):3982.

67. Fairbridge NA, Southall TM, Ayre DC, Komatsu Y, Raquet PI, Brown RJ, et al. Loss of CD24 in mice leads to metabolic dysfunctions and a reduction in white adipocyte tissue. PLoS One. 2015;10(11):e0141966.

68. Hsieh P, Dang V, Vollger MR, Mao Y, Huang TH, Dishuck PC, et al. Evidence for opposing selective forces operating on human-specific duplicated TCAF genes in Neanderthals and humans. Nature communications. 2021;12(1):5118.

69. Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 2014;512(7513):194–7.

70. Shenkin A. The key role of micronutrients☆. Clinical Nutrition. 2006 Feb;25(1):1–13.

71. Ross AC, Caballero B, Cousins RJ, Tucker KL. Modern nutrition in health and disease [Internet]. Jones & Bartlett Learning; 2020 [cited 2024 Sep 27]. Available from: https://books.google.com/books?hl=en&lr=&id=9zZvEAAAQBAJ&oi=fnd&pg=PP1&dq=Modern+nutrition+in+health+and+disease&ots=PpP6oNNd7c&sig=f-2e_mB7J2nbIdk0etoUABMpQOs

72. Ivask M, Volke V, Raasmaja A, Kõks S. High-fat diet associated sensitization to metabolic stress in Wfs1 heterozygous mice. Molecular Genetics and Metabolism. 2021;134(1–2):203–11.

73. Trinkaus E. Neandertal limb features and cold adaptation. In: Aspects of Human Evolution. London: Taylor and Francis TTD; 1981. p. 187–224.

74. Holliday TW. Postcranial evidence of cold adaptation in European Neandertals. American J Phys Anthropol. 1997 Oct;104(2):245–58.

75. Vernot B, Akey JM. Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. Science. 2014 Feb 28;343(6174):1017–21.

76. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 2014 Mar 20;507(7492):354–7.

77. Dobrovolskaya MV, Tiunov AV. The Neanderthals of Okladnikov Cave Altai: Environment and Diet Based on Isotopic Analysis*. Archaeology, Ethnology and Anthropology of Eurasia. 2013 Mar;41(1):78–88.

78. Dobrovolskaya MV, Tiunov AV. Stable isotope (13C/12C and 15N/14N) evidence for Late Pleistocene hominines' Palaeodiets in Gorny Altai. In: Proceedings of the International Symposium. Denisova Cave, Altai: Novosibirsk: Publishing Department of the Institute of Archaeology and Ethnography SB RAS; 2011.

79. Salazar-García DC, Power RC, Rudaya N, Kolobova K, Markin S, Krivoshapkin A, et al. Dietary evidence from Central Asian Neanderthals: A combined isotope and plant microremains approach at Chagyrskaya Cave (Altai, Russia). Journal of Human Evolution. 2021 Jul;156:102985.

80. Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Francioli LC, et al. A structural variation reference for medical and population genetics. Nature. 2020 May 28;581(7809):444–51.

81. Ben-Dor M, Gopher A, Barkai R. Neandertals' large lower thorax may represent adaptation to high protein diet. American J Phys Anthropol. 2016 Jul;160(3):367–78.

82. Snodgrass JJ, Leonard WR, Robertson ML. The Energetics of Encephalization in Early Hominids. In: Hublin JJ, Richards MP, editors. The Evolution of Hominin Diets [Internet]. Dordrecht: Springer Netherlands; 2009 [cited 2024 Sep 27]. p. 15–29. (Vertebrate Paleobiology and Paleoanthropology). Available from: http://link.springer.com/10.1007/978-1-4020-9699-0_2

83. Fellows Yates JA, Velsko IM, Aron F, Posth C, Hofman CA, Austin RM, et al. The evolution and changing ecology of the African hominid oral microbiome. Proc Natl Acad Sci USA. 2021 May 18;118(20):e2021655118.

84. Franklin MP, Sathyanarayan A, Mashek DG. Acyl-CoA Thioesterase 1 (ACOT1) Regulates PPARα to Couple Fatty Acid Flux With Oxidative Capacity During Fasting. Diabetes. 2017 Aug 1;66(8):2112–23.

85. Kaleta C, De Figueiredo LF, Schuster S. Against the stream: relevance of gluconeogenesis from fatty acids for natives of the arctic regions. International Journal of Circumpolar Health. 2012 Jan 31;71(1):18436.

86. Pereira PMDCC, Vicente AFDRB. Meat nutritional composition and nutritive role in the human diet. Meat Science. 2013 Mar;93(3):586–92.

87. Dhital S, Lin AHM, Hamaker BR, Gidley MJ, Muniandy A. Mammalian Mucosal α-Glucosidases Coordinate with α-Amylase in the Initial Starch Hydrolysis Stage to Have a Role in Starch Digestion beyond Glucogenesis. Blachier F, editor. PLoS ONE. 2013 Apr 25;8(4):e62546.

88. Khrameeva EE, Bozek K, He L, Yan Z, Jiang X, Wei Y, et al. Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. Nat Commun. 2014 Apr 1;5(1):3584.

89. Melamed Y, Kislev ME, Geffen E, Lev-Yadun S, Goren-Inbar N. The plant component of an Acheulian diet at Gesher Benot Ya'aqov, Israel. Proc Natl Acad Sci USA. 2016 Dec 20;113(51):14674–9.

90. Zilhão J, Angelucci DE, Igreja MA, Arnold LJ, Badal E, Callapez P, et al. Last Interglacial Iberian Neandertals as fisher-hunter-gatherers. Science. 2020 Mar 27;367(6485):eaaz7943.

91. Ding X, Kaminsky LS. Human Extrahepatic Cytochromes P450: Function in Xenobiotic Metabolism and Tissue-Selective Chemical Toxicity in the Respiratory and Gastrointestinal Tracts. Annu Rev Pharmacol Toxicol. 2003 Apr;43(1):149–73.

92. Condemi S, Savatier F. L'énigme Denisova. Paris: Albin Michel; 2024.

93. Skoglund P, Jakobsson M. Archaic human ancestry in East Asia. Proc Natl Acad Sci USA. 2011 Nov 8;108(45):18301–6.

94. Ungar PS, Sponheimer M. Hominin Diets. In: Begun DR, editor. A Companion to Paleoanthropology [Internet]. 1st ed. Wiley; 2013 [cited 2024 Sep 27]. p. 165–82. Available from: https://onlinelibrary.wiley.com/doi/10.1002/9781118332344.ch9

95. Larbey C, Mentzer SM, Ligouis B, Wurz S, Jones MK. Cooked starchy food in hearths ca. 120 kya and 65 kya (MIS 5e and MIS 4) from Klasies River Cave, South Africa. Journal of Human Evolution. 2019 Jun;131:210–27.

96. Zhu H, Ren Q, Yan Z, Xu S, Luo J, Wu X, et al. Human HAND1 inhibits the conversion of cholesterol to steroids in trophoblasts. Journal of Genetics and Genomics. 2022 Apr;49(4):350–63.

97. Giannuzzi G, Schmidt PJ, Porcu E, Willemin G, Munson KM, Nuttle X, et al. The Human-Specific BOLA2 Duplication Modifies Iron Homeostasis and Anemia Predisposition in Chromosome 16p11.2 Autism Individuals. The American Journal of Human Genetics. 2019 Nov;105(5):947–58.

98. Reiner AP, Carlson CS, Rieder MJ, Siscovick DS, Liu K, Chandler WL, et al. Coagulation factor VII gene haplotypes, obesity-related traits, and cardiovascular risk in young women. Journal of Thrombosis and Haemostasis. 2007;5(1):42–9.

99. Bernardi F, Mariani G. Biochemical, molecular and clinical aspects of coagulation factor VII and its role in hemostasis and thrombosis. haematol. 2021 Jan 7;106(2):351–62.

100. Fairley S, Lowy-Gallego E, Perry E, Flicek P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. Nucleic Acids Research. 2020 Jan 8;48(D1):D941–7.

101. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature. 2010 Oct 28;467(7319):1061–73.

102. Xie C, Tammi MT. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics. 2009 Dec;10(1):80.

103. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 2014 Feb;15(2):121–32.

104. Condemi S, Weniger GC, editors. Continuity and Discontinuity in the Peopling of Europe: One Hundred Fifty Years of Neanderthal Study [Internet]. Dordrecht: Springer Netherlands; 2011 [cited 2024 Oct 2]. (Vertebrate Paleobiology and Paleoanthropology). Available from: http://link.springer.com/10.1007/978-94-007-0492-3

105. Hublin JJ, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. Nature. 2017 Jun 8;546(7657):289–92.

106. Mendez FL, Poznik GD, Castellano S, Bustamante CD. The Divergence of Neandertal and Modern Human Y Chromosomes. The American Journal of Human Genetics. 2016 Apr;98(4):728–34.

107. Posth C, Wißing C, Kitagawa K, Pagani L, Van Holstein L, Racimo F, et al. Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. Nat Commun. 2017 Jul 4;8(1):16046.

108. Mounier A, Marchal F, Condemi S. Is Homo heidelbergensis a distinct species? New insight on the Mauer mandible. Journal of Human Evolution. 2009 Mar;56(3):219–46.

109. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A Draft Sequence of the Neandertal Genome. Science. 2010 May 7;328(5979):710–22.

110. Aiello LC, Wells JCK. Energetics and the Evolution of the Genus *HOMO*. Annu Rev Anthropol. 2002 Oct;31(1):323–38.

111. Peyrégne S, Slon V, Mafessoni F, De Filippo C, Hajdinjak M, Nagel S, et al. Nuclear DNA from two early Neandertals reveals 80,000 years of genetic continuity in Europe. Sci Adv. 2019 Jun;5(6):eaaw5873.

112. Bocquet-Appel JP, Degioanni A. Neanderthal Demographic Estimates. Current Anthropology. 2013 Dec;54(S8):S202–13.

113. Kolobova KA, Roberts RG, Chabai VP, Jacobs Z, Krajcarz MT, Shalagina AV, et al. Archaeological evidence for two separate dispersals of Neanderthals into southern Siberia. Proc Natl Acad Sci USA. 2020 Feb 11;117(6):2879–85.

114. Bokelmann L, Hajdinjak M, Peyrégne S, Brace S, Essel E, De Filippo C, et al. A genetic analysis of the Gibraltar Neanderthals. Proc Natl Acad Sci USA. 2019 Jul 30;116(31):15610–5.

115. Conard NJ, Richter J, editors. Neanderthal Lifeways, Subsistence and Technology [Internet]. Dordrecht: Springer Netherlands; 2011 [cited 2024 Sep 27]. (Vertebrate Paleobiology and Paleoanthropology; vol. 19). Available from: http://link.springer.com/10.1007/978-94-007-0415-2

116. Pearce E, Stringer C, Dunbar RIM. New insights into differences in brain organization between Neanderthals and anatomically modern humans. Proc R Soc B. 2013 May 7;280(1758):20130168.

117. Gunz P, Neubauer S, Maureille B, Hublin JJ. Brain development after birth differs between Neanderthals and modern humans. Current Biology. 2010 Nov;20(21):R921–2.

118. Kyriacou A, Bruner E. Innovation and the Evolution of Human Behavior. Brain Evolution, Innovation, and Endocranial Variations in Fossil Hominids. PA. 2011:130–43.

119. Gilpin W, Feldman MW, Aoki K. An ecocultural model predicts Neanderthal extinction through competition with modern humans. Proc Natl Acad Sci USA. 2016 Feb 23;113(8):2134–9.

120. Kochiyama T, Ogihara N, Tanabe HC, Kondo O, Amano H, Hasegawa K, et al. Reconstructing the Neanderthal brain using computational anatomy. Sci Rep. 2018 Apr 26;8(1):6296.

121. Aitken MJ, Aitken MJ, editors. The origin of modern humans and the impact of chronometric dating: a discussion. Princeton, NJ: Princeton Univ. Press; 1993. 248 p.

122. Barton CM, Riel-Salvatore J, Anderies JM, Popescu G. Modeling Human Ecodynamics and Biocultural Interactions in the Late Pleistocene of Western Eurasia. Hum Ecol. 2011 Dec;39(6):705–25.

123. Zilhão J, Angelucci DE, Badal-García E, d'Errico F, Daniel F, Dayet L, et al. Symbolic use of marine shells and mineral pigments by Iberian Neandertals. Proc Natl Acad Sci USA. 2010 Jan 19;107(3):1023–8.

124. Hoffmann DL, Standish CD, Pike AWG, García-Diez M, Pettitt PB, Angelucci DE, et al. Dates for Neanderthal art and symbolic behaviour are reliable. Nat Ecol Evol. 2018 Jun 25;2(7):1044–5.

125. Zollikofer CPE, Ponce De León MS, Vandermeersch B, Lévêque F. Evidence for interpersonal violence in the St. Césaire Neanderthal. Proc Natl Acad Sci USA. 2002 Apr 30;99(9):6444–8.

126. Houldcroft CJ, Underdown SJ. Neanderthal genomics suggests a pleistocene time frame for the first epidemiologic transition. American J Phys Anthropol. 2016 Jul;160(3):379–88.

127. Sørensen B. Demography and the extinction of European Neanderthals. Journal of Anthropological Archaeology. 2011 Mar;30(1):17–29.

128. Frost GS, Walton GE, Swann JR, Psichas A, Costabile A, Johnson LP, et al. Impacts of Plant-Based Foods in Ancestral Hominin Diets on the Metabolism and Function of Gut Microbiota *In Vitro*. Huffnagle GB, editor. mBio. 2014 Jul;5(3):e00853-14.

129. Harris DN, Ruczinski I, Yanek LR, Becker LC, Becker DM, Guio H, et al. Evolution of Hominin Polyunsaturated Fatty Acid Metabolism: From Africa to the New World. Das S, editor. Genome Biology and Evolution. 2019 May 1;11(5):1417–30.

130. Leonard WR, Robertson ML, Snodgrass JJ. What did humans evolve to eat? Metabolic implications of major trends in hominid evolution. In: Human Diet and Nutrition in Biocultural Perspective: Past Meets Present. Berghahn Books; 2010. p. 14–34.

131. Pontzer H. Ecological Energetics in Early *Homo*. Current Anthropology. 2012 Dec;53(S6):S346–58.

132. Aiello LC. Brains and guts in human evolution: The Expensive Tissue Hypothesis. Braz J Genet. 1997 Mar;20(1):141–8.

133. Zink KD, Lieberman DE. Impact of meat and Lower Palaeolithic food processing techniques on chewing in humans. Nature. 2016 Mar 24;531(7595):500–3.

134. Slon V, Mafessoni F, Vernot B, De Filippo C, Grote S, Viola B, et al. The genome of the offspring of a Neanderthal mother and a Denisovan father. Nature. 2018 Sep;561(7721):113–6.

135. Krause J, Fu Q, Good JM, Viola B, Shunkov MV, Derevianko AP, et al. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature. 2010 Apr 8;464(7290):894–7.

136. Sawyer S, Renaud G, Viola B, Hublin JJ, Gansauge MT, Shunkov MV, et al. Nuclear and mitochondrial DNA sequences from two Denisovan individuals. Proc Natl Acad Sci USA. 2015 Dec 22;112(51):15696–700.

137. Bennett EA, Crevecoeur I, Viola B, Derevianko AP, Shunkov MV, Grange T, et al. Morphology of the Denisovan phalanx closer to modern humans than to Neanderthals. Sci Adv. 2019 Sep 6;5(9):eaaw3950.

138. Chen F, Welker F, Shen CC, Bailey SE, Bergmann I, Davis S, et al. A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. Nature. 2019 May;569(7756):409–12.

139. Zhang D, Xia H, Chen F, Li B, Slon V, Cheng T, et al. Denisovan DNA in Late Pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau. Science. 2020 Oct 30;370(6516):584–7.

140. Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, et al. Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. The American Journal of Human Genetics. 2011 Oct;89(4):516–28.

141. Mondal M, Casals F, Xu T, Dall'Olio GM, Pybus M, Netea MG, et al. Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. Nat Genet. 2016 Sep;48(9):1066–70.

142. Sankararaman S, Mallick S, Patterson N, Reich D. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. Current Biology. 2016 May;26(9):1241–7.

143. Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science. 2016 Apr 8;352(6282):235–9.

144. Larena M, McKenna J, Sanchez-Quinto F, Bernhardsson C, Ebeo C, Reyes R, et al. Philippine Ayta possess the highest level of Denisovan ancestry in the world. Current Biology. 2021 Oct;31(19):4219-4230.e10.

145. Larena M, Sanchez-Quinto F, Sjödin P, McKenna J, Ebeo C, Reyes R, et al. Multiple migrations to the Philippines during the last 50,000 years. Proc Natl Acad Sci USA. 2021 Mar 30;118(13):e2026132118.

146. Zavala EI, Jacobs Z, Vernot B, Shunkov MV, Kozlikin MB, Derevianko AP, et al. Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. Nature. 2021 Jul 15;595(7867):399–403.

147. Bergström A, Stringer C, Hajdinjak M, Scerri EML, Skoglund P. Origins of modern human ancestry. Nature. 2021 Feb 11;590(7845):229–37.

148. Bosch MD, Mannino MA, Prendergast AL, O'Connell TC, Demarchi B, Taylor SM, et al. New chronology for Ksâr 'Akil (Lebanon) supports Levantine route of modern human dispersal into Europe. Proc Natl Acad Sci USA. 2015 Jun 23;112(25):7683–8.

149. Zilhão J, d'ERRICO F, Banks WE, Teyssandier N. A Data-Driven Paradigm Shift for the Middle-to-Upper Palaeolithic Transition and the Neandertal Debate. Quaternary Environments and Humans. 2024 Nov;100037.

150. Anghelinu M, Niţă L, Sitlivy V, Uthmeier T, Băltean I. Looking around Peştera Cu Oase: The beginnings of Upper Paleolithic in Romania. Quaternary International. 2012 Oct;274:136–57.

151. Hajdinjak M, Mafessoni F, Skov L, Vernot B, Hübner A, Fu Q, et al. Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry. Nature. 2021 Apr 8;592(7853):253–7.

152. Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas AS, Manica A, Moltke I, et al. Genomic structure in Europeans dating back at least 36,200 years. Science. 2014 Nov 28;346(6213):1113–8.

153. Dinnis R, Bessudnov A, Reynolds N, Devièse T, Pate A, Sablin M, et al. New data for the Early Upper Paleolithic of Kostenki (Russia). Journal of Human Evolution. 2019 Feb;127:21–40.

154. Prüfer K, Posth C, Yu H, Stoessel A, Spyrou MA, Deviese T, et al. A genome sequence from a modern human skull over 45,000 years old from Zlatý kůň in Czechia. Nat Ecol Evol. 2021 Apr 7;5(6):820–5.

155. Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, et al. The genetic history of Ice Age Europe. Nature. 2016 Jun 9;534(7606):200–5.

156. Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, Mallick S, et al. The genomic history of southeastern Europe. Nature. 2018 Mar;555(7695):197–203.

157. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015 Jun 11;522(7555):207–11.

158. Sánchez-Quinto F, Lalueza-Fox C. Almost 20 years of Neanderthal palaeogenetics: adaptation, admixture, diversity, demography and extinction. Phil Trans R Soc B. 2015 Jan 19;370(1660):20130374.

159. Jacobs Z, Li B, Shunkov MV, Kozlikin MB, Bolikhovskaya NS, Agadjanian AK, et al. Timing of archaic hominin occupation of Denisova Cave in southern Siberia. Nature. 2019 Jan;565(7741):594–9.

160. Douka K, Slon V, Jacobs Z, Ramsey CB, Shunkov MV, Derevianko AP, et al. Age estimates for hominin fossils and the onset of the Upper Palaeolithic at Denisova Cave. Nature. 2019 Jan;565(7741):640–4.

161. Brown S, Higham T, Slon V, Pääbo S, Meyer M, Douka K, et al. Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. Sci Rep. 2016 Mar 29;6(1):23559.

162. Viola B, Markin SV, Buzhilova AP, Mednikova MB, Cabec AL, Shunkov MV. New Neanderthal remains from Chagyrskaya Cave (Altai mountains, Russian Federation). Am J Phys Anthropol. 2012;147(54):293–4.

163. Derevianko AP, Markin SV, Shunkov MV. The Sibiryachikha Facies of the Middle Paleolithic of the Altai. Archaeology, Ethnology and Anthropology of Eurasia. 2013 Mar;41(1):89–103.

164. Vasiliev SK. Large Mammal Fauna from the Pleistocene Deposits of Chagyrskaya Cave Northwestern Altai (based on 2007–2011 Excavations). Archaeology, Ethnology and Anthropology of Eurasia. 2013 Mar;41(1):28–44.

165. Rudaya N, Vasiliev S, Viola B, Talamo S, Markin S. Palaeoenvironments during the period of the Neanderthals settlement in Chagyrskaya cave (Altai Mountains, Russia). Palaeogeography, Palaeoclimatology, Palaeoecology. 2017 Feb;467:265–76.

166. Glantz M, Van Arsdale A, Temirbekov S, Beeton T. How to survive the glacial apocalypse: Hominin mobility strategies in late Pleistocene Central Asia. Quaternary International. 2018 Feb;466:82–92.

167. Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S. Neandertal DNA Sequences and the Origin of Modern Humans. Cell. 1997 Jul;90(1):19–30.

168. Ovchinnikov IV, Götherström A, Romanova GP, Kharitonov VM, Lidén K, Goodwin W. Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature. 2000 Mar;404(6777):490–3.

169. Golovanova LV, Doronichev VB, Cleghorn NE, Koulkova MA, Sapelko TV, Shackley MS. Significance of Ecological Factors in the Middle to Upper Paleolithic Transition. Current Anthropology. 2010 Oct;51(5):655–91.

170. Devièse T, Karavanić I, Comeskey D, Kubiak C, Korlević P, Hajdinjak M, et al. Direct dating of Neanderthal remains from the site of Vindija Cave and implications for the Middle to Upper Paleolithic transition. Proc Natl Acad Sci USA. 2017 Oct 3;114(40):10606–11.

171. Karriger WM, Schmidt CW, Smith FH. Dental Microwear Texture Analysis of Croatian Neandertal Molars. PaleoAnthropology. 2016;172−184.

172. Richards MP, Pettitt PB, Trinkaus E, Smith FH, Paunović M, Karavanić I. Neanderthal diet at Vindija and Neanderthal predation: The evidence from stable isotopes. Proc Natl Acad Sci USA. 2000 Jun 20;97(13):7663–6.

173. Semal P, Rougier H, Crevecoeur I, Jungels C, Flas D, Hauzeur A, et al. New data on the late Neandertals: Direct dating of the Belgian Spy fossils. American J Phys Anthropol. 2009 Apr;138(4):421–8.

174. Rougier H, Semal P. Spy Cave: 125 Years of Multidisciplinary Research at the Betche Aux Rotches (Jemeppe-sur-Sambre, Province of Namur, Belgium). Société Royale Belge d'Anthropologie et de Préhistoire; 2013. 380 p. (Anthropologica et Praehistorica; vol. 123).

175. Williams FL, Schmidt CW, Droke JL, Willman JC, Semal P, Becam G, et al. Dietary reconstruction of Spy I using dental microwear texture analysis. Comptes Rendus Palevol. 2019 Dec;18(8):1083–94.

176. El Zaatari S, Grine FE, Ungar PS, Hublin JJ. Ecogeographic variation in Neandertal dietary habits: Evidence from occlusal molar microwear texture analysis. Journal of Human Evolution. 2011 Oct;61(4):411–24.

177. Hardy K, Buckley S, Collins MJ, Estalrrich A, Brothwell D, Copeland L, et al. Neanderthal medics? Evidence for food, cooking, and medicinal plants entrapped in dental calculus. Naturwissenschaften. 2012 Aug;99(8):617–26.

178. Radini A, Nikita E, Buckley S, Copeland L, Hardy K. Beyond food: The multiple pathways for inclusion of materials into ancient dental calculus. American J Phys Anthropol. 2017 Jan;162(S63):71–83.

179. Hardy K. Plant use in the Lower and Middle Palaeolithic: Food, medicine and raw materials. Quaternary Science Reviews. 2018 Jul;191:393–405.

180. Rougier H, Crevecoeur I, Beauval C, Posth C, Flas D, Wißing C, et al. Neandertal cannibalism and Neandertal bones used as tools in Northern Europe. Sci Rep. 2016 Jul 6;6(1):29005.

181. Wißing C, Rougier H, Crevecoeur I, Germonpré M, Naito YI, Semal P, et al. Isotopic evidence for dietary ecology of late Neandertals in North-Western Europe. Quaternary International. 2016 Aug;411:327–45.

182. Hervella M, Svensson EM, Alberdi A, Günther T, Izagirre N, Munters AR, et al. The mitogenome of a 35,000-year-old Homo sapiens from Europe supports a Palaeolithic back-migration to Africa. Sci Rep. 2016 May 19;6(1):25501.

183. Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, et al. An early modern human from Romania with a recent Neanderthal ancestor. Nature. 2015 Aug;524(7564):216–9.

184. Moorjani P, Sankararaman S, Fu Q, Przeworski M, Patterson N, Reich D. A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. Proc Natl Acad Sci USA. 2016 May 17;113(20):5652–7.

185. Wong EHM, Khrunin A, Nichols L, Pushkarev D, Khokhrin D, Verbenko D, et al. Reconstructing genetic history of Siberian and Northeastern European populations. Genome Res. 2017 Jan;27(1):1–14.

186. Abramova ZA. L'art paléolithique d'Europe orientale et de Sibérie. Grenoble: J. Millon; 1995. (L'homme des origines).

187. Sikora M, Pitulko VV, Sousa VC, Allentoft ME, Vinner L, Rasmussen S, et al. The population history of northeastern Siberia since the Pleistocene. Nature. 2019 Jun;570(7760):182–8.

188. Flegontov P, Changmai P, Zidkova A, Logacheva MD, Altınışık NE, Flegontova O, et al. Genomic study of the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. Sci Rep. 2016 Feb 11;6(1):20768.

189. Heuertz M. Documents préhistoriques du territoire luxembourgeois. Le milieu naturel. L'homme et son œuvre. Luxembourg, 1, 295 (1969). Vol. 1. Musée d'Histoire Naturelle et de La Société des Naturalistes luxembourgeois; 1969. 295 p.

190. Stäuble H. Häuser und absolute Datierung den ältesten Bandkeramik. Vol. 117. Bonn, Habelt: Universitätsforsch. Prähist. Arch.; 2005.

191. Price TD, Wahl J, Knipper C, Burger-Heinrich E, Kurtz G, Bentley RA. Das bandkeramische Gräberfeld vom „Viesenhäuser Hof" bei Stuttgart-Mühlhausen: Neue Untersuchungsergebnisse zum Migrationsverhalten im frühen Neolithikum. In: Fundberichte aus Baden-Württemberg. 2003. p. 23–58.

192. Schubert M, Ginolhac A, Lindgreen S, Thompson JF, AL-Rasheid KA, Willerslev E, et al. Improving ancient DNA read mapping against modern reference genomes. BMC Genomics. 2012 Dec;13(1):178.

193. Oliva A, Tobler R, Llamas B, Souilmi Y. Additional evaluations show that specific *BWA-aln* settings still outperform *BWA-mem* for ancient DNA data alignment. Ecology and Evolution. 2021 Dec;11(24):18743–8.

194. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.

195. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021 Jan 29;10(2):giab008.

196. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

197. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841–2.

198. Chen YC, Liu T, Yu CH, Chiang TY, Hwang CC. Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. Xu Y, editor. PLoS ONE. 2013 Apr 29;8(4):e62856.

199. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016 Jul 8;44(W1):W160–5.

200. Benjamini Y, Speed TP. Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Research. 2012 May 1;40(10):e72–e72.

201. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep;489(7414):57–74.

202. Amemiya HM, Kundaje A, Boyle AP. The ENCODE blacklist: identification of problematic regions of the genome. Scientific reports. 2019;9(1):9354.

203. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. Hancock J, editor. Bioinformatics. 2018 Mar 1;34(5):867–8.

204. Frankish A, Carbonell-Sala S, Diekhans M, Jungreis I, Loveland JE, Mudge JM, et al. GENCODE: reference annotation for the human and mouse genomes in 2023. Nucleic Acids Research. 2023 Jan 6;51(D1):D942–9.

205. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009 Aug;4(8):1184–91.

206. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Research. 2018 Jan 4;46(D1):D754–61.

207. Digre A, Lindskog C. The human protein atlas—Integrated omics for single cell mapping of the human proteome. Protein Science. 2023 Feb;32(2):e4562.

208. Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. Protein Science. 2018 Jan;27(1):233–44.

209. Pontén F, Jirström K, Uhlen M. The Human Protein Atlas—a tool for pathology. The Journal of Pathology. 2008 Dec;216(4):387–93.

210. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou L, Mi H. PANTHER : Making genome-scale phylogenetics accessible to all. Protein Science. 2022 Jan;31(1):8–22.

211. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Research. 2022 Jul 5;50(W1):W216–21.

212. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics. 2009 Nov 15;25(22):3045–6.

213. Hunt MC, Rautanen A, Westin MAK, Svensson LT, Alexson SEH. Analysis of the mouse and human acyl-CoA thioesterase (ACOT) gene clusters shows that convergent, functional evolution results in a reduced number of human peroxisomal ACOTs [1]. FASEB j. 2006 Sep;20(11):1855–64.

214. Brocker C, Carpenter C, Nebert DW, Vasiliou V. Evolutionary divergence and functions of the human acyl-CoA thioesterase gene (ACOT) family. Hum Genomics. 2010;4(6):411.

215. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria; 2013. (R Foundation for Statistical Computing). Available from: http://www.R-project.org/

216. Kassambara A, Mundt F. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses [Internet]. 2020. (Comprehensive R Archive Network (CRAN)). Available from: https://CRAN.R-project.org/package=factoextra

217. Sievert C. Interactive Web-Based Data Visualization with R, plotly, and shiny [Internet]. 1st ed. Chapman and Hall/CRC; 2020 [cited 2024 Sep 28]. Available from: https://www.taylorfrancis.com/books/9780429824210

218. Fabre V, Condemi S, Degioanni A. Genetic Evidence of Geographical Groups among Neanderthals. Awadalla P, editor. PLoS ONE. 2009 Apr 15;4(4):e5151.

219. Dalen L, Orlando L, Shapiro B, Brandstrom-Durling M, Quam R, Gilbert MTP, et al. Partial Genetic Turnover in Neandertals: Continuity in the East and Population Replacement in the West. Molecular Biology and Evolution. 2012 Aug 1;29(8):1893–7.

220. Lipson M, Ribot I, Mallick S, Rohland N, Olalde I, Adamski N, et al. Ancient West African foragers in the context of African population history. Nature. 2020 Jan 30;577(7792):665–70.

221. Gamage N, Barnett A, Hempel N, Duggleby RG, Windmill KF, Martin JL, et al. Human Sulfotransferases and Their Role in Chemical Metabolism. Toxicological Sciences. 2006 Mar 1;90(1):5–22.

222. Dooley TP. Cloning of the human phenol sulfotransferase gene family: three genes implicated in the metabolism of catecholamines, thyroid hormones and drugs. Chemico-Biological Interactions. 1998 Feb;109(1–3):29–41.

223. Hildebrandt MAT, Salavaggione OE, Martin YN, Flynn HC, Jalal S, Wieben ED, et al. Human SULT1A3 pharmacogenetics: gene duplication and functional genomic studies. Biochemical and Biophysical Research Communications. 2004 Sep;321(4):870–8.

224. Coughtrie MWH. Function and organization of the human cytosolic sulfotransferase (SULT) family. Chemico-Biological Interactions. 2016 Nov;259:2–7.

225. Dajani R, Cleasby A, Neu M, Wonacott AJ, Jhoti H, Hood AM, et al. X-ray Crystal Structure of Human Dopamine Sulfotransferase, SULT1A3. Journal of Biological Chemistry. 1999 Dec;274(53):37862–8.

226. Alex KD, Pehek EA. Pharmacologic mechanisms of serotonergic regulation of dopamine neurotransmission. Pharmacology & Therapeutics. 2007 Feb;113(2):296–320.

227. Bairam AF, Rasool MI, Alherz FA, Abunnaja MS, El Daibani AA, Gohal SA, et al. Sulfation of catecholamines and serotonin by SULT1A3 allozymes. Biochemical Pharmacology. 2018 May;151:104–13.

228. Svendsen JM, Smogorzewska A, Sowa ME, O'Connell BC, Gygi SP, Elledge SJ, et al. Mammalian BTBD12/SLX4 Assembles A Holliday Junction Resolvase and Is Required for DNA Repair. Cell. 2009 Jul;138(1):63–77.

229. Hanhineva K, Törrönen R, Bondia-Pons I, Pekkinen J, Kolehmainen M, Mykkänen H, et al. Impact of Dietary Polyphenols on Carbohydrate Metabolism. IJMS. 2010 Mar 31;11(4):1365–402.

230. Waheed Janabi AH, Kamboh AA, Saeed M, BiBi J, Naveed M, Huixia L. Flavonoid-rich foods (FRF): A promising nutraceutical approach against lifespan-shortening diseases. Iranian Journal of Basic Medical Sciences [Internet]. 2019 Dec [cited 2024 Oct 2];(Online First). Available from: https://doi.org/10.22038/ijbms.2019.35125.8353

231. Sims HF, Lowe ME. The human colipase gene: isolation, chromosomal location, and tissue-specific expression. Biochemistry. 1992 Aug 11;31(31):7120–5.

232. Miled N. Digestive lipases: From three-dimensional structure to physiology. Biochimie. 2000 Nov;82(11):973–86.

233. Natarajan P, Peloso GM, Zekavat SM, Montasser M, Ganna A, Chaffin M, et al. Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. Nat Commun. 2018 Aug 23;9(1):3391.

234. Davies W, White D, Lewis M, Stringer C. Evaluating the transitional mosaic: frameworks of change from Neanderthals to Homo sapiens in eastern Europe. Quaternary Science Reviews. 2015 Jun;118:211–42.

235. Morin E. Rethinking the emergence of bone grease procurement. Journal of Anthropological Archaeology. 2020 Sep;59:101178.

236. Calo J, Blanco AM, Soengas JL. Dietary lipid sensing through fatty acid oxidation and chylomicron formation in the gastrointestinal tract of rainbow trout. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology. 2024 Aug;294:111638.

237. Pivovarova O, Jürchott K, Rudovich N, Hornemann S, Ye L, Möckel S, et al. Changes of Dietary Fat and Carbohydrate Content Alter Central and Peripheral Clock in Humans. The Journal of Clinical Endocrinology & Metabolism. 2015 Jun;100(6):2291–302.

238. Gao T, McKenna B, Li C, Reichert M, Nguyen J, Singh T, et al. Pdx1 Maintains β Cell Identity and Function by Repressing an α Cell Program. Cell Metabolism. 2014 Feb;19(2):259–71.

239. Chen C, Leavitt T, Sibley E. Intestinal Pdx1 mediates nutrient metabolism gene networks and maternal expression is essential for perinatal growth in mice. Biochemical and Biophysical Research Communications. 2012 Aug;424(3):549–53.

240. Zhang M, Yang C, Zhu M, Qian L, Luo Y, Cheng H, et al. Saturated fatty acids entrap PDX1 in stress granules and impede islet beta cell function. Diabetologia. 2021 May;64(5):1144–57.

241. Liu SS, Wu F, Jin YM, Chang WQ, Xu TM. HDAC11: a rising star in epigenetics. Biomedicine & Pharmacotherapy. 2020 Nov;131:110607.

242. Sun L, Marin De Evsikova C, Bian K, Achille A, Telles E, Seto E. HDAC11 Deficiency Prevents High-Fat Diet-Induced Obesity and Metabolic Syndrome [Internet]. 2018 [cited 2024 Oct 2]. Available from: http://biorxiv.org/lookup/doi/10.1101/310383

243. Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, Chisholm RL, et al. The phenotypic legacy of admixture between modern humans and Neandertals. Science. 2016 Feb 12;351(6274):737–41.

244. Geng C, Zhang Y, Gao Y, Tao W, Zhang H, Liu X, et al. Mst1 regulates hepatic lipid metabolism by inhibiting Sirt1 ubiquitination in mice. Biochemical and Biophysical Research Communications. 2016 Mar;471(4):444–9.

245. Bigham AW, Wilson MJ, Julian CG, Kiyamu M, Vargas E, Leon-Velarde F, et al. Andean and Tibetan patterns of adaptation to high altitude. American J Hum Biol. 2013 Mar;25(2):190–7.

246. Brutsaert TD, Kiyamu M, Elias Revollendo G, Isherwood JL, Lee FS, Rivera-Ch M, et al. Association of *EGLN1* gene with high aerobic capacity of Peruvian Quechua at high altitude. Proc Natl Acad Sci USA. 2019 Nov 26;116(48):24006–11.

247. Duan Y, Zeng L, Zheng C, Song B, Li F, Kong X, et al. Inflammatory Links Between High Fat Diets and Diseases. Front Immunol. 2018 Nov 13;9:2649.

248. Han S, Lone MA, Schneiter R, Chang A. Orm1 and Orm2 are conserved endoplasmic reticulum membrane proteins regulating lipid homeostasis and protein quality control. Proc Natl Acad Sci USA. 2010 Mar 30;107(13):5851–6.

249. Trivedi NN, Tamraz B, Chu C, Kwok PY, Caughey GH. Human subjects are protected from mast cell tryptase deficiency despite frequent inheritance of loss-of-function mutations. Journal of Allergy and Clinical Immunology. 2009 Nov;124(5):1099-1105.e4.

250. Lyons JJ, Yu X, Hughes JD, Le QT, Jamil A, Bai Y, et al. Elevated basal serum tryptase identifies a multisystem disorder associated with increased TPSAB1 copy number. Nat Genet. 2016 Dec;48(12):1564–9.

251. Li F, Wen H, Bukhari I, Liu B, Guo C, Ren F, et al. Relationship Between CNVs and Immune Cells Infiltration in Gastric Tumor Microenvironment. Front Genet. 2022 Jun 8;13:869967.

252. Ryaboshapkina M, Hammar M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. Sci Rep. 2017 Sep 27;7(1):12361.

253. Wang Y, Menon AK, Maki Y, Liu YS, Iwasaki Y, Fujita M, et al. Genome-wide CRISPR screen reveals CLPTM1L as a lipid scramblase required for efficient glycosylphosphatidylinositol biosynthesis. Proc Natl Acad Sci USA. 2022 Apr 5;119(14):e2115083119.

254. Frey-Jakobs S, Hartberger JM, Fliegauf M, Bossen C, Wehmeyer ML, Neubauer JC, et al. ZNF341 controls STAT3 expression and thereby immunocompetence. Sci Immunol. 2018 Jun 8;3(24):eaat4941.

255. Law SK, Dodds AW, Porter RR. A comparison of the properties of two classes, C4A and C4B, of the human complement component C4. The EMBO Journal. 1984 Aug;3(8):1819–23.

256. Gatenby PA, Barbosa JE, Lachmann PJ. Differences between C4A and C4B in the handling of immune complexes: the enhancement of CR1 binding is more important than the inhibition of immunoprecipitation. Clinical and Experimental Immunology. 2008 Jun 28;79(2):158–63.

257. Wang C, Ling T, Zhong N, Xu LG. N4BP3 Regulates RIG-I-Like Receptor Antiviral Signaling Positively by Targeting Mitochondrial Antiviral Signaling Protein. Front Microbiol. 2021 Nov 22;12:770600.

258. Okumura R, Kurakawa T, Nakano T, Kayama H, Kinoshita M, Motooka D, et al. Lypd8 promotes the segregation of flagellated microbiota and colonic epithelia. Nature. 2016 Apr;532(7597):117–21.

259. Hsu CC, Okumura R, Takeda K. Human LYPD8 protein inhibits motility of flagellated bacteria. Inflamm Regener. 2017 Dec;37(1):23.

260. Pombinho R, Pinheiro J, Resende M, Meireles D, Jalkanen S, Sousa S, et al. Stabilin-1 plays a protective role against Listeria monocytogenes infection through the regulation of cytokine and chemokine production and immune cell recruitment. Virulence. 2021 Dec 31;12(1):2088–103.

261. Manta CP, Leibing T, Friedrich M, Nolte H, Adrian M, Schledzewski K, et al. Targeting of Scavenger Receptors Stabilin-1 and Stabilin-2 Ameliorates Atherosclerosis by a Plasma Proteome Switch Mediating Monocyte/Macrophage Suppression. Circulation. 2022 Dec 6;146(23):1783–99.

262. Goel R, Boylan B, Gruman L, Newman PJ, North PE, Newman DK. The proinflammatory phenotype of PECAM-1-deficient mice results in atherogenic diet-induced steatohepatitis. American Journal of Physiology-Gastrointestinal and Liver Physiology. 2007 Dec;293(6):G1205–14.

263. Xue Y, Yin Y, Trabi EB, Xie F, Lin L, Mao S. Transcriptome analysis reveals the effect of high-grain pelleted and non-pelleted diets on ruminal epithelium of Hu-lamb. Animal. 2021 Jul;15(7):100278.

264. Plubell DL, Fenton AM, Wilmarth PA, Bergstrom P, Zhao Y, Minnier J, et al. GM-CSF driven myeloid cells in adipose tissue link weight gain and insulin resistance via formation of 2-aminoadipate. Sci Rep. 2018 Jul 31;8(1):11485.

265. Manivasagam S, Klein RS. Type III Interferons: Emerging Roles in Autoimmunity. Front Immunol. 2021 Nov 26;12:764062.

266. Reimúndez A, Fernández-Peña C, García G, Fernández R, Ordás P, Gallego R, et al. Deletion of the Cold Thermoreceptor TRPM8 Increases Heat Loss and Food Intake Leading to Reduced Body Temperature and Obesity in Mice. J Neurosci. 2018 Apr 11;38(15):3643–56.

267. Kostetsky EYa, Velansky PV. Phospholipids of sea worms, mollusks, and arthropods. Russ J Mar Biol. 2009 May;35(3):187–99.

268. Pirini M, Manuzzi MP, Pagliarani A, Trombetti F, Borgatti AR, Ventrella V. Changes in fatty acid composition of Mytilus galloprovincialis (Lmk) fed on microalgal and wheat germ diets. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology. 2007 Aug;147(4):616–26.

269. Řezanka T, Křesinová Z, Kolouchová I, Sigler K. Lipidomic analysis of bacterial plasmalogens. Folia Microbiol. 2012 Sep;57(5):463–72.

270. Pham TH, Manful CF, Pumphrey RP, Hamilton MC, Adigun OA, Vidal NP, et al. Big game cervid meat as a potential good source of plasmalogens for functional foods. Journal of Food Composition and Analysis. 2021 Mar;96:103724.

271. Singh V, Mani I, Chaudhary DK. ATP4A gene regulatory network for fine-tuning of proton pump and ion channels. Syst Synth Biol. 2013 Jun;7(1–2):23–32.

272. Sargsyan A, Doridot L, Hannou SA, Tong W, Srinivasan H, Ivison R, et al. HGFAC is a ChREBP-regulated hepatokine that enhances glucose and lipid homeostasis. JCI Insight. 2023 Jan 10;8(1):e153740.

273. Xu H, Ghishan FK, Kiela PR. SLC9 Gene Family: Function, Expression, and Regulation. In: Terjung R, editor. Comprehensive Physiology [Internet]. 1st ed. Wiley; 2018 [cited 2024 Oct 2]. p. 555–83. Available from: https://onlinelibrary.wiley.com/doi/10.1002/cphy.c170027

274. Crujeiras AB, Diaz-Lagares A, Sandoval J, Milagro FI, Navas-Carretero S, Carreira MC, et al. DNA methylation map in circulating leukocytes mirrors subcutaneous adipose tissue methylation pattern: a genome-wide analysis from non-obese and obese patients. Sci Rep. 2017 Feb 17;7(1):41903.

275. Timirci-Kahraman O, Yilmaz U, Yilmaz N, Cevik A, Horozoglu C, Celik F, et al. A Study of Short- and Long-term mRNA Levels of the *Retn, Iapp* , and *Drd5* Genes in Obese Mice Induced with High-fat Diet. In Vivo. 2018;32(4):813–7.

276. Zhang FF, Liu YH, Wang DW, Liu TS, Yang Y, Guo JM, et al. Obesity-induced reduced expression of the lncRNA ROIT impairs insulin transcription by downregulation of Nkx6.1 methylation. Diabetologia. 2020 Apr;63(4):811–24.
277. Giroud M, Tsokanos FF, Caratti G, Kotschi S, Khani S, Jouffe C, et al. HAND2 is a novel obesity-linked adipogenic transcription factor regulated by glucocorticoid signalling. Diabetologia. 2021 Aug;64(8):1850–65.
278. Pessoa Rodrigues C, Chatterjee A, Wiese M, Stehle T, Szymanski W, Shvedunova M, et al. Histone H4 lysine 16 acetylation controls central carbon metabolism and diet-induced obesity in mice. Nat Commun. 2021 Oct 27;12(1):6212.
279. Li X, Perelman D, Leong AK, Fragiadakis G, Gardner CD, Snyder MP. Distinct factors associated with short-term and long-term weight loss induced by low-fat or low-carbohydrate diet intervention. Cell Reports Medicine. 2022 Dec;3(12):100870.
280. Hou Q, Yan F, Li X, Liu H, Yang X, Dong X. ATP5me alleviates high glucose-induced myocardial cell injury. International Immunopharmacology. 2024 Mar;129:111626.
281. Lee JH, Park A, Oh KJ, Lee SC, Kim WK, Bae KH. The Role of Adipose Tissue Mitochondria: Regulation of Mitochondrial Function for the Treatment of Metabolic Diseases. IJMS. 2019 Oct 4;20(19):4924.
282. Lu Q, Zhu L. The Role of Semaphorins in Metabolic Disorders. IJMS. 2020 Aug 6;21(16):5641.
283. Vietor I, Cikes D, Piironen K, Vasakou T, Heimdörfer D, Gstir R, et al. The negative adipogenesis regulator Dlk1 is transcriptionally regulated by Ifrd1 (TIS7) and translationally by its orthologue Ifrd2 (SKMc15). eLife. 2023 Aug 21;12:e88350.
284. Shahid M, Javed AA, Chandra D, Ramsey HE, Shah D, Khan MF, et al. IEX-1 deficiency induces browning of white adipose tissue and resists diet-induced obesity. Sci Rep. 2016 Apr 11;6(1):24135.
285. Ravaud C, Esteve D, Villageois P, Bouloumie A, Dani C, Ladoux A. IER3 Promotes Expansion of Adipose Progenitor Cells in Response to Changes in Distinct Microenvironmental Effectors. Stem Cells. 2015 Aug 1;33(8):2564–73.
286. Minchin JEN, Dahlman I, Harvey CJ, Mejhert N, Singh MK, Epstein JA, et al. Plexin D1 determines body fat distribution by regulating the type V collagen microenvironment in visceral adipose tissue. Proc Natl Acad Sci USA. 2015 Apr 7;112(14):4363–8.

Supplementary Tables

Genome	Identifier	Repository site	Coverage	Contamination	Reference
Den D3		http://cdna.eva.mpg.de/denisova/	31.59x	<0.5%	Meyer et al. 2012
Altai Neandertal D5	ENA PRJEB1265	https://www.ebi.ac.uk/ena/browser/home	52x	1%	Prüfer et al. 2014
Chagyrskaya 8		http://ftp.eva.mpg.de/neandertal/Chagyrskaya/VCF/	28.36x	0.7%	Mafessoni et al. 2020
Mezmaiskaya 1	ENA PRJEB1757	http://cdna.eva.mpg.de/neandertal/Mezmaiskaya1/Pruefer_etal_2017_bam/	2.04	<1%	Prüfer et al. 2017
Spy 94a	ENA PRJEB21883	https://www.ebi.ac.uk/ena/browser/home	1.08	1.75%	Hajdinjak et al. 2018
Vindija 33.19	ENA PRJEB21882	https://www.ebi.ac.uk/ena/browser/home	31.32	<1%	Prüfer et al. 2017
Mezmaiskaya 2	ENA PRJEB21881	https://www.ebi.ac.uk/ena/browser/home	1.85	0.83%	Hajdinjak et al. 2018
Les Cottés Z4-1514	ENA PRJEB21875	https://www.ebi.ac.uk/ena/browser/home	2.94	0.18%	Hajdinjak et al. 2018
Goyet Q56.1	ENA PRJEB21870	https://www.ebi.ac.uk/ena/browser/home	2.30	0.89%	Hajdinjak et al. 2018
Ust'-Ishim	ENA PRJEB6622	https://www.ebi.ac.uk/ena/browser/home	42.72	<0.13%	Fu et al. 2014
Peştera Muierii 1	ENA PRJEB33172	https://www.ebi.ac.uk/ena/browser/home	17.00	1.5%	Svensson et al. 2021
Mal'ta	NCBI SRA SRP029640	https://www.ncbi.nlm.nih.gov/sra/SRX346836[accn]	1.21	1.6-2%	Raghavan et al. 2014
Loschbour, Motala 12, Stuttgart	ENA PRJEB6272	https://www.ebi.ac.uk/ena/browser/home	22.13, 2.46, 19.28	0.4% (Loschburg, Stuttgart) 0.35% (Motala 12)	Lazaridis et al. 2014
Human reference	GRCH38/hg38	https://zenodo.org/records/5146236			Schneider et al. 2017
Human Blacklisted regions	GRCH38/hg38	https://www.encodeproject.org/files/ENCFF356LFX/			Amemiya et al. 2019

Table S1. **Download information for the 15 ancient human genomes and the reference genome analyzed in this study**. From top to bottom, genomes are listed according to dating from the oldest to the most recent within each population: Denisovan (grey), Neandertal (orange), *Homo sapiens* (purple).

Genome	Identifier	Total Amount of Used Reads	Total Amount of Mapped Reads after Filtering	Percentage of Reads that Mapped after Filtering
Den 3	ENA (PRJEB3092)	4,789,230,696	1,329,869,335	27.77%
Altai Neandertal D5	ENA (PRJEB1265)	2,273,303,350	1,464,996,457	64.44%
Chagyrskaya 8	http://cdna.eva.mpg.de/neandertal/Chagyrskaya/	1,208,110,335	984,904,699	81.52%
Mezmaiskaya 1	ENA (PRJEB21195)	1,401,739,418	131,887,968	9.41%
Spy 94 a	ENA (PRJEB21883)	72,809,832	66,872,902	91.85%
Vindijia 33.19	http://cdna.eva.mpg.de/neandertal/Vindija/bam/Pruefer_etal_2017/Vindija33.19/	1,754,339,206	1,353,844,718	77.17%
Mezmaiskaya 2	ENA (SAMEA104168735, SAMEA104168734, SAMEA104168736)	111,777,820	103,837,327	92.90%
Les Cottés Z4-1514	ENA (PRJEB21875)	189,224,733	166,455,607	87.97%
Goyet Q56.1	ENA (PRJEB21870)	116,417,514	109,663,410	94.20%
Ust'-Ishim	ENA (PRJEB6622)	2,039,353,632	1,572,569,967	77.11%
Peştera Muierii 1	ENA (PRJEB33172)	611,030,108	498,336,516	81.56%
Mal'ta	NCBI (SRA SRP029640)	79,765,126	40,065,643	50.23%
Loschbour	ENA (PRJEB6272)	862,379,535	693,299,437	80.39%
Motala 12	ENA (PRJEB6272)	94,818,771	94,463,217	99.63%
Stuttgart	ENA (PRJEB6272)	791,983,938	643,277,473	81.22%

Table S1 (continued) Download information for the 15 ancient human genomes analyzed in this study. Number of reads available for mapping (first column), total reads mapped to hg38 after filtering (see Methods), and the percentage of reads mapped after filtering.

Fossil remains	Fossil details *anatomy* sex	Geographical location	Cave and layer	Time period (ybp - calibrated age)	Relevant publications for the genome published
Denisova D3	*Distal phalanx of the fifth finger.* Female	Central Asia: Central Siberia, Altai Moutain	Denisova Cave Layer 11.2, East Gallery	76200-51600	Meyer et al. Science 2012 Reich et al. Nature 2010 Douka et al. Nature 2019
Altai Neandertal D5	*Phalanx foot toe* Female	Central Asia: Central Siberia, Altai Moutain	Denisova Cave Layer 11.4	120000	Prüfer et al. Nature 2014
Chagyrskaya 8	*Distal manual phalanx* Female	Central Asia: Central Siberia, Charysh River, foothills of the Altai Moutain.	Chagyrskaya Cave	80000-60000	Mafessoni et al. PNAS 2020
Mezmaiskaya 1	*Almost complete skeleton* Still breast-feeding Neonate	Eastern Europe: Southern Russia, foothills of the North Caucasus	Mezmaiskaya Cave	70000-60000	Prüfer et al. Science 2017
Spy 94a	*Upper right molar associated to a maxillary fragment* Male	Western Europe: Belgium, Namur province	Cave	49800-42300	Haidinjak et al. Nature 2018 Devièse et al. PNAS 2021
Vindija 33.19	*Fragments of long bones* Female	Southern Europe: Croatia, Dinaric Alps	Vindija Cave	47400-44300	Prüfer et al. Science 2017
Mezmaiskaya 2	*Skull fragment* Male?	Eastern Europe: Foothills of the North Caucasus	Mezmaiskaya Cave	44.600-42.960	Haidinjak et al. Nature 2018
Les Cottés Z4-1514	*Tooth* NA	Western Europe: France, Saint-Pierre-de-Maillé	"Les Cottés" cave	43200-42500	Haidinjak et al. Nature 2018
Goyet Q56-1	*Right Femur* NA	Western Europe: Belgium, Namur province	"third cavern" of Goyet cave	42800-42100	Haidinjak et al. Nature 2018
Ust'-ishim	*Femora* Male	Central Asia: North-Western Siberia, Omsk province	Cave on the bank of the Irtysh River	45930-42900	Fu et al. Nature 2014
Peștera Muierii 1	*Temporal bone* Female	Eastern Europe: Romania, Carpatian mountains	"Women's Cave"	35257 +/-259	Svensson et al. Current Biology 2021
Malt'a 1	*Skeleton burial* Male	Central Asia: Central - Eastern Siberia, Bolshaya Belaya river	Open air site	25600	Raghavan et al. Nature 2014 Fiedel & Kuzmin, 2007
Loschbour	*Skeleton* Male	Central Europe: Luxembourg, Mullerthal	Rock shelter	8200-7900	Lazaridis et al. Nature 2014
Motala 12	*Maxilla (burial)* Male	Northern Europe: Sweden, Motala	Necropolis - Kanaljorden	8000-7600	Lazaridis et al. Nature 2014
Stuttgart	*Tooth (burial)*	Central Europe: South-Western, Germany	Stuttgart-Mühlhausen	7000	Lazaridis et al. Nature 2014

Table S2. Description of the fossils of the 15 individuals whose genomes are analyzed in this study. From top to bottom, genomes are listed according to dating from the oldest to the most recent within each population: Denisovan (grey), Neandertal (orange), *Homo sapiens* (purple).

Region	Populations	Identifier	Depository site	Sex	Source
Africa	Gambian1	SAMEA3302631	https://www.internationalgenome.org/data-portal/population/GambianSGDP	M	SGDP (Mallick et al., 2016)
	Gambian2	SAMEA3302886	https://www.internationalgenome.org/data-portal/population/GambianSGDP	F	SGDP (Mallick et al., 2016)
	Mandenka1	HGDP00905	https://www.internationalgenome.org/data-portal/population/MandenkaHGDP	M	HGDP (Bergström et al., 2020)
	Mandenka2	HGDP01201	https://www.internationalgenome.org/data-portal/population/MandenkaHGDP	F	HGDP (Bergström et al., 2020)
	Mandenka3	HGDP00912	https://www.internationalgenome.org/data-portal/population/MandenkaHGDP	M	HGDP (Bergström et al., 2020)
	BantuTsawana1	SAMEA3302630	https://www.internationalgenome.org/data-portal/population/BantuTswanaSGDP	M	SGDP (Mallick et al., 2016)
	BantuTsawana2	SAMEA3302849	https://www.internationalgenome.org/data-portal/population/BantuTswanaSGDP	M	SGDP (Mallick et al., 2016)
	KhomaniSan1	SAMEA3302800	https://www.internationalgenome.org/data-portal/population/KhomaniSanSGDP	F	SGDP (Mallick et al., 2016)
	KhomaniSan2	SAMEA3302708	https://www.internationalgenome.org/data-portal/population/KhomaniSanSGDP	F	SGDP (Mallick et al., 2016)
	BantuKenya1	HGDP01412	https://www.internationalgenome.org/data-portal/population/BantuKenyaHGDP	M	HGDP (Bergström et al., 2020)
	BantuKenya2	HGDP01416	https://www.internationalgenome.org/data-portal/population/BantuKenyaHGDP	M	HGDP (Bergström et al., 2020)
	BantuKenya3	HGDP01408	https://www.internationalgenome.org/data-portal/population/BantuKenyaHGDP	M	HGDP (Bergström et al., 2020)
	Somali1	SAMEA3302884	https://www.internationalgenome.org/data-portal/population/SomaliSGDP	F	SGDP (Mallick et al., 2016)
Eurasia	Icelandic1	SAMEA3302859	https://www.internationalgenome.org/data-portal/population/IcelandicSGDP	F	SGDP (Mallick et al., 2016)
	Icelandic2	SAMEA3302837	https://www.internationalgenome.org/data-portal/population/IcelandicSGDP	F	SGDP (Mallick et al., 2016)
	Saami1	SAMEA3302851	https://www.internationalgenome.org/data-portal/population/SaamiSGDP	M	SGDP (Mallick et al., 2016)
	Saami2	SAMEA3302875	https://www.internationalgenome.org/data-portal/population/SaamiSGDP	F	SGDP (Mallick et al., 2016)
	Mansi1	SAMEA3302762	https://www.internationalgenome.org/data-portal/population/MansiSGDP	M	SGDP (Mallick et al., 2016)
	Mansi2	SAMEA3302654	https://www.internationalgenome.org/data-portal/population/MansiSGDP	F	SGDP (Mallick et al., 2016)
	Tubalar1	SAMEA3302781	https://www.internationalgenome.org/data-portal/population/TubalarSGDP	F	SGDP (Mallick et al., 2016)
	Tubalar2	SAMEA3302764	https://www.internationalgenome.org/data-portal/population/TubalarSGDP	F	SGDP (Mallick et al., 2016)
	Altaian	SAMEA3302753	https://www.internationalgenome.org/data-portal/population/AltaianSGDP	M	SGDP (Mallick et al., 2016)
	Mongola1	SAMEA3302653	https://www.internationalgenome.org/data-portal/population/MongolaSGDP	M	SGDP (Mallick et al., 2016)
	Mongola2	SAMEA3302743	https://www.internationalgenome.org/data-portal/population/MongolaSGDP	F	SGDP (Mallick et al., 2016)
	Oroqen1	HGDP01206	https://www.internationalgenome.org/data-portal/population/OroqenHGDP	M	HGDP (Bergström et al., 2020)
	Oroqen2	HGDP01209	https://www.internationalgenome.org/data-portal/population/OroqenHGDP	F	HGDP (Bergström et al., 2020)
	Oroqen3	HGDP01207	https://www.internationalgenome.org/data-portal/population/OroqenHGDP	M	HGDP (Bergström et al., 2020)
	Even1	SAMEA3302720	https://www.internationalgenome.org/data-portal/population/EvenSGDP	F	SGDP (Mallick et al., 2016)
	Even2	SAMEA3302715	https://www.internationalgenome.org/data-portal/population/EvenSGDP	M	SGDP (Mallick et al., 2016)
	Even3	SAMEA3302752	https://www.internationalgenome.org/data-portal/population/EvenSGDP	M	SGDP (Mallick et al., 2016)
	Yakut1	HGDP00945	https://www.internationalgenome.org/data-portal/population/YakutHGDP	M	HGDP (Bergström et al., 2020)
	Yakut2	HGDP00947	https://www.internationalgenome.org/data-portal/population/YakutHGDP	M	HGDP (Bergström et al., 2020)
	Eskimo Chaplin	SAMEA3302660	https://www.internationalgenome.org/data-portal/population/EskimoChaplinSGDP	M	SGDP (Mallick et al., 2016)
	EskimoNaukan1	SAMEA3302856	https://www.internationalgenome.org/data-portal/population/EskimoNaukanSGDP	F	SGDP (Mallick et al., 2016)
	EskimoNaukan2	SAMEA3302831	https://www.internationalgenome.org/data-portal/population/EskimoNaukanSGDP	F	SGDP (Mallick et al., 2016)

Table S3. Dowloading information for the 64 modern genomes, organised in 5 populations (Africa, Eurasia, iddle East, South Asia and Oceania), in this study.

Region	Populations	Identifier	Depository site	Sex	Source
Middle East	Palestinian1	SAMEA3302652	https://www.internationalgenome.org/data-portal/population/PalestinianSGDP	M	HGDP (Bergström et al., 2020)
	Palestinian2	SAMEA3302646	https://www.internationalgenome.org/data-portal/population/PalestinianSGDP	F	HGDP (Bergström et al., 2020)
	Palestinian3	SAMEA3302897	https://www.internationalgenome.org/data-portal/population/PalestinianSGDP	F	HGDP (Bergström et al., 2020)
	Iranian1	SAMEA3302612	https://www.internationalgenome.org/data-portal/population/IranianSGDP	M	HGDP (Bergström et al., 2020)
	Iranian2	SAMEA3302874	https://www.internationalgenome.org/data-portal/population/IranianSGDP	M	HGDP (Bergström et al., 2020)
	Balochi1	HGDP00056	https://www.internationalgenome.org/data-portal/population/BalochiHGDP	M	HGDP (Bergström et al., 2020)
	Balochi2	HGDP00057	https://www.internationalgenome.org/data-portal/population/BalochiHGDP	M	HGDP (Bergström et al., 2020)
	Balochi3	HGDP00062	https://www.internationalgenome.org/data-portal/population/BalochiHGDP	M	HGDP (Bergström et al., 2020)
South Asia	Xibo1	HGDP01244	https://www.internationalgenome.org/data-portal/population/XiboHGDP	M	HGDP (Bergström et al., 2020)
	Xibo2	HGDP01245	https://www.internationalgenome.org/data-portal/population/XiboHGDP	M	HGDP (Bergström et al., 2020)
	Xibo3	HGDP01243	https://www.internationalgenome.org/data-portal/population/XiboHGDP	M	HGDP (Bergström et al., 2020)
	Lahu1	HGDP01317	https://www.internationalgenome.org/data-portal/population/LahuHGDP	M	HGDP (Bergström et al., 2020)
	Lahu2	HGDP01318	https://www.internationalgenome.org/data-portal/population/LahuHGDP	M	HGDP (Bergström et al., 2020)
	Lahu3	HGDP01319	https://www.internationalgenome.org/data-portal/population/LahuHGDP	M	HGDP (Bergström et al., 2020)
	Cambodian1	HGDP00714	https://www.internationalgenome.org/data-portal/population/CambodianHGDP	M	HGDP (Bergström et al., 2020)
	Cambodian2	HGDP00711	https://www.internationalgenome.org/data-portal/population/CambodianHGDP	M	HGDP (Bergström et al., 2020)
	Cambodian3	HGDP00715	https://www.internationalgenome.org/data-portal/population/CambodianHGDP	M	HGDP (Bergström et al., 2020)
Oceania	Dusun1	SAMEA3302688	https://www.internationalgenome.org/data-portal/population/DusunSGDP	F	SGDP (Mallick et al., 2016)
	Dusun2	SAMEA3302759	https://www.internationalgenome.org/data-portal/population/DusunSGDP	F	SGDP (Mallick et al., 2016)
	Igorot1	SAMEA3302696	https://www.internationalgenome.org/data-portal/population/IgorotSGDP	M	SGDP (Mallick et al., 2016)
	Igorot2	SAMEA3302724	https://www.internationalgenome.org/data-portal/population/IgorotSGDP	F	SGDP (Mallick et al., 2016)
	PapuanSepik1	HGDP00544	https://www.internationalgenome.org/data-portal/population/PapuanSepikHGDP	F	HGDP (Bergström et al., 2020)
	PapuanSepik2	SAMEA3302892	https://www.internationalgenome.org/data-portal/population/PapuanSGDP	M	SGDP (Mallick et al., 2016)
	PapuanSepik3	SAMEA3302626	https://www.internationalgenome.org/data-portal/population/PapuanSGDP	M	SGDP (Mallick et al., 2016)
	PapuanHighlands1	SAMEA3302726	https://www.internationalgenome.org/data-portal/population/PapuanSGDP	M	SGDP (Mallick et al., 2016)
	PapuanHighlands2	SAMEA3302721	https://www.internationalgenome.org/data-portal/population/PapuanSGDP	M	SGDP (Mallick et al., 2016)
	PapuanHighlands3	SAMEA3302738	https://www.internationalgenome.org/data-portal/population/PapuanSGDP	F	SGDP (Mallick et al., 2016)
	Australian1	SAMEA3302719	https://www.internationalgenome.org/data-portal/population/AustralianSGDP	F	SGDP (Mallick et al., 2016)
	Australian2	SAMEA3302681	https://www.internationalgenome.org/data-portal/population/AustralianSGDP	M	SGDP (Mallick et al., 2016)

Table S3 (continued). Dowloading information for the 64 modern genomes, organised in 5 populations (Africa, Eurasia, iddle East, South Asia and Oceania), in this study.

Gene Name	GeneID	Digestion-related tissue with highest rna expression (nTPM)	Digestion-related tissue with highest protein expression (level: high)
AMY1A	ENSG00000237763	Proximal Digestive Tract: Salivary Gland (8033)	Proximal Digestive Tract: Salivary Gland
AMY1B	ENSG00000174876	Proximal Digestive Tract: Salivary Gland (10158)	Proximal Digestive Tract: Salivary Gland
AMY1C	ENSG00000187733	Proximal Digestive Tract: Salivary Gland (8147)	Proximal Digestive Tract: Salivary Gland Pancreas
SULT1A3	ENSG00000261052	Gastrointestinal Tract: Duodenum (233)	Gastrointestinal Tract: Small Intestin
SULT1A4	ENSG00000213648	Gastrointestinal Tract: Small Intestin (60.1)	Gastrointestinal Tract: Small Intestin
SLX1A	ENSG00000132207	Liver & Gallbladder: Liver (57.9)	Gastrointestinal Tract: Duodenum
SLX1B	ENSG00000181625	Gastrointestinal Tract: Duodenum (43.9)	Gastrointestinal Tract: Duodenum
UGT2B28	ENSG00000135226	NA	Gastrointestinal Tract: Colon
ACOT1	ENSG00000184227	Liver & Gallbladder: Liver (43.2)	NA
CLPS	ENSG00000137392	Pancreas: 253763	NA
CLPSL1	ENSG00000204140	NA	NA
LPA	ENSG00000198670	Liver & Gallbladder: Liver (74.5)	NA
F7	ENSG00000057593	Liver & Gallbladder: Liver (156)	NA
ZNG1E	ENSG00000147996	Liver & Gallbladder: Liver (40.9)	NA
ZNG1F	ENSG00000215126	Liver & Gallbladder: Liver (7.1)	Liver & Gallbladder: Liver
ZNG1C	ENSG00000196873	NA	NA
RFPL4A	ENSG00000223638	Liver & Gallbladder: Liver (3.3)	NA
RFPL4AL1	ENSG00000229292	Liver & Gallbladder: Liver (0.6)	NA
PDX1	ENSG00000139515	Gastrointestinal Tract: Duodenum (40.6)	Gastrointestinal Tract: Duodenum Pancreas
ACOX3	ENSG00000087008	Proximal Digestive Tract: esophagus (25.8)	Gastrointestinal Tract: Colon
HAND1	ENSG00000113196	Gastrointestinal Tract: Colon (63.9)	NA
WFS1	ENSG00000109501	NA	Pancreas
H2AC11	ENSG00000196787	NA	Gastrointestinal Tract: Duodenum Pancreas
MST1	ENSG00000173531	Liver & Gallbladder: Liver (905)	NA
H1-2	ENSG00000187837	NA	Gastrointestinal Tract: Colon Liver & Gallbladder: Gallbladder Proximal Digestive Tract: Esophagus Pancreas
CD24	ENSG00000272398	NA	Gastrointestinal Tract: Colon Proximal Digestive Tract: Esophagus
BOLA2	ENSG00000183336	NA	NA
BOLA2B	ENSG00000169627	NA	NA
TPSAB1	ENSG00000172236	Gastrointestinal Tract: Small Intestine (144)	NA
TPSB2	ENSG00000197253	Gastrointestinal Tract: Small Intestine (142)	NA
IFNL2	ENSG00000183709	Pancreas (0.4)	NA
IFNL3	ENSG00000197110	NA	NA
C4A	ENSG00000244731	Liver & Gallbladder: Liver (328)	NA
C4B	ENSG00000224389	Liver & Gallbladder: Liver (375)	NA
LYPD8	ENSG00000259823	Gastrointestinal Tract: Colon (209)	NA
PECAM1	ENSG00000261371	NA	Gastrointestinal Tract: Colon
GRTP1	ENSG00000139835	Liver & Gallbladder: Liver (81.8)	Gastrointestinal Tract: Duodenum Liver & Gallbladder: Gallbladder
CLPTM1L	ENSG00000049656	Liver & Gallbladder: Liver (131)	Pancreas
ZNF431	ENSG00000196705	Proximal Digestive Tract: Esophagus (18.7)	NA
N4BP3	ENSG00000145911	Proximal Digestive Tract: Esophagus (15.2)	NA
STAB1	ENSG00000010327	NA	Gastrointestinal Tract: Colon Proximal Digestive Tract: Esophagus
ORM2	ENSG00000228278	Liver & Gallbladder: Liver (8653)	NA
ORM1	ENSG00000229314	Liver & Gallbladder: Liver (41893)	NA
UQCRFS1	ENSG00000169021	Proximal Digestive Tract: Tongue	Gastrointestinal Tract: Colon Liver & Gallbladder: Gallbladder Proximal Digestive Tract: Esophagus Pancreas

Table S4. Primary and secondary tissues expressing the 50 genes. The genes listed in **Table 1** and **Table 2** are expressed across various tissues. For gene clusters, each gene is considered individually, totalling 62 genes in this table. We specify the tissues where each gene or gene copy in a cluster shows strong expression, with RNA and protein expression data in columns three and four, respectively. This information is sourced from the Human Protein Atlas Database (www.proteinatlas.org/), which provides RNA and protein expression data alongside additional functional details. Ensembl gene ID are listed in the second column, and genes are color coded according to **Tables 1** and **2**.

TCAF2	ENSG00000170379	Proximal Digestive Tract: Esophagus (8.9)	NA
TCAF2C	ENSG00000283528	Proximal Digestive Tract: Esophagus (0.3)	NA
TMEM86B	ENSG00000180089	Liver & Gallbladder: Liver (16.3)	NA
ATP4B	ENSG00000186009	Gastrointestinal Tract: Stomach (1067)	NA
HGFAC	ENSG00000109758	Liver & Gallbladder: Liver (212)	NA
SLC9A3	ENSG00000066230	Gastrointestinal Tract: Colon (91)	Gastrointestinal Tract: Colon Liver & Gallbladder: Gallbladder
FGFRL1	ENSG00000127418	Pancreas (69.5)	NA
DRD5	ENSG00000169676	Gastrointestinal Tract: Stomach (4.3)	NA
IDUA	ENSG00000127415	NA	Liver & Gallbladder: Gallbladder Pancreas
ATP5ME	ENSG00000169020	NA	Gastrointestinal Tract: Colon Liver & Gallbladder: Gallbladder Proximal Digestive Tract: Esophagus Pancreas
SEMA3B	ENSG00000012171	NA	Gastrointenstinal Tract: Duodenum Liver & Gallbladder: Gallbladder Pancreas
IFRD2	ENSG00000214706	NA	Liver & Gallbladder: Liver
IER3 (IEX-1)	ENSG00000137331	Proximal Digestive Tract: Esophagus (251)	Pancreas
NKX6−1	ENSG00000163623	Proximal Digestive Tract: Esophagus (8.9)	Pancreas
HAND2	ENSG00000164107	Gastrointestinal Tract: Colon (58.5)	NA
PLXND1	ENSG00000004399	NA	Gastrointestinal Tract: Colon Proximal Digestive Tract: Oral Mucosa
H4C5	ENSG00000276966	NA	Gastrointestinal Tract: Colon Liver & Gallbladder: Gallbladder Proximal Digestive Tract: Oral Mucosa
H4C16	ENSG00000197837	Proximal Digestive Tract: Esophagus (2.6)	Gastrointestinal Tract: Colon Liver & Gallbladder: Gallbladder Pancreas Proximal Digestive Tract: Esophagus

Table S4 (continued)

Genes (ENS ID ~ Gene Name)	Count	Neanderthal	Ancient *sapiens*
Digestion			
GO:0007586~digestion	**5**		
ENSG00000204140 ~ CLPSL1		+	+
ENSG00000137392 ~ CLPS		+	+
Carbohydrates			
GO:0044723~single-organism carbohydrate metabolic process	**9**		
ENSG00000237763/ENSG00000174876/ENSG00000187733 ~ AMY1A/B/C		-	
ENSG00000135226 ~ UGT2B28		-	
ENSG00000173531 ~ MST1		+	+
ENSG00000139515 ~ PDX1		+	+
ENSG00000127415 ~ IDUA		+	+
ENSG00000137331 ~ IER3 (IEX-1)			+
GO:0005975~carbohydrate metabolic process	**9**		
ENSG00000237763/ENSG00000174876/ENSG00000187733 ~ AMY1A/B/C		-	
ENSG00000137331 ~ IER3 (IEX-1)			+
ENSG00000127415 ~ IDUA		+	+
ENSG00000135226 ~ UGT2B28		-	
ENSG00000139515 ~ PDX1		+	+
ENSG00000173531 ~ MST1		+	+
GO:0009311~oligosaccharide metabolic process	**4**		
ENSG00000237763/ENSG00000174876/ENSG00000187733 ~ AMY1A/B/C		-	
ENSG00000127415 ~ IDUA		+	+
GO:0010677~negative regulation of cellular carbohydrate metabolic process	**3**		
ENSG00000137331 ~ IER3 (IEX-1)			+
ENSG00000173531 ~ MST1		+	+
GO:0045912~negative regulation of carbohydrate metabolic process	**3**		
ENSG00000137331 ~ IER3 (IEX-1)			+
ENSG00000173531 ~ MST1		+	+
GO:0009756~carbohydrate mediated signaling	**2**		
ENSG00000163623 ~ NKX6−1		+	+
ENSG00000139515 ~ PDX1		+	+
GO:0005996~monosaccharide metabolic process	**4**		
ENSG00000139515 ~ PDX1		+	+
ENSG00000173531 ~ MST1		+	+
ENSG00000135226 ~ UGT2B28		-	
GO:0044262~cellular carbohydrate metabolic process	**4**		
ENSG00000137331 ~ IER3 (IEX-1)			+
ENSG00000127415 ~ IDUA		+	+
ENSG00000173531 ~ MST1		+	+
GO:0010182~sugar mediated signaling pathway	**2**		
ENSG00000163623 ~ NKX6−1		+	+
ENSG00000139515 ~ PDX1		+	+
GO:0010255~glucose mediated signaling pathway	**2**		
ENSG00000163623 ~ NKX6−1		+	+
ENSG00000139515 ~ PDX1		+	+
GO:0009757~hexose mediated signaling	**2**		
ENSG00000163623 ~ NKX6−1		+	+
ENSG00000139515 ~ PDX1		+	+

Table S5. **Gene Ontology terms associated with the 50 genes presenting differential CNV in archaic populations compared to the reference human genome.** The first column report the GO term and the GO description for the genes presenting a gain (third column) or loss (fourth column) of copy number in *Hn* and *Hs* populations compared to the reference human genome. The genes ZNG1E/F/C, H2AC11, GRTP1, CLPTM1L, FGFRL1, SEMA3B, IFRD2, and PLXND1 are currently missing from any documented digestion-related GO term and are not listed in the table.

Lipids			
GO:0006629~lipid metabolic process	**10**		
ENSG00000087008 ~ ACOX3		+	+
ENSG00000198670 ~ LPA		+	
ENSG00000180089 ~ TMEM86B		-	-
ENSG00000137392 ~ CLPS		+	+
ENSG00000204140 ~ CLPSL1		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
ENSG00000184227 ~ ACOT1		-	
ENSG00000135226 ~ UGT2B28		-	
GO:0016803~ether hydrolase activity	**1**		
ENSG00000180089 ~ TMEM86B		-	-
GO:0008202~steroid metabolic process	**1**		
ENSG00000135226 ~ UGT2B28		-	
GO:0006631~fatty acid metabolic process	**1**		
ENSG00000184227 ~ ACOT1		-	
GO:0044255~cellular lipid metabolic process	**1**		
ENSG00000180089 ~ TMEM86B		-	-
GO:0008202~steroid metabolic process	**1**		
ENSG00000135226 ~ UGT2B28		-	
Proteins - Liver			
GO:0016491~oxidoreductase activity	**1**		
ENSG00000169021 ~ UQCRFS1		+	+
GO:0016627~oxidoreductase activity, acting on the CH-CH group of donors	**1**		
ENSG00000087008 ~ ACOX3		+	+
GO:1990204~oxidoreductase complex	**1**		
ENSG00000139515 ~ PDX1		+	+
GO:0004252~serine-type endopeptidase activity	**2**		
ENSG00000109758~HGFAC		+	+
ENSG00000172236 ~ TPSAB1/B2		-	-
PC00220~transferase (transaminase activity)	**1**		
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
Small Molecules			
GO:0055086~nucleobase-containing small molecule metabolic process	**6**		
ENSG00000137331 ~ IER3 (IEX-1)			+
ENSG00000169676 ~ DRD5		+	+
ENSG00000169021 ~ UQCRFS1		+	+
ENSG00000169020 ~ ATP5ME		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0009150~purine ribonucleotide metabolic process	**6**		
ENSG00000137331 ~ IER3 (IEX-1)			+
ENSG00000169676 ~ DRD5		+	+
ENSG00000169021 ~ UQCRFS1		+	+
ENSG00000169020 ~ ATP5ME		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0045934~negative regulation of nucleobase-containing compound metabolic process	**10**		
ENSG00000196705 ~ ZNF431		+	+
ENSG00000137331 ~ IER3 (IEX-1)			+
ENSG00000113196 ~ HAND1			+

Table S5 (continued)

ENSG00000132207/ENSG00000181625 ~ SLX1A/B		-	
ENSG00000163623 ~ NKX6−1		+	+
ENSG00000139515 ~ PDX1		+	+
ENSG00000109501 ~ WFS1		+	+
ENSG00000187837 ~ H1-2			+
GO:0009259~ribonucleotide metabolic process	**6**		
ENSG00000137331 ~ IER3 (IEX-1)			+
ENSG00000169676 ~ DRD5		+	+
ENSG00000169021 ~ UQCRFS1		+	+
ENSG00000169020 ~ ATP5ME		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0034035~purine ribonucleoside bisphosphate metabolic process	**2**		
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0006082~organic acid metabolic process	**8**		
ENSG00000087008 ~ ACOX3		+	+
ENSG00000137331 ~ IER3 (IEX-1)			+
ENSG00000127415 ~ IDUA		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
ENSG00000184227 ~ ACOT1		-	
ENSG00000135226 ~ UGT2B28		-	
GO:0045721~negative regulation of gluconeogenesis	**2**		
ENSG00000173531 ~ MST1		+	+
GO:0044281~small molecule metabolic process	**15**		
ENSG00000180089 ~ TMEM86B		-	-
ENSG00000139515 ~ PDX1		+	+
ENSG00000169021 ~ UQCRFS1		+	+
ENSG00000169020 ~ ATP5ME		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
ENSG00000173531 ~ MST1		+	+
ENSG00000135226 ~ UGT2B28		-	
ENSG00000087008 ~ ACOX3		+	+
ENSG00000137331 ~ IER3 (IEX-1)			+
ENSG00000169676 ~ DRD5		+	+
ENSG00000127415 ~ IDUA		+	+
ENSG00000184227 ~ ACOT1		-	
GO:0044282~small molecule catabolic process	**5**		
ENSG00000087008 ~ ACOX3		+	+
ENSG00000127415 ~ IDUA		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0060359~response to ammonium ion	**3**		
ENSG00000169676 ~ DRD5		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0042445~hormone metabolic process	**1**		
ENSG00000135226 ~ UGT2B28		-	
Molecules & Ion Transport			
GO:0098657~import into cell	**1**		
ENSG00000066230 ~ SLC9A3		+	+
GO:1990573~potassium ion import across plasma membrane	**1**		
ENSG00000186009 ~ ATP4B		+	+

Table S5 (continued)

GO:0071805~potassium ion transmembrane transport	**1**		
ENSG00000066230 ~ SLC9A3		+	+
GO:0015079~potassium ion transmembrane transporter activity	**1**		
ENSG00000066230 ~ SLC9A3		+	+
GO:0051139~metal cation:proton antiporter activity	**1**		
ENSG00000066230 ~ SLC9A3		+	+
GO:0043271~negative regulation of monoatomic ion transport	**2**		
ENSG00000170379 ~ TACF2		-	
ENSG00000283528 ~ TACF2C		-	
GO:0051222~positive regulation of protein transport	**2**		
ENSG00000170379 ~ TACF2		-	
ENSG00000283528 ~ TACF2C		-	
GO:0033157~regulation of intracellular protein transport	**2**		
ENSG00000170379 ~ TACF2		-	
ENSG00000283528 ~ TACF2C		-	
GO:0034765~regulation of monoatomic ion transmembrane transport	**2**		
ENSG00000170379 ~ TACF2		-	
ENSG00000283528 ~ TACF2C		-	
GO:0032412~regulation of monoatomic ion transmembrane transporter activity	**2**		
ENSG00000170379 ~ TACF2		-	
ENSG00000283528 ~ TACF2C		-	
GO:0035725~sodium ion transmembrane transport	**2**		
ENSG00000186009 ~ ATP4B		+	+
ENSG00000066230 ~ SLC9A3		+	+
GO:0015081~sodium ion transmembrane transporter activity	**1**		
ENSG00000066230 ~ SLC9A3		+	+
GO:1902495~transmembrane transporter complex	**1**		
ENSG00000186009 ~ ATP4B		+	+
Nitrogen			
GO:0043279~response to alkaloid	**3**		
ENSG00000169676 ~ DRD5		+	+
ENSG00000139515 ~ PDX1		+	+
ENSG00000163623 ~ NKX6−1		+	+
GO:0010243~response to organonitrogen compound	**7**		
ENSG00000169676 ~ DRD5		+	+
ENSG00000057593 ~ F7		+	+
ENSG00000163623 ~ NKX6−1		+	+
ENSG00000139515 ~ PDX1		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
ENSG00000186009 ~ ATP4B		+	+
GO:0071870~cellular response to catecholamine stimulus	**3**		
ENSG00000169676 ~ DRD5		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0071868~cellular response to monoamine stimulus	**3**		
ENSG00000169676 ~ DRD5		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0071869~response to catecholamine	**3**		
ENSG00000169676 ~ DRD5		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	

Table S5 (continued)

GO:0071867~response to monoamine	**3**		
ENSG00000169676 ~ DRD5		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0006584~catecholamine metabolic process	**3**		
ENSG00000164107 ~ HAND2			+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0009712~catechol-containing compound metabolic process	**3**		
ENSG00000164107 ~ HAND2			+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0006637~acyl-CoA metabolic process	**1**		
ENSG00000184227 ~ ACOT1		-	
GO:1904431~positive regulation of t-circle formation	**2**		
ENSG00000132207/ENSG00000181625 ~ SLX1A/B		-	
Glucoronidation			
GO:0009812~flavonoid metabolic process	**2**		
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0009410~response to xenobiotic stimulus	**5**		
ENSG00000163623 ~ NKX6−1		+	+
ENSG00000139515 ~ PDX1		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
ENSG00000135226 ~ UGT2B28		-	
GO:0006805~xenobiotic metabolic process	**3**		
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
ENSG00000135226 ~ UGT2B28		-	
Sulfur			
GO:0097428~protein maturation by iron-sulfur cluster transfer	**2**		
ENSG00000183336/ENSG00000169627 ~ BOLA2/B		-	
GO:0006790~sulfur compound metabolic process	**6**		
ENSG00000127415 ~ IDUA		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
ENSG00000184227 ~ ACOT1		-	
ENSG00000183336/ENSG00000169627 ~ BOLA2/B		-	
GO:0051923~sulfation	**2**		
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
Homeostasis			
GO:0006873~cellular ion homeostasis	**7**		
ENSG00000066230 ~ SLC9A3		+	+
ENSG00000169676 ~ DRD5		+	+
ENSG00000186009 ~ ATP4B		+	+
ENSG00000183336/ENSG00000169627 ~ BOLA2/B		-	
ENSG00000109501 ~ WFS1		+	+
ENSG00000272398 ~ CD24			-
GO:0055082~cellular chemical homeostasis	**9**		
ENSG00000066230 ~ SLC9A3		+	+
ENSG00000169676 ~ DRD5		+	+
ENSG00000163623 ~ NKX6−1		+	+
ENSG00000139515 ~ PDX1		+	+
ENSG00000186009 ~ ATP4B		+	+
ENSG00000183336/ENSG00000169627 ~ BOLA2/B		-	

Table S5 (continued)

ENSG00000109501 ~ WFS1		+	+
ENSG00000272398 ~ CD24			-
GO:0019725~cellular homeostasis	**9**		
ENSG00000066230 ~ SLC9A3		+	+
ENSG00000169676 ~ DRD5		+	+
ENSG00000163623 ~ NKX6−1		+	+
ENSG00000139515 ~ PDX1		+	+
ENSG00000186009 ~ ATP4B		+	+
ENSG00000183336/ENSG00000169627 ~ BOLA2/B		-	
ENSG00000109501 ~ WFS1		+	+
ENSG00000272398 ~ CD24			-
GO:0048878~chemical homeostasis	**10**		
ENSG00000087008 ~ ACOX3		+	+
ENSG00000066230 ~ SLC9A3		+	+
ENSG00000169676 ~ DRD5		+	+
ENSG00000163623 ~ NKX6−1		+	+
ENSG00000139515 ~ PDX1		+	+
ENSG00000186009 ~ ATP4B		+	+
ENSG00000183336/ENSG00000169627 ~ BOLA2/B		-	
ENSG00000109501 ~ WFS1		+	+
ENSG00000272398 ~ CD24			-
GO:0030003~cellular cation homeostasis	**7**		
ENSG00000066230 ~ SLC9A3		+	+
ENSG00000169676 ~ DRD5		+	+
ENSG00000186009 ~ ATP4B		+	+
ENSG00000183336/ENSG00000169627 ~ BOLA2/B		-	
ENSG00000109501 ~ WFS1		+	+
ENSG00000272398 ~ CD24			-
GO:0006873~cellular ion homeostasis	**7**		
ENSG00000066230 ~ SLC9A3		+	+
ENSG00000169676 ~ DRD5		+	+
ENSG00000186009 ~ ATP4B		+	+
ENSG00000183336/ENSG00000169627 ~ BOLA2/B		-	
ENSG00000109501 ~ WFS1		+	+
ENSG00000272398 ~ CD24			-
GO:0055080~cation homeostasis	**7**		
ENSG00000066230 ~ SLC9A3		+	+
ENSG00000169676 ~ DRD5		+	+
ENSG00000186009 ~ ATP4B		+	+
ENSG00000183336/ENSG00000169627 ~ BOLA2/B		-	
ENSG00000109501 ~ WFS1		+	+
ENSG00000272398 ~ CD24			-
GO:0098771~inorganic ion homeostasis	**7**		
ENSG00000066230 ~ SLC9A3		+	+
ENSG00000169676 ~ DRD5		+	+
ENSG00000186009 ~ ATP4B		+	+
ENSG00000183336/ENSG00000169627 ~ BOLA2/B		-	
ENSG00000109501 ~ WFS1		+	+
ENSG00000272398 ~ CD24			-

Table S5 (continued)

GO:0006875~cellular metal ion homeostasis	**6**		
ENSG00000169676 ~ DRD5		+	+
ENSG00000186009 ~ ATP4B		+	+
ENSG00000183336/ENSG00000169627 ~ BOLA2/B			
ENSG00000109501 ~ WFS1		+	+
ENSG00000272398 ~ CD24			-
GO:0050801~ion homeostasis	**7**		
ENSG00000066230 ~ SLC9A3		+	+
ENSG00000169676 ~ DRD5		+	+
ENSG00000186009 ~ ATP4B		+	+
ENSG00000183336/ENSG00000169627 ~ BOLA2/B		-	
ENSG00000109501 ~ WFS1		+	+
ENSG00000272398 ~ CD24			-
GO:0055065~metal ion homeostasis	**6**		
ENSG00000169676 ~ DRD5		+	+
ENSG00000186009 ~ ATP4B		+	+
ENSG00000183336/ENSG00000169627 ~ BOLA2/B		-	
ENSG00000109501 ~ WFS1		+	+
ENSG00000272398 ~ CD24			-
GO:0055074~calcium ion homeostasis	**1**		
ENSG00000109501 ~ WFS1		+	+
Catabolism			
GO:0019336~phenol-containing compound catabolic process	**2**		
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0018958~phenol-containing compound metabolic process	**3**		
ENSG00000164107 ~ HAND2			+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0044712~single-organism catabolic process	**9**		
ENSG00000087008 ~ ACOX3		+	+
ENSG00000137331 ~ IER3 (IEX-1)			+
ENSG00000137392 ~ CLPS		+	+
ENSG00000204140 ~ CLPSL1		+	+
ENSG00000127415 ~ IDUA		+	+
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
ENSG00000109501 ~ WFS1		+	+
GO:0034310~primary alcohol catabolic process	**2**		
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0006068~ethanol catabolic process	**2**		
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
GO:0006067~ethanol metabolic process	**2**		
ENSG00000261052/ENSG00000213648 ~ SULT1A3/4		-	
Immunity - Antiobiotic			
GO:0006955~immune response	**13**		
ENSG00000183709/ENSG00000197110 ~ IFNL2/3		-	-
ENSG00000223638 /ENSG00000229292 ~ RFPL4A/L1		+	+
ENSG00000173531 ~ MST1		+	+
ENSG00000244731/ENSG00000224389 ~ C4A/B		+	+
ENSG00000261371 ~ PECAM1			-
ENSG00000145911 ~ N4BP3		+	+

Table S5 (continued)

ENSG00000272398 ~ CD24			-
GO:0006952~defense response	**16**		
ENSG00000183709/ENSG00000197110 ~ IFNL2/3		-	-
ENSG00000223638 /ENSG00000229292 ~ RFPL4A/L1		+	+
ENSG00000173531 ~ MST1		+	+
ENSG00000244731/ENSG00000224389 ~ C4A/B		+	+
ENSG00000228278 ~ ORM2		+	
ENSG00000229314 ~ ORM1		+	
ENSG00000172236 ~ TPSAB1/B2		-	-
ENSG00000145911 ~ N4BP3		+	+
ENSG00000137331 ~ IER3 (IEX-1)			+
ENSG00000259823 ~ LYPD8			-
ENSG00000010327 ~ STAB1		+	+
GO:0002376~immune system process	**17**		
ENSG00000183709/ENSG00000197110 ~ IFNL2/3		-	-
ENSG00000223638 /ENSG00000229292 ~ RFPL4A/L1		+	+
ENSG00000173531 ~ MST1		+	+
ENSG00000244731/ENSG00000224389 ~ C4A/B		+	+
ENSG00000261371 ~ PECAM1			-
ENSG00000276966/ENSG00000197837 ~ H4C16/5			+
ENSG00000145911 ~ N4BP3		+	+
ENSG00000057593 ~ F7		+	+
ENSG00000164107 ~ HAND2			+
ENSG00000272398 ~ CD24			-
GO:0045087~innate immune response	**10**		
ENSG00000244731/ENSG00000224389 ~ C4A/B		+	+
ENSG00000145911 ~ N4BP3		+	+
ENSG00000183709/ENSG00000197110 ~ IFNL2/3		-	-
ENSG00000223638 /ENSG00000229292 ~ RFPL4A/L1		+	+
ENSG00000173531 ~ MST1		+	+
GO:0006959~humoral immune response	**4**		
ENSG00000244731/ENSG00000224389 ~ C4A/B		+	+
GO:0098542~defense response to other organism	**6**		
ENSG00000183709/ENSG00000197110 ~ IFNL2/3		-	-
ENSG00000259823 ~ LYPD8			-
ENSG00000010327 ~ STAB1		+	+
GO:0009617~response to bacterium	**8**		
ENSG00000244731/ENSG00000224389 ~ C4A/B		+	+
ENSG00000137392 ~ CLPS		+	+
ENSG00000259823 ~ LYPD8			-
ENSG00000010327 ~ STAB1		+	+
ENSG00000186009 ~ ATP4B		+	+
ENSG00000272398 ~ CD24			-
GO:2000675~negative regulation of type B pancreatic cell apoptotic process	**2**		
ENSG00000109501 ~ WFS1		+	+
ENSG00000139515 ~ PDX1		+	+
GO:0002682~regulation of immune system process	**12**		
ENSG00000276966/ENSG00000197837 ~ H4C16/5			+
ENSG00000244731/ENSG00000224389 ~ C4A/B		+	+
ENSG00000057593 ~ F7		+	+
ENSG00000228278 ~ ORM2		+	
ENSG00000229314 ~ ORM1		+	
ENSG00000183709/ENSG00000197110 ~ IFNL2/3		-	-
ENSG00000173531 ~ MST1		+	+
ENSG00000272398 ~ CD24			-

Table S5 (continued)

Genes	GRCh38/hg38	Denisova 3	Altai Neandertal D5	Chagyrskaya 8	Mezmaiskaya 1	Spy 94a	Vindija 33.19	Mezmaiskaya 2	Les Cottés Z4-1514	Goyet Q56.1	Ust'-Ishim	Peştera Muierii 1	Mal'ta	Loschbour	Motala 12	Stuttgart	Neandertal	H. sapiens
AMY1A/B/C	3	1.58	1.60	1.36	1.85	2.09	1.36	2.33	2.55	2.26	2.94	3.19	1.65	5.06	2.07	6.10	-	
SULT1A3/4	2	1.04	1.07	0.98	0.95	0.95	1.01	0.98	0.85	0.86	2.31	2.45	3.35	2.34	1.83	2.47	-	
SLX1A/B	2	1.03	0.96	0.92	1.02	1.14	0.99	0.93	1.03	1.05	2.19	2.47	3.11	2.34	1.80	2.39	-	
UGT2B28	1	0.82	0.14	0.12	0.27	0.19	0.17	0.21	0.19	0.14	0.83	0.84	0.35	0.75	0.45	0.46	-	
LPA	1	1.42	1.38	1.38	1.53	1.89	1.47	2.08	1.86	2.01	1.37	2.08	1.98	1.41	1.39	1.26	+	
ZNG1E/F/C	3	3.92	4.05	3.51	3.38	4.18	3.67	3.64	4.23	4.16	3.34	4.92	2.21	4.12	3.44	3.56	+	
ACOT1	1	0.28	0.27	0.27	0.32	0.35	0.35	0.40	0.42	0.31	0.77	0.70	0.84	0.22	0.65	0.77	-	
CLPS	1	2.31	2.42	2.46	2.77	2.86	2.65	3.59	3.03	3.56	1.36	1.89	3.29	2.65	1.80	2.01	+	+
CLPSL1	1	1.58	1.64	1.63	1.80	1.63	1.75	2.06	2.18	1.91	1.15	1.58	2.11	1.77	1.51	1.49	+	+
RFPL4A/L1	2	2.94	2.83	2.88	2.87	2.40	2.75	2.57	3.22	2.26	3.84	3.96	3.44	3.60	4.04	2.42	+	+
F7	1	1.93	2.07	2.03	1.84	1.87	1.84	1.94	1.43	1.91	1.96	1.52	2.08	2.03	2.32	2.27	+	+
PDX1	1	1.71	1.89	1.88	2.08	1.40	1.97	1.82	2.00	1.51	1.96	1.59	2.03	2.25	2.16	1.91	+	+
WFS1	1	1.59	1.80	1.74	1.77	1.65	1.69	1.61	1.52	1.76	1.69	1.36	1.84	1.86	1.62	1.98	+	+
H2AC11	1	1.52	1.65	1.46	1.68	0.18	1.70	2.09	1.73	1.60	1.82	1.60	2.15	1.78	2.15	1.98	+	+
ACOX3	1	1.43	1.56	1.53	1.55	1.40	1.56	1.43	1.43	1.48	1.51	1.36	1.67	1.75	1.49	1.80		+
HAND1	1	1.37	1.44	1.31	1.42	0.96	1.41	1.52	1.72	1.61	1.41	1.30	2.15	1.77	1.66	1.51		+
MST1	1	1.99	2.39	1.94	2.27	2.22	1.97	2.19	2.42	2.34	1.77	1.78	1.54	2.34	2.15	2.24	+	+
H1-2	1	1.25	1.62	1.38	1.85	1.22	1.61	1.22	1.83	1.09	1.49	1.56	2.26	2.06	2.71	1.83		+
CD24	1	1.11	0.35	0.25	1.08	0.79	0.25	1.33	1.21	1.30	0.39	0.98	0.58	0.40	0.31	0.39		-
BOLA2/2B	2	0.83	0.90	0.66	0.95	0.94	0.76	0.90	0.83	0.76	2.01	2.13	4.24	2.22	1.75	1.92	-	
CLPTM1L	1	1.37	1.52	1.47	1.53	1.58	1.51	1.60	1.58	1.59	1.47	1.30	1.46	1.66	1.59	1.77	+	
ORM1	1	1.05	2.01	1.85	2.64	1.20	1.72	4.02	2.66	2.39	1.04	0.83	1.09	1.03	1.19	3.45	+	
ORM2	1	1.05	1.87	1.86	2.67	1.33	1.78	3.68	2.66	2.16	1.05	0.93	0.86	1.07	0.93	1.99	+	
TPSAB1/B2	2	1.32	1.49	1.33	1.16	1.73	1.33	1.07	1.08	1.26	1.41	1.24	1.17	1.54	1.31	1.53	-	-
IFNL2/3	2	1.14	1.26	1.17	1.20	1.10	1.11	1.28	1.14	1.40	1.13	1.07	1.26	0.93	1.20	1.19	-	-
C4A/B	2	2.86	1.46	2.80	2.54	3.27	2.73	2.62	1.83	2.96	2.74	2.63	4.51	3.26	2.81	3.19	+	+
GRTP1	1	1.57	1.73	1.83	1.69	1.45	1.77	1.62	1.56	1.62	1.71	1.53	1.56	2.03	1.75	2.04	+	+
UQCRFS1	1	1.62	1.57	2.46	1.95	2.09	1.46	3.94	2.82	3.37	1.52	1.57	1.38	1.60	1.60	1.58	+	+
ZNF431	1	1.63	1.65	1.69	1.49	1.69	1.63	1.56	1.62	1.61	1.62	1.64	1.64	1.81	1.56	1.63	+	+
N4BP3	1	1.45	1.55	1.48	1.64	1.35	1.60	1.55	1.45	1.51	1.52	1.28	1.72	1.62	1.47	1.73	+	+
STAB1	1	1.49	1.62	1.51	1.65	1.42	1.53	1.52	1.27	1.76	1.48	1.36	1.50	1.67	1.55	1.80	+	+
LYPD8	1	0.87	0.35	0.32	0.86	0.73	0.31	0.83	0.77	0.77	0.33	0.86	0.36	0.31	0.33	0.30		-
PECAM1	1	0.89	0.31	0.21	0.87	0.87	0.24	0.86	0.88	0.86	0.32	0.86	0.16	0.32	0.28	0.32		-
TCAF2	1	0.26	0.27	0.24	0.22	0.22	0.25	0.20	0.25	0.23	0.59	0.60	0.57	0.55	0.55	0.40	-	
TCAF2C	1	0.40	0.47	0.39	0.39	0.31	0.39	0.36	0.45	0.34	1.13	1.07	1.10	1.02	0.95	0.76	-	
TMEM86B	1	0.42	0.41	0.38	0.31	0.30	0.39	0.39	0.37	0.29	0.40	0.29	0.47	0.31	0.33	0.40	-	-
ATP4B	1	1.74	2.15	2.06	1.81	1.64	1.95	1.84	1.46	1.74	1.89	1.66	2.22	2.42	1.97	2.25	+	+
HGFAC	1	1.56	1.69	1.76	1.63	1.90	1.68	1.60	1.48	1.66	1.71	1.45	1.69	1.73	1.67	2.26	+	+
SLC9A3	1	1.49	1.62	1.56	1.48	1.81	1.62	1.47	1.49	1.62	1.47	1.36	1.76	1.78	1.56	1.73	+	+
SEMA3B	1	1.36	1.60	1.49	1.53	1.33	1.62	1.72	1.62	1.89	1.47	1.41	1.40	1.80	1.44	1.89	+	
NKX6−1	1	1.25	1.41	1.44	1.69	1.50	1.61	1.60	1.62	1.33	1.52	1.28	1.48	1.70	1.70	1.68	+	+
FGFRL1	1	1.53	1.63	1.51	1.64	1.36	1.75	1.78	1.54	1.68	1.65	1.40	1.72	1.71	1.59	2.07	+	+
DRD5	1	1.62	1.77	1.80	2.02	2.33	1.82	1.62	1.59	1.26	1.77	1.60	2.01	2.02	1.66	1.76	+	+
IDUA	1	1.58	1.74	1.76	1.79	1.68	1.79	1.75	1.57	1.81	1.73	1.57	1.91	1.81	1.73	2.09	+	+
ATP5ME	1	1.42	1.60	1.62	1.81	1.87	1.71	1.51	1.85	1.62	1.64	1.66	2.30	1.89	1.45	1.90	+	+
H4C5/16	2	2.27	2.23	1.89	2.24	1.62	2.49	1.89	2.30	2.18	2.57	2.05	3.78	2.38	2.63	2.67		+
IFRD2	1	1.32	1.51	1.23	1.71	0.98	1.52	1.37	1.20	1.73	1.38	1.23	1.81	1.54	1.50	1.70		+
IER3 (IEX-1)	1	1.50	1.65	1.49	0.85	0.49	1.75	1.39	1.31	1.12	1.56	1.38	3.23	1.87	1.23	1.82		+
HAND2	1	1.40	1.47	1.49	1.61	1.38	1.69	1.55	1.41	1.55	1.58	1.30	1.59	1.80	1.49	1.61		+
PLXND1	1	1.34	1.46	1.43	1.56	1.27	1.42	1.45	1.36	1.49	1.39	1.25	1.51	1.55	1.53	1.70		+

Table S6. **Estimated haploid copy number for genes related to digestion in ancient genomes and human population CNV trends (Complement to Table 1 and Table 2).** From left to right, the exact estimated gene copy number is reported for all genes and all genomes from *Den*, *Hn* and *Hs* populations. Genomes within a population are ordered from the oldest to the youngest. Blue shades correspond to the number of gene copies after mapping sequencing reads on the reference genome of modern humans GRGh38/hg38. Colors as in **Tables 1** and **2**.

		Africa													Eurasia																				
Genes	GRCh38/hg38	Gambian1	Gambian2	Mandenka1	Mandenka2	Mandenka3	Somali1	BantuKenya1	BantuKenya2	BantuKenya3	BantuTswana1	BantuTswana2	KhomaniSan1	KhomaniSan2	Icelandic1	Icelandic2	Saami1	Saami2	Mansi1	Mansi2	Tubalar1	Tubalar2	Altaian1	Mongola1	Mongola2	Oroqen1	Oroqen2	Oroqen3	Even1	Even2	Even3	Yakut1	Yakut2	EskimoChaplin1	EskimoNaukan1	EskimoNaukan2
AMY1B	3	1.86	1.89	3.79	2.84	3.21	3.05	3.69	2.79	4.12	2.82	2.74	6.14	1.93	1.84	7.51	2.02	2.36	3.02	2.94	3.48	4.85	4.42	4.27	4.24	2.07	3	3.39	3.07	2.98	4	2.63	2.77	2.06	1.93	2.96
SULT1A4	2	2.9	2.89	2.13	2.56	3.22	3.87	2.54	3.15	2.98	3.42	2.88	2.86	1.96	3.08	4.15	2.47	2.4	1.89	3.54	1.95	2.5	3.51	4.1	3.38	3.24	3.38	3.39	2.57	2.97	3.1	2.56	3.08	2.86	2.97	2.46
SLX1A	2	3.14	3.01	2.01	2.48	3.15	3.79	2.54	2.93	2.82	3.51	3.01	2.95	1.96	3.01	4.29	2.52	2.5	1.87	3.62	1.94	2.46	3.92	4.2	3.57	3.18	3.41	3.08	2.62	2.95	2.99	2.5	2.95	2.94	2.82	2.47
UGT2B28	1	0.01	0.48	0.48	0.47	0.48	0.47	0	0.5	0.97	0.01	0.5	0.52	0.01	0.98	0.99	0.49	0.51	0.53	1.03	0.5	0.96	0.94	0.51	0.96	0.98	0.44	0.94	1.01	0.98	0.49	0.98	0.98	1.04	0.96	1.02
LPA	1	1.58	1.5	1.62	1.77	1.44	2.01	1.76	1.81	1.37	1.63	1.54	1.32	1.58	1.59	2.18	1.47	1.26	1.56	1.54	1.67	1.13	1.71	2.74	2.2	1.64	1.77	1.75	1.69	1.47	1.6	1.81	1.78	1.5	1.39	1.79
ZNG1E	3	3.48	2.79	3.23	3.48	3.63	4.32	3.27	3.26	3.22	3.15	2.97	2.65	3.41	3.6	4.67	3.13	3.08	3.23	3.55	3.43	3.51	5.26	4.22	5.2	3.77	4.85	5.26	3.14	3.12	3.5	3.77	3.69	2.86	3.52	2.34
ACOT1	1	0.6	0.59	1.3	1.12	0.94	1.04	0.1	0.1	1.08	0.07	1.32	0.1	0.06	0.58	0.6	0.5	0.1	0.09	0.09	0.08	0.09	0.13	0.12	0.1	0.07	0.11	0.54	0.08	1	0.09	0.58	0.11	0.08	0.09	0.1
CLPS	1	2.4	1.49	2.24	1.07	2.23	2.4	1.69	2.6	1.5	1.77	1.39	1.4	0.96	1.1	1.47	1.87	1.94	1.91	0.89	1.48	0.91	1.52	1.47	1.52	1.06	1.69	1.5	1.97	1.96	1.48	2.18	1.5	1.4	1.02	0.92
CLPSL1	1	1.79	1.23	1.49	0.98	1.57	1.7	1.25	1.82	1.19	1.46	1.15	1.25	0.95	1	1.26	1.44	1.47	1.41	0.9	1.16	0.92	1.18	1.22	1.25	1.07	1.47	1.36	1.42	1.44	1.19	1.58	1.32	1.08	0.9	0.89
RFPL4A	2	2.33	2.33	2.29	2.81	2.17	2.42	2.73	2.32	2.26	2.35	2.25	2.33	3.03	2.24	2.41	3.49	2.36	2.3	3.01	2.85	2.21	2.35	4.2	3.62	2.12	2.74	2.25	3.57	3.42	3.36	2.8	3.53	2.89	3.34	2.9
F7	1	0.92	0.93	0.96	0.91	0.98	0.89	0.96	0.96	0.94	0.91	0.93	0.88	0.94	0.98	0.94	0.89	0.89	0.9	0.83	0.92	0.91	0.91	0.88	0.95	0.88	0.98	0.95	0.93	0.91	0.89	0.97	0.95	0.9	1.01	0.91
PDX1	1	0.99	1.02	1	1.1	1.07	0.92	0.96	1.08	0.94	1.01	0.94	0.91	0.97	1	0.97	0.93	0.97	0.9	0.98	1	1.02	0.93	0.99	1.01	1.06	1.05	1.07	1.04	0.97	0.94	1.03	1.1	0.97	0.95	1.01
WFS1	1	0.96	0.94	1.04	1.05	1.03	0.94	1.03	1.08	1.01	0.9	0.95	0.93	0.94	0.97	0.95	0.93	0.91	0.88	0.98	0.92	0.93	0.94	0.95	0.95	1.01	1.05	1.01	0.95	0.98	0.99	1.01	1.04	0.99	0.99	0.93
H2AC11	1	0.79	1.37	0.98	1.09	1.12	1.02	0.91	1	0.97	1.03	0.83	1.11	0.9	0.94	1.02	0.99	0.95	0.85	0.87	0.95	1.16	1.1	0.83	0.89	0.88	0.94	0.97	0.98	0.95	0.84	0.93	0.76	1.04	0.96	0.9
ACOX3	1	0.96	0.96	1.02	1.01	1.02	0.94	1.01	1.01	1	0.95	0.94	0.92	0.94	0.96	0.92	0.93	0.93	0.92	0.94	0.93	0.94	0.97	0.96	0.96	1.01	1.07	1.03	0.95	0.98	0.92	1.02	1.03	0.96	1	0.96
HAND1	1	1.06	1.01	1.14	0.99	1	0.91	1	1.08	1.01	1	0.96	0.94	0.94	0.95	0.99	0.96	0.92	1	0.98	0.94	0.95	0.93	1	0.94	0.97	1.13	1.06	0.93	0.95	1.01	1.03	1.04	1	1	0.94
MST1	1	1.08	1.01	1.04	1.12	1.1	1.08	1.11	0.97	1.04	1.05	0.96	1.03	1.06	1.02	1.05	0.97	1.06	0.95	1.11	1.01	1.07	1.15	1.03	1.1	1.11	1.24	1.12	1.11	1.06	1.06	1.11	1.13	1.05	1.06	0.99
H1-2	1	1.03	0.98	1.05	1.14	1.18	0.87	1.17	0.94	1.39	1.03	0.86	1.03	0.98	1.02	1	0.98	0.97	0.96	0.98	1	0.98	1.06	0.62	0.95	0.92	1.05	0.81	0.92	0.96	0.82	0.98	1.13	0.95	1.03	1.2
CD24	1	0.96	0.95	1.07	1.05	1.06	0.83	1.02	1.09	0.97	0.95	0.99	0.97	0.93	1.04	0.94	0.97	0.97	0.91	0.92	0.92	1	0.92	0.98	0.95	1.07	1.14	1.06	0.87	0.89	0.86	1.07	0.98	0.93	0.88	0.93
BOLA2B	2	2.88	3.08	2.14	2.68	3.27	4.06	2.6	2.98	3.72	3.35	2.95	2.87	1.93	3.02	4.34	2.37	2.65	2.06	3.53	1.85	2.5	4.06	4.43	3.86	3.67	3.74	3.59	2.73	2.85	2.98	2.62	2.93	2.94	2.89	2.65
CLPTM1L	1	1.06	0.97	1.08	1.06	1.09	1.07	1.09	1.11	1.07	1.06	1.04	0.97	1.04	1.09	1.06	1	0.97	1.11	1.01	1.06	1.11	1.15	1.1	1.14	1.15	1.27	1.19	1.19	1.11	1.12	1.14	1.18	1.11	1.16	1.11
ORM1	1	0.98	1.08	1.15	1.02	1.52	0.88	1.02	1.56	1.19	1.01	1	0.96	0.92	0.9	1.03	0.97	0.96	0.95	1.07	0.91	1.07	1.01	1.03	0.99	1.48	1.05	0.93	0.94	0.97	0.9	1.01	1.11	1.46	1.04	1
ORM2	1	0.89	0.92	1.01	1.09	1.07	0.58	1.07	0.98	1.01	0.98	0.93	1.06	1	1.02	1.06	1.02	1.02	0.96	1.49	0.99	0.99	0.96	1.02	1.03	1.13	0.96	1.1	0.98	0.92	1.05	0.99	1.01	1.12	1.04	1.03
TPSAB1	2	1.81	2.02	2.01	1.93	1.36	2.02	2.07	1.97	2.02	1.9	1.54	1.98	1.74	1.75	2.06	1.71	1.84	1.8	1.76	1.55	2	2.17	2.03	2.13	1.97	1.94	1.91	1.49	1.96	1.84	1.89	1.6	1.93	1.83	1.72
IFNL2	2	2.07	1.98	1.95	2.11	2.17	2.07	2	2.23	1.97	2.03	1.89	2.03	2.1	2.05	2.15	2.05	2.04	2.02	2.05	2.07	1.99	2.02	2.16	2.04	1.97	2.07	2.15	2.09	2.09	1.96	1.82	2.12	2.03	1.93	2.03
C4B	2	1.66	1.66	1.97	2.02	1.55	1.97	1.55	2.52	1.79	1.56	1.57	1.42	1.62	1.85	3.24	1.08	1.93	2.23	1.86	1.77	1.63	2.45	2.57	2.22	2.33	2.33	2	1.7	1.84	1.37	2.02	2.26	1.85	1.91	1.73
GRTP1	1	0.92	0.94	0.98	0.98	1.05	0.91	0.98	1	0.94	0.9	0.89	0.9	0.93	0.99	0.93	0.91	0.92	0.87	0.91	0.94	0.93	0.91	0.94	0.94	1.01	1.11	1.04	0.93	0.96	0.93	1.01	1.02	0.91	0.94	0.95
UQCRFS1	1	1.08	1.12	1.11	1.15	1.04	1.08	1.13	1.07	1.19	1.18	1.15	1.14	1.14	1.08	1.14	1.15	1.12	1.17	1.18	1.15	1.15	1.15	1.13	1.12	1.08	1.01	1.05	1.08	1.09	1.16	1.15	1.06	1.24	1.18	1.16
ZNF431	1	1.06	1.1	1.03	1.05	1.01	1.05	1.03	1.02	1.09	1.07	1.07	1.11	1.11	1.08	1.06	1.12	1.11	1.13	1.17	1.11	1.1	1.07	1.06	1.06	1.03	1.02	1.01	1.12	1.12	1.17	1.09	1.06	1.16	1.12	1.13
N4BP3	1	0.91	0.92	0.97	1.02	1.01	0.92	0.97	0.9	0.95	0.95	0.95	0.93	0.93	0.97	0.97	0.91	0.93	0.92	0.91	0.91	0.92	0.95	0.94	0.94	1.08	1.1	1.07	0.95	0.91	0.95	0.97	0.99	0.92	0.97	0.9
STAB1	1	0.97	0.95	1.02	1.06	1.07	0.92	1.02	1.09	1	0.93	0.95	0.93	0.95	0.98	0.96	0.91	0.92	0.94	1.01	0.96	0.94	0.93	0.99	0.96	1.08	1.14	1.09	1.01	0.98	1.01	1.07	1.11	1	1.03	0.96
LYPD8	1	0.95	0.87	0.88	0.94	0.95	0.86	1.03	0.99	0.92	0.9	0.97	0.95	0.88	0.94	0.9	0.91	1.4	0.98	0.93	0.96	0.9	0.91	0.89	0.9	0.93	0.85	0.94	0.87	1.43	0.89	0.94	0.91	0.96	0.95	0.94
PECAM1	1	0.98	1	0.99	1	1	0.95	0.99	1.03	0.98	1	0.99	0.98	1	1.01	0.96	0.99	0.97	0.95	0.95	0.99	0.97	0.95	0.97	0.95	1	1.05	1.01	0.91	0.94	0.95	1.02	1	0.94	0.98	0.98
TCAF2	1	0.52	0.6	0.52	0.54	0.36	0.61	0.45	0.43	0.38	0.33	0.52	0.46	0.44	0.36	0.51	0.36	0.47	0.34	0.34	0.28	0.37	0.52	0.53	0.39	0.4	0.3	0.44	0.38	0.36	0.36	0.39	0.34	0.28	0.28	0.28
TCAF2C	1	0.95	1.09	0.96	0.8	0.6	1.06	0.7	0.65	0.57	0.57	0.78	0.77	0.6	0.51	0.83	0.53	0.72	0.56	0.6	0.39	0.43	0.9	0.83	0.61	0.38	0.29	0.63	0.54	0.55	0.55	0.37	0.52	0.4	0.39	0.4
TMEM86B	1	1.03	0.96	0.96	0.98	0.9	0.99	0.93	1.01	1.07	0.87	0.95	0.98	1	1	0.87	0.93	0.96	1.06	1	0.93	0.98	0.9	1.04	0.9	1.04	0.97	1.09	1.01	1.02	0.99	1.03	0.94	0.92	1.03	0.92
ATP4B	1	0.99	0.94	1.05	1.03	1.08	0.98	1.02	1.03	1.05	0.93	0.97	0.89	1.01	0.99	0.97	0.96	0.95	0.99	1.01	0.98	0.99	1.04	0.92	0.97	1.07	1.11	1.08	1.01	1	1.01	1.11	1.07	1	0.99	0.94
HGFAC	1	1.02	0.96	0.94	1.05	0.97	0.89	1.05	1	0.93	0.98	0.96	0.96	0.96	0.99	0.92	0.92	0.9	0.94	0.97	1	0.98	0.93	0.94	0.97	1	1.04	1.06	0.98	1.05	0.92	0.96	0.96	1.01	1.05	0.93
SLC9A3	1	0.91	0.88	0.98	0.97	1.01	0.88	0.97	1	0.98	0.88	0.92	0.88	0.9	0.93	0.93	0.86	0.88	0.9	0.91	0.91	0.93	0.94	0.9	0.94	1.01	1.04	1.03	0.95	0.94	0.92	0.98	1.02	0.93	0.97	0.9
SEMA3B	1	0.91	0.93	1	1.07	1.02	0.92	0.98	1.02	0.98	0.92	0.89	0.9	0.91	0.93	0.92	0.89	0.9	0.92	1	0.87	0.94	0.91	0.94	0.93	1.09	1.18	1.03	0.96	0.92	0.99	1.03	1.1	0.97	1.01	0.91
NKX6-1	1	0.92	0.96	0.94	0.95	1.04	0.91	0.97	1.02	1	0.93	0.93	0.93	0.98	0.97	0.85	0.94	0.92	0.91	0.96	0.96	0.95	0.93	0.87	0.95	0.95	0.9	1	0.93	0.98	0.98	0.89	0.89	0.98	0.94	0.95
FGFRL1	1	0.87	0.88	0.98	0.94	1.02	0.74	0.96	0.92	0.94	0.88	0.89	0.87	0.84	0.92	0.89	0.82	0.87	0.87	0.91	0.88	0.91	0.86	0.87	0.93	0.98	1.08	1.04	0.91	0.92	0.9	0.97	0.95	0.94	0.91	0.86
DRD5	1	0.96	0.9	0.91	0.89	1.06	0.81	0.97	1.14	0.98	0.96	0.92	0.96	1.03	0.87	0.84	0.95	0.92	0.91	0.92	1	0.9	0.94	0.94	0.94	1.04	1	0.96	0.9	0.94	0.98	0.97	0.92	0.99	0.92	0.98
IDUA	1	0.93	0.93	1.03	1.02	1.07	0.79	1.02	0.97	0.96	0.88	0.9	0.87	0.91	0.99	0.93	0.89	0.86	0.88	0.91	0.93	0.94	0.93	0.92	0.95	1.06	1.13	1.02	0.94	0.94	0.92	0.97	0.99	0.95	0.96	0.91
ATP5ME	1	0.82	0.91	1.04	1.05	1.19	0.84	1.02	0.85	1.04	0.8	0.95	0.76	0.84	0.86	0.79	0.81	0.79	0.87	0.82	0.89	0.96	0.97	0.82	0.94	1.01	1.07	1.03	0.89	0.92	0.83	0.95	1.07	0.96	0.94	0.87
H4C16	2	1.79	1.98	2.17	2.28	2.02	1.72	2.11	1.79	2.34	2.02	1.98	1.8	1.98	1.95	1.96	1.82	1.93	1.92	2.07	2.3	2.06	2.13	2.08	1.85	2.16	2.17	1.9	1.77	1.84	1.97	1.9	1.77	1.9	1.58	1.97
IFRD2	1	0.99	1.05	1.03	1.09	1.05	0.96	0.95	1.06	1.04	0.89	0.96	0.94	0.99	0.99	0.95	0.96	0.92	0.94	0.97	0.97	0.91	0.99	0.94	0.99	1.07	1.14	1.03	0.99	0.94	0.97	1.09	1.03	1.02	1.05	0.97
IER3	1	0.94	1.02	1.08	1.16	1.12	0.95	1.09	1.25	1.12	0.97	0.95	1.03	0.93	0.96	0.99	0.86	0.85	0.94	0.83	0.91	0.87	1.09	1.01	0.9	1.12	1.16	1.11	0.99	0.91	0.84	1.09	1.15	0.96	1.02	0.93
HAND2	1	0.92	0.95	0.94	0.97	0.93	0.9	0.9	0.96	1	0.87	0.91	0.95	0.94	0.89	0.96	0.91	0.9	0.9	0.9	0.97	0.93	0.94	0.88	0.92	0.9	0.96	1.04	0.93	0.96	0.93	0.97	0.96	0.98	0.89	0.93
PLXND1	1	0.96	0.94	1.04	1.05	1.04	0.94	1.05	1.04	1.03	0.94	0.93	0.92	0.97	0.96	0.96	0.93	0.94	0.93	0.97	0.94	0.95	0.93	0.97	0.98	1.04	1.07	1.06	0.94	0.97	0.95	1.03	1.06	0.98	1.02	0.95

Table S7. **Estimated haploid copy number for genes related to digestion in modern genomes (Complement to Table 1 and Table 2).** From left to right, the exact estimated gene copy number is reported for all genes and all genomes from ethnic groups in Africa and Eurasia. Genomes within a population are ordered by ethnic group. Blue shades correspond to the number of gene copies after mapping sequencing reads on the reference genome of modern humans GRGh38/hg38. Colors as in **Tables 1** and **2**.

		Middle East								Oceania												South Asia								
Genes	GRCh38/hg38	Palestinian1	Palestinian2	Palestinian3	Iranian1	Iranian2	Balochi1	Balochi2	Balochi3	Dusun1	Dusun2	Igorot1	Igorot2	PapuanHighlands1	PapuanHighlands2	PapuanHighlands3	PapuanSepik1	PapuanSepik2	PapuanSepik3	Australian1	Australian2	Xibo1	Xibo2	Xibo3	Lahu1	Lahu2	Lahu3	Cambodian1	Cambodian2	Cambodian3
AMY1B	3	2.31	4.95	6.19	2.95	6.78	0.92	0.9	4.58	3.85	2.92	2.92	4.85	2.73	2.73	5.48	3.59	4.34	2.76	3.75	1.9	4.44	4.22	2.78	4.08	4.04	3.99	1.7	3.58	3.4
SULT1A4	2	2.89	2.7	3.38	2.88	5.24	2.99	3.06	2.36	2.88	2.35	3.43	2.45	2.47	2.94	4.65	2.49	3.43	2.93	3	3.03	2.99	3.67	2.46	3.35	3.29	2.18	2.67	2.52	2.63
SLX1A	2	2.99	2.7	3.38	2.98	5.16	2.93	3.03	2.38	3.01	2.37	3.51	2.49	2.5	3.01	4.76	2.48	3.15	2.96	2.83	3.07	2.91	3.47	2.62	3.31	3.25	2.13	2.54	2.55	2.45
UGT2B28	1	1	0.93	0.96	0.01	0.51	0.48	0.98	0.51	0.97	0.98	0.98	0.93	1	0.99	0.52	0.98	0.95	1.01	0.5	1	0.91	0.92	0.44	0.89	0.98	0.95	0.45	0.95	0.98
LPA	1	1.62	1.54	1.94	1.49	2.09	1.49	1.55	1.6	1.68	1.44	1.52	1.83	1.58	1.78	1.99	1.81	2.33	1.73	1.77	1.6	1.8	1.75	1.79	1.83	1.91	1.58	1.97	1.4	1.82
ZNG1E	3	3.09	4.58	4.58	3.05	4.16	3.17	3.08	3.67	3.83	3.13	3.39	3.09	3.91	3.95	5.12	3.35	4.72	3.55	3.06	3.05	3.1	3.6	3.64	3.73	3.38	3.4	3.3	3.18	3.79
ACOT1	1	0.59	0.13	0.08	0.52	1.75	1.03	0.52	0.09	0.09	0.09	0.1	0.08	0.56	1.04	1.08	0.99	1.05	1.06	0.99	1.01	0.11	0.12	1.02	0.06	0.09	0.09	0.62	0.56	0.56
CLPS	1	1.49	0.94	2.52	1.39	1	1.11	1.72	1.49	1.51	0.89	0.91	0.91	1.59	1.42	1.93	1.55	1.4	2.39	2.21	1.53	1.69	1.15	1.23	1.42	1.35	1.72	1.07	1.63	1.82
CLPSL1	1	1.24	0.94	1.77	1.13	0.9	0.96	1.31	1.15	1.2	0.95	0.94	0.95	1.24	1.26	1.46	1.26	1.15	1.75	1.47	1.23	1.3	0.99	0.96	1.36	1.32	1.39	1.03	1.27	1.3
RFPL4A	2	2.93	2.32	2.29	2.25	3.51	2.94	3.46	3.46	3.34	3.96	2.89	3.48	3.51	2.87	3.59	3.47	3.66	3.46	2.94	2.85	3.92	4.3	4.03	2.29	2.78	3.9	3.77	3.95	3.95
F7	1	0.89	0.92	0.97	0.85	0.93	0.91	0.97	0.84	0.95	0.97	0.86	0.89	0.94	0.9	0.87	0.91	0.88	0.95	1.38	1.22	0.96	0.92	0.89	0.95	0.91	0.96	1	0.95	0.96
PDX1	1	1	0.97	0.97	0.96	0.98	1.02	1.08	0.97	0.93	0.94	0.94	0.98	0.97	0.96	0.96	1	0.85	1	0.86	0.85	1.04	1.04	0.98	1.06	1.04	0.93	1.04	1.02	1.06
WFS1	1	0.94	0.94	0.95	0.94	0.93	1.03	1.04	1.05	0.96	0.97	0.96	0.96	0.93	0.96	0.95	1.06	0.9	0.93	1.03	1.08	1.07	1.06	1.02	1.01	0.97	1.01	1.04	1.06	1.03
H2AC11	1	0.95	0.87	0.91	1.03	1.02	0.84	1.06	0.99	0.95	0.92	0.87	1.12	0.95	1	0.95	1.02	1.04	0.99	1.21	1.18	1.23	1.13	0.95	1.06	1.13	0.98	0.98	1.2	0.91
ACOX3	1	0.96	0.96	0.96	0.94	0.95	1.01	1.05	0.97	0.96	0.95	0.94	0.96	0.94	0.97	0.94	1.02	0.93	0.94	0.97	1.03	1.07	1.06	1.02	1	0.99	1.01	1.03	1.04	1.04
HAND1	1	0.92	1	0.91	0.98	1.04	1.08	1.09	1.02	0.98	0.95	0.91	0.96	0.94	0.98	0.91	0.98	0.96	0.98	0.79	0.78	1.04	1.07	1.02	0.96	1.07	1.06	1.09	1.08	1.08
MST1	1	1.08	1.1	1.08	0.97	1.07	1.07	1.12	0.93	1	1.03	1.08	1.01	1.06	1.05	0.99	1.1	0.96	1.05	1.15	1.19	1.14	1.12	1.05	1.12	1.13	1.11	1.15	1.09	1.04
H1-2	1	1.16	0.83	1.08	1.02	0.86	0.94	0.98	0.92	1.03	1.08	1.01	1.04	0.97	0.86	1.01	0.95	1	0.99	0.82	1.04	1.14	1.08	1.11	1	1.03	0.99	0.97	1.09	0.85
CD24	1	0.97	0.95	0.96	0.97	0.92	0.98	1.1	0.97	0.98	0.86	0.97	0.9	0.98	0.91	0.95	1.13	0.99	0.93	0.84	0.83	1.08	1.05	0.97	0.99	0.98	1	1.06	1.02	1.09
BOLA2B	2	2.96	2.83	3.62	2.87	5.54	3.21	3.29	2.69	2.77	2.01	3.44	2.34	2.6	2.95	4.94	2.47	3.38	2.82	2.96	3.55	3.08	3.71	2.61	3.48	3.13	2.15	2.73	2.71	2.73
CLPTM1L	1	1.12	1.12	1.08	1.07	1.11	1.08	1.15	1.1	1.11	1.08	1.11	1.11	1.13	1.13	1.09	1.1	1.18	1.13	1.33	1.35	1.12	1.17	1.08	1.11	1.12	1.16	1.16	1.17	1.11
ORM1	1	1.87	1.05	1.04	0.98	1.07	1.09	1.07	1.06	0.93	1.1	1.12	0.98	0.96	1.09	1.48	1.01	1	1	1.06	1.01	1.02	1.08	1.05	0.97	0.98	1.1	1.08	1.02	1.06
ORM2	1	1.24	1	1.02	1.01	0.99	1.09	1.06	1.06	0.95	1.49	0.96	0.98	0.99	1.4	1.07	1	0.93	0.99	1.02	1.03	1.11	1.05	0.61	0.98	1.51	1.46	1.1	1.05	1.08
TPSAB1	2	1.86	2.27	1.92	2.02	1.9	1.89	1.96	1.77	1.98	1.93	1.9	1.94	1.76	1.96	2.24	2.04	2.1	1.83	2.01	1.79	1.78	1.98	1.74	1.94	1.96	1.92	1.91	2.02	1.88
IFNL2	2	2.01	2.06	1.92	2	2.08	2.01	2.26	2.21	1.95	1.96	2.08	2.14	2.02	2.07	2.06	2.12	2.03	2.1	2.34	2.15	2.14	2.29	2.28	2.15	2.05	2.02	2.12	2.08	2.3
C4B	2	2.01	2.48	2.56	1.92	2.67	1.64	1.71	1.75	2.57	2.16	2.44	1.7	1.8	1.7	3.28	1.93	2.97	2.34	1.88	2.42	2.77	2.06	1.93	1.76	2.22	2.41	2.83	2.08	2.9
GRTP1	1	0.92	0.94	0.93	0.92	0.9	0.96	1.04	0.93	0.95	0.93	0.9	0.91	0.94	0.94	0.91	1.01	0.87	0.93	0.94	0.96	1.01	1.04	0.94	0.99	1	1.03	1.03	1.01	1.04
UQCRFS1	1	1.15	1.08	1.08	1.1	1.15	1.25	1.01	1.19	1.12	1.16	1.09	1.15	1.07	1.05	1.14	1.14	1.1	1.09	1.11	1.12	1.11	1.09	1.08	1.09	1.11	1.03	1.08	1.16	1.13
ZNF431	1	1.08	1.05	1.05	1.11	1.11	1.05	1.01	1.05	1.11	1.07	1.11	1.1	1.06	1.05	1.03	1.06	1.05	1.08	1.03	1.01	1.02	1.05	1.04	1.03	1.04	0.99	1.02	1.02	1.04
N4BP3	1	0.93	0.98	0.92	0.93	0.87	0.98	1.03	0.94	0.96	0.93	0.91	0.94	0.94	0.96	0.91	0.97	0.87	0.93	0.96	1.03	0.98	1.01	0.9	1.03	0.98	1.07	0.98	1.01	1.06
STAB1	1	0.94	0.96	0.98	0.96	0.95	1.02	1.09	1.02	0.93	0.96	0.96	0.96	0.98	0.97	0.94	1.06	0.91	0.95	1.08	1.08	1.07	1.03	1.04	1.08	1.02	1.08	1.06	1.06	1.06
LYPD8	1	0.95	0.93	0.93	0.95	0.96	0.92	0.95	1.13	0.94	0.95	0.93	0.94	0.95	0.97	0.92	0.9	0.99	0.96	1.07	1.05	0.95	0.96	0.94	0.92	0.88	0.95	0.97	0.94	0.92
PECAM1	1	0.98	0.96	0.94	0.98	0.92	0.98	1.01	1.01	0.98	0.97	0.98	0.97	0.99	1.01	0.96	0.97	0.94	0.99	1	0.99	1.03	1.02	1.01	1.01	1	0.99	1.04	1.01	1.03
TCAF2	1	0.35	0.81	0.65	0.28	0.67	0.44	0.33	0.4	0.53	0.39	0.35	0.45	0.55	0.33	0.65	0.53	0.54	0.47	0.36	0.4	0.36	0.36	0.29	0.38	0.46	0.37	0.36	0.35	0.3
TCAF2C	1	0.58	1.41	1.11	0.39	1.22	0.73	0.52	0.62	0.9	0.51	0.59	0.77	0.99	0.58	1.13	0.91	0.89	0.77	0.55	0.73	0.38	0.5	0.26	0.52	0.73	0.58	0.45	0.52	0.32
TMEM86B	1	1	1.05	0.88	1.02	0.99	0.86	1.04	1.01	1.1	0.88	0.9	0.87	0.95	0.88	1.01	1.05	0.95	0.93	1.12	1.05	0.99	0.91	1.03	0.89	0.98	0.96	1.07	1.01	1.04
ATP4B	1	0.9	0.97	0.99	0.94	0.96	1.06	1	1.06	0.94	0.93	0.94	0.97	0.99	0.98	0.88	1.1	0.89	0.99	1.05	0.99	1.13	1.1	1.01	1.09	0.98	1.06	1.16	1.1	1.07
HGFAC	1	0.98	0.9	1	0.93	0.9	0.93	0.99	0.97	0.99	0.96	0.98	0.92	1.01	1.01	0.97	1.04	0.95	0.95	1.22	1.1	0.96	0.99	0.94	1.02	1.02	0.96	0.98	1.06	0.94
SLC9A3	1	0.91	0.92	0.95	0.9	0.92	0.98	1.01	0.96	0.92	0.94	0.91	0.92	0.9	0.92	0.9	0.97	0.89	0.9	1.02	0.98	1.02	1	0.91	0.96	0.98	1.05	0.99	1	1
SEMA3B	1	0.92	0.92	0.95	0.87	0.89	0.92	1.03	0.99	0.93	0.9	0.94	0.91	0.96	0.96	0.92	0.94	0.84	0.92	0.99	1.02	0.97	1.01	0.93	1.07	0.99	1.07	1	0.98	0.98
NKX6-1	1	0.96	0.86	0.95	0.94	1	0.89	0.91	0.94	0.95	0.94	0.94	0.89	0.96	0.95	0.89	0.96	0.87	0.95	0.69	0.81	0.94	0.89	0.95	0.98	1.01	0.96	0.94	0.95	0.88
FGFRL1	1	0.9	0.89	0.91	0.85	0.84	0.93	0.99	0.93	0.89	0.89	0.88	0.91	0.88	0.92	0.88	1	0.84	0.85	0.93	1	0.97	0.96	0.94	1	1.03	1.08	1.01	0.96	0.98
DRD5	1	1.01	0.96	0.97	0.91	0.93	1.01	1.1	0.93	0.97	0.99	0.99	1.05	0.89	0.92	0.96	1.04	0.87	0.92	0.89	1.07	0.97	1.07	0.97	1.03	1.02	1.02	1.05	1.07	1.07
IDUA	1	0.94	0.89	0.92	0.89	0.91	0.9	1.01	0.95	0.96	0.91	0.93	0.93	0.91	0.93	0.9	1.01	0.87	0.89	0.91	1.04	1.01	1.05	0.96	1.02	1.01	1.04	0.97	0.99	1.02
ATP5ME	1	0.93	0.89	0.87	0.84	0.91	0.99	1.21	0.98	0.92	0.84	0.92	0.89	0.88	0.95	0.84	1.17	0.75	0.84	0.72	0.82	1.13	1.1	1.01	1	1.11	0.96	1.12	1.04	0.99
H4C16	2	1.93	1.98	1.89	1.86	2.14	1.88	2.01	2.06	1.86	1.77	1.85	2	1.92	1.84	1.76	1.94	1.81	1.77	2	1.99	1.78	2	1.97	1.78	1.89	2.04	1.92	2.26	1.91
IFRD2	1	0.94	0.97	0.98	0.94	0.92	0.98	1.04	0.97	1.01	0.97	0.97	0.91	0.99	0.97	0.96	1.04	0.85	0.9	1.03	1.13	1.08	1.06	0.96	1.03	1.01	1.18	1.01	1.17	1.07
IER3	1	1.01	0.97	1.07	1.1	0.93	0.88	0.98	1.04	0.98	0.91	0.89	1.05	1.05	1.01	0.91	0.93	0.95	1.06	0.98	1.05	1.11	1.16	1.01	1	1.03	1.2	1.03	1.14	1.02
HAND2	1	0.99	1	0.81	0.92	0.95	0.92	0.93	1	0.9	0.86	0.9	0.9	0.92	0.94	0.88	0.94	0.82	0.91	0.67	0.7	1.06	1	0.98	0.97	1.02	0.98	0.99	1.03	1.02
PLXND1	1	0.96	0.95	0.99	0.95	0.96	1.05	1.07	1.02	0.97	0.96	0.94	0.96	0.96	0.98	0.94	1.03	0.91	0.96	1.02	1.04	1.06	1.04	1.04	1.02	1.02	1.03	1.04	1.04	1.03

Table S7 (continued). The table presents CNV estimates for individuals from the Middle East, Oceania and South Asia.

Supplementary Figures

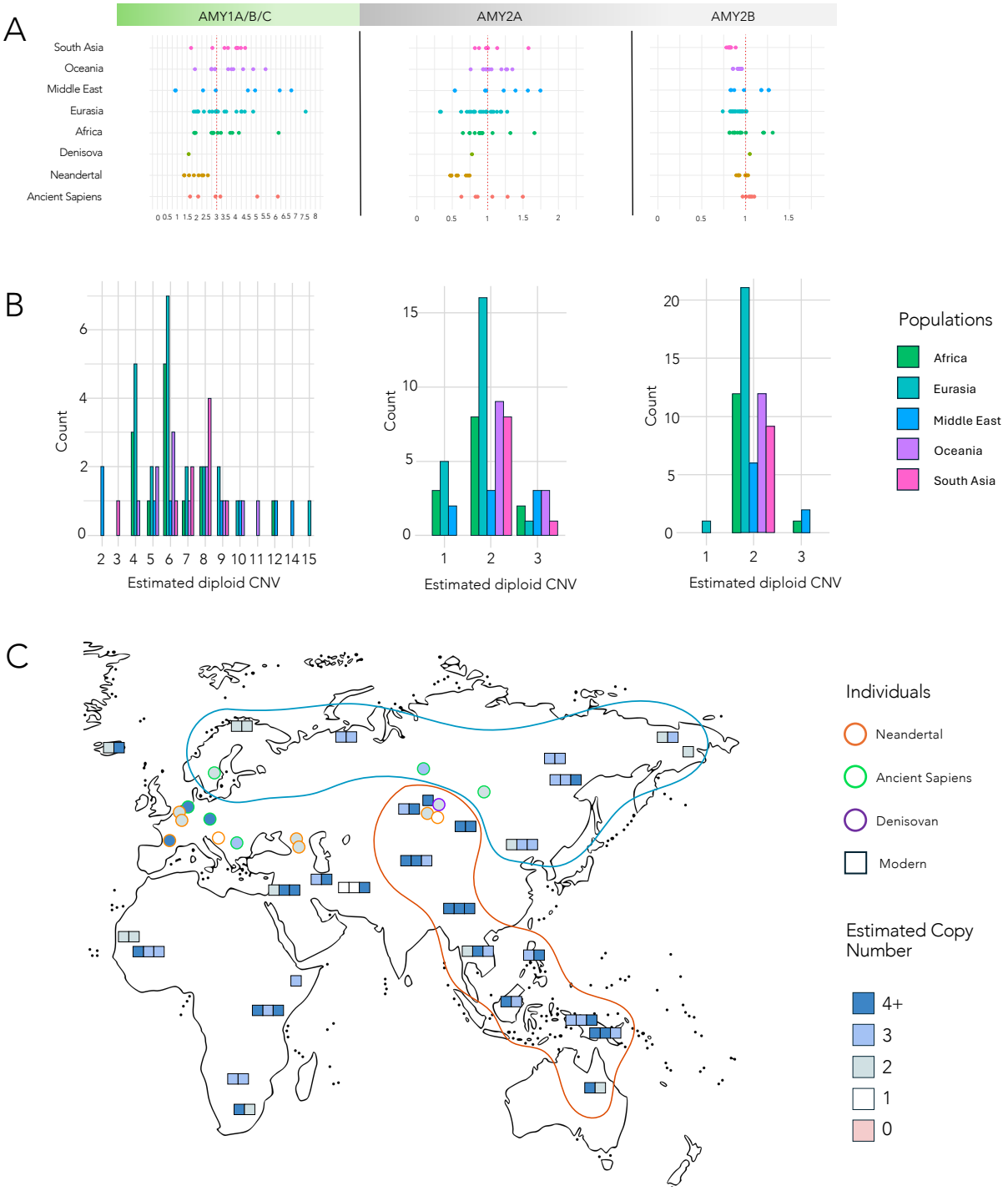


Figure S1. CNV estimations of the *AMY1A/B/C* gene cluster, and the *AMY2A* and *AMY2B* genes in ancient and modern populations. **A.** Haploid estimation of CNV for the *AMY1A/B/C* gene cluster, along with the *AMY2A* and *AMY2B* genes. Note that the CNV for the *AMY2A* and *AMY2B* genes is estimated to be 1. (Complement to **Figure 3**.) **B.** Distribution of diploid copy number for the 64 modern individuals analyzed in this study. Diploid CNV was calculated using 2x read estimates for each gene rounded to the nearest integer. As expected, pair diploid CNVs show higher numbers of individuals, as remarked in (Usher *et al.* 2015) **C.** Visualisation of *AMY1A/B/C* haploid CNVs (colors as shown in **Tables 1** and **2**) for both modern (square) and ancient (circle) individuals. Different populations are identified by colored borders: *Hn* in orange, *Hs* in green, *Den* in purple, and modern individuals in black. Most modern genomes from the Eurasian belt region (blue shape) exhibit low CNVs, similar to ancient individuals, while modern genomes from the East Asia, South Asia, and Oceania corridor (red shape) show higher CNVs. The CNV for each individual within the ethnic groups considered in this study is represented by a colored square (compare to **Figure 1B**).

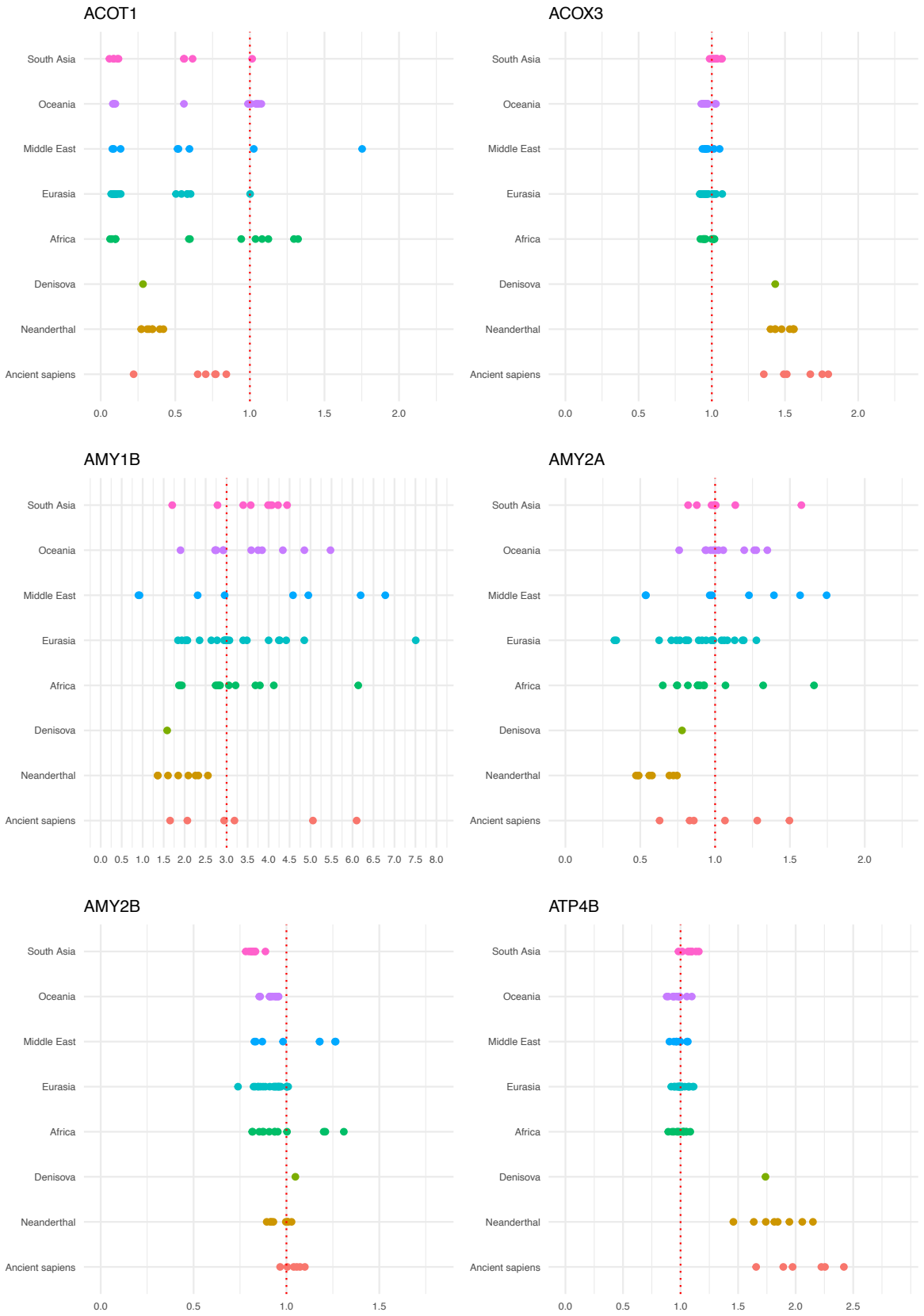


Figure S2. Comparison of CNVs across all modern and ancient individuals in this study, grouped by population. All referenced genes (including diet-related, accessory, and others) are listed in alphabetical order.

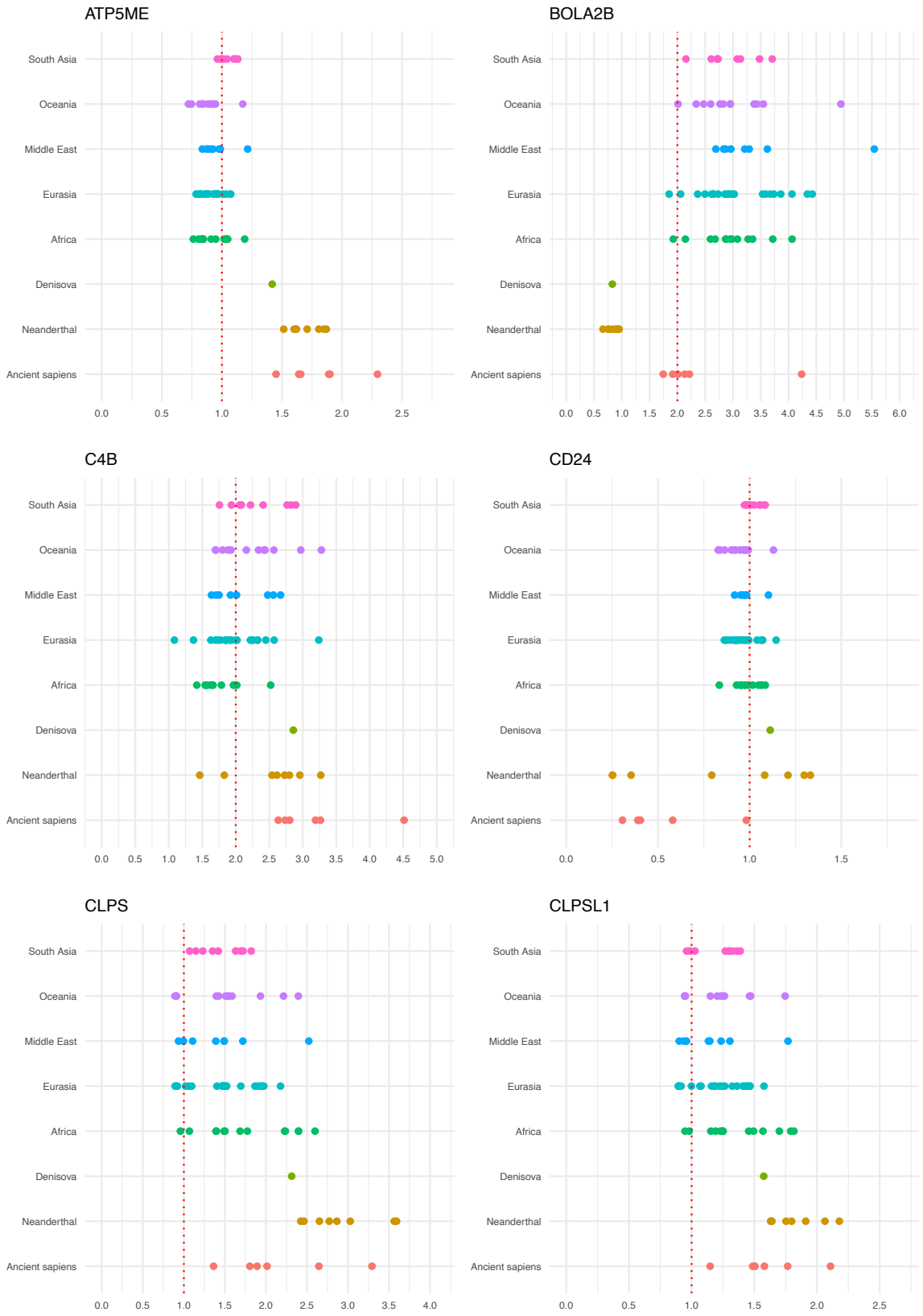


Figure S2 (continued).

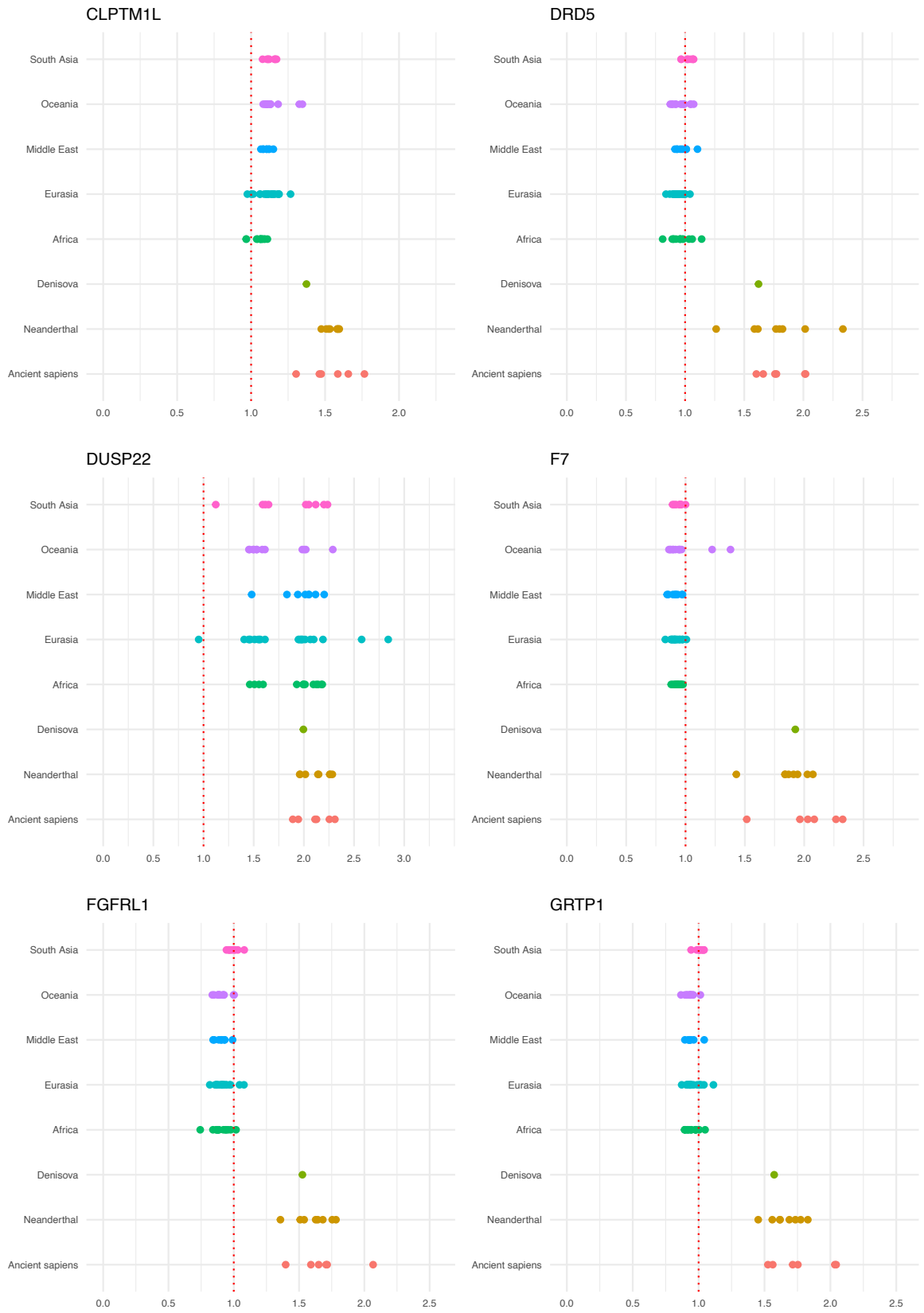


Figure S2 (continued).

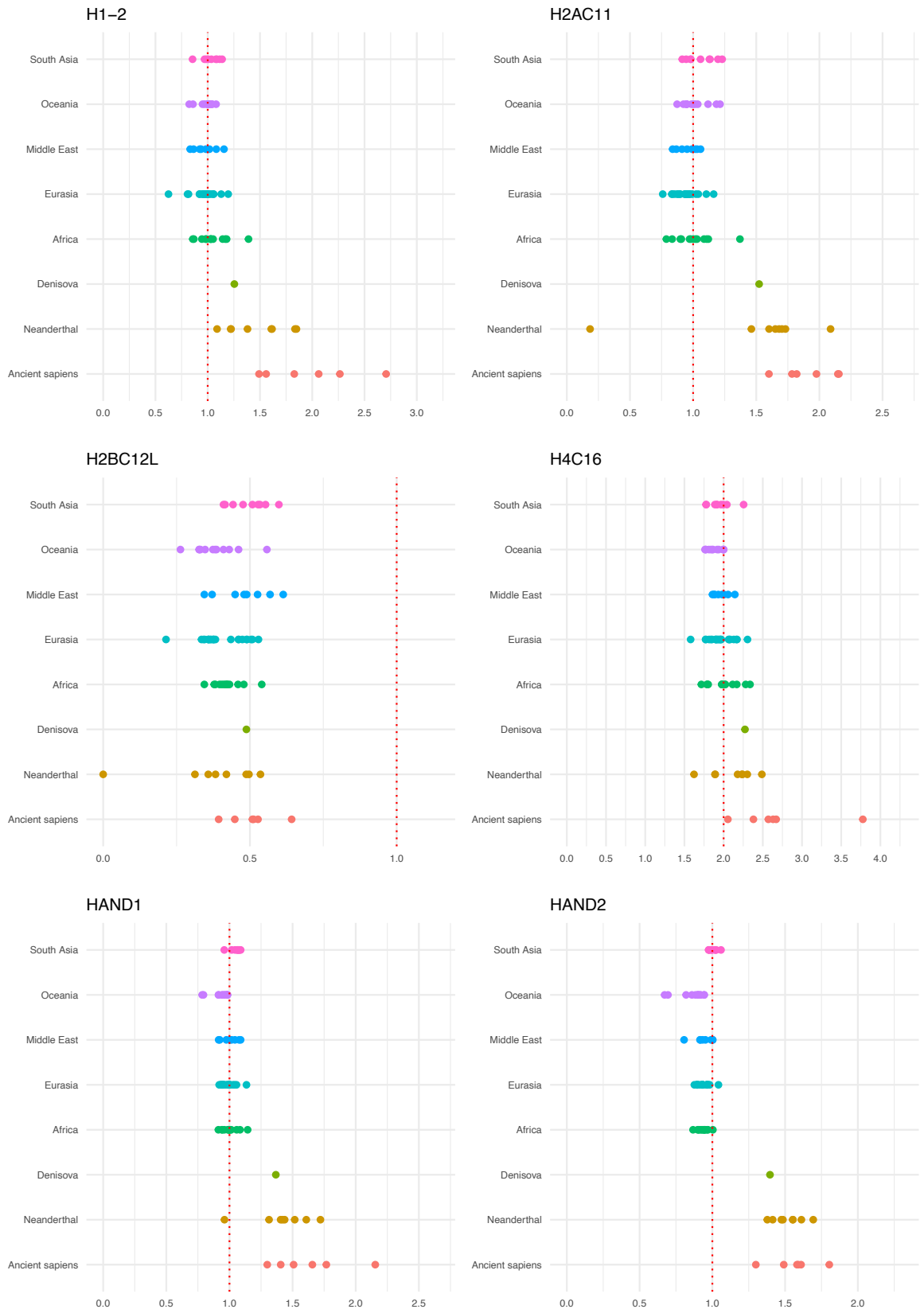


Figure S2 (continued).

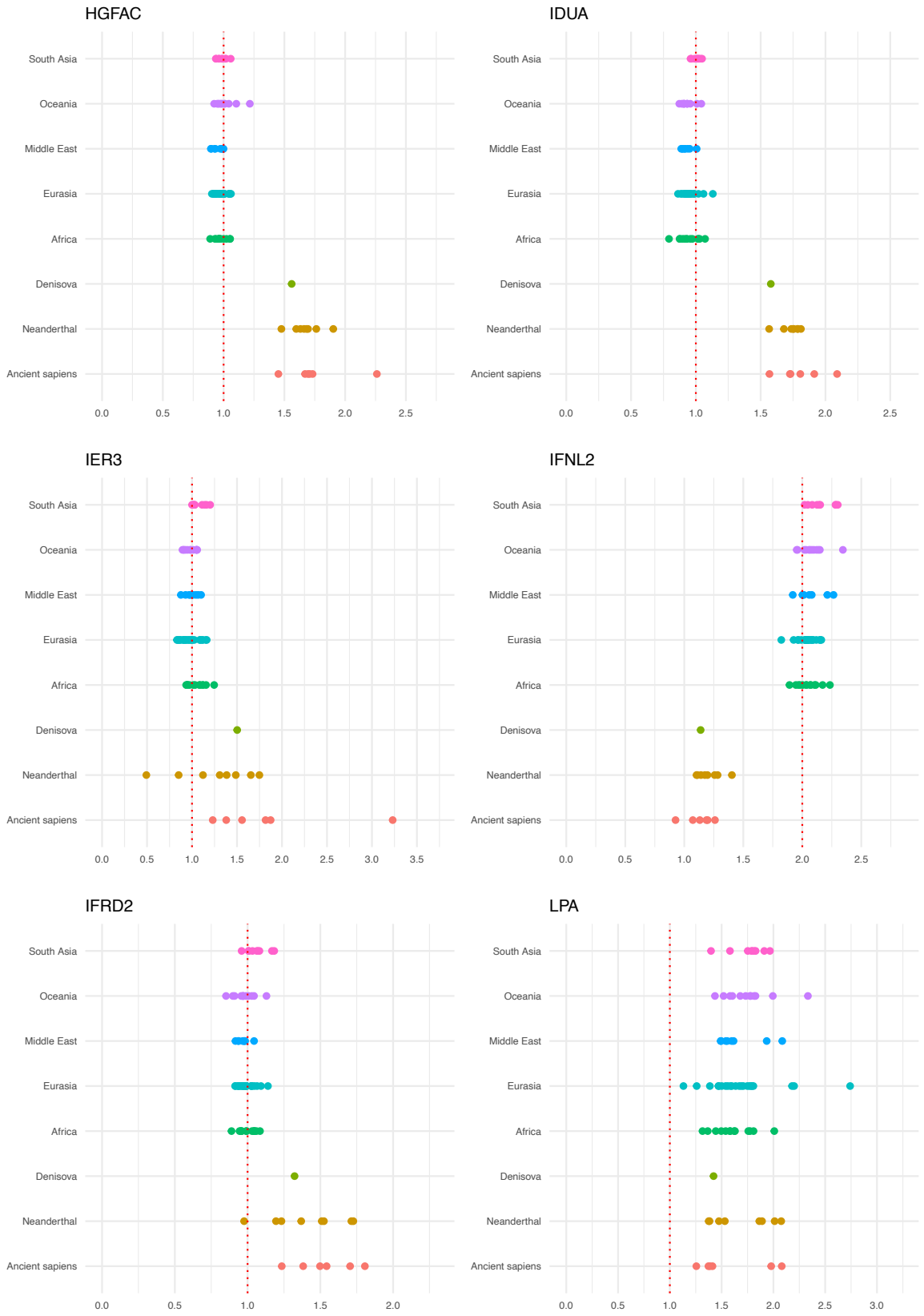


Figure S2 (continued).

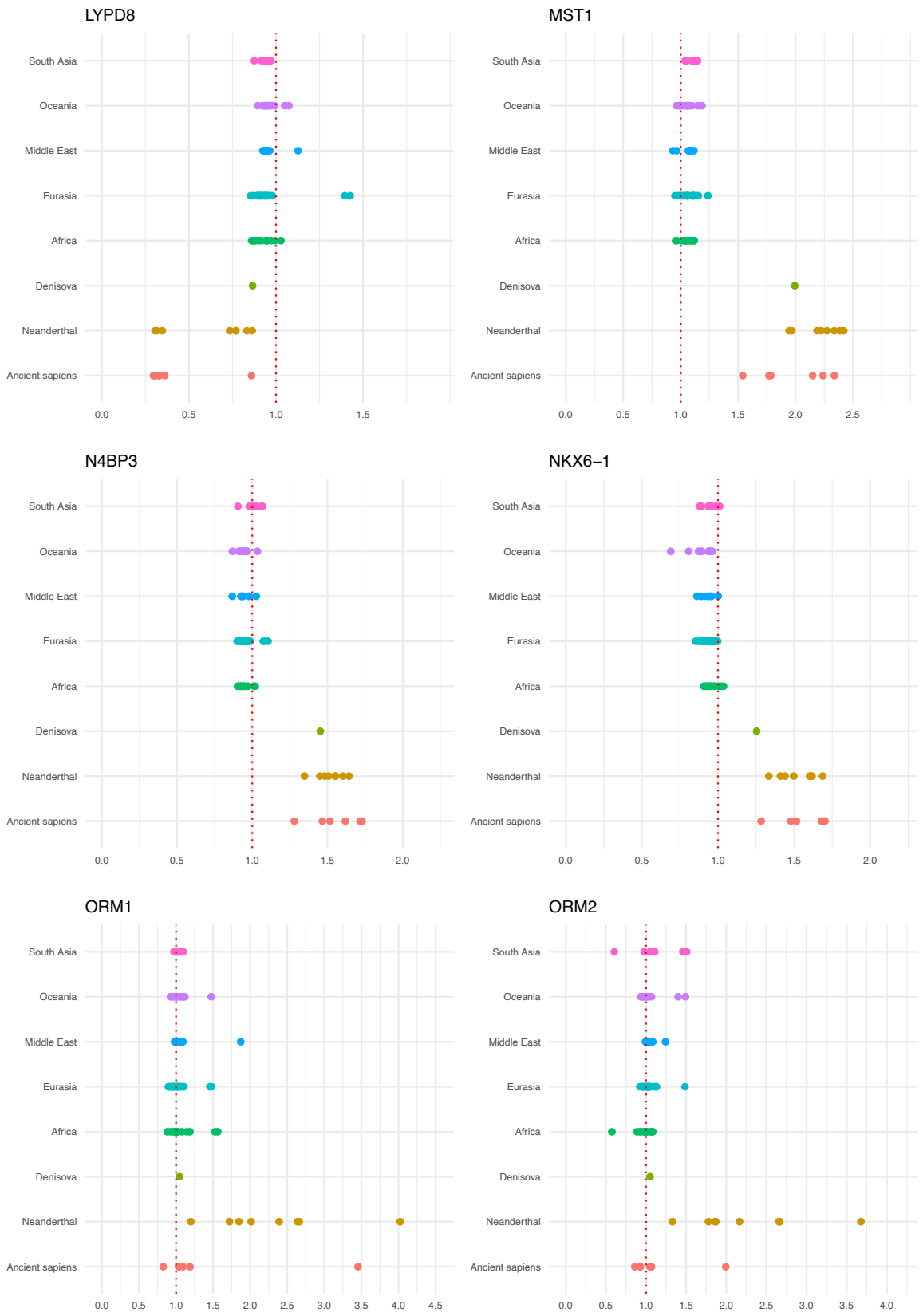


Figure S2 (continued).

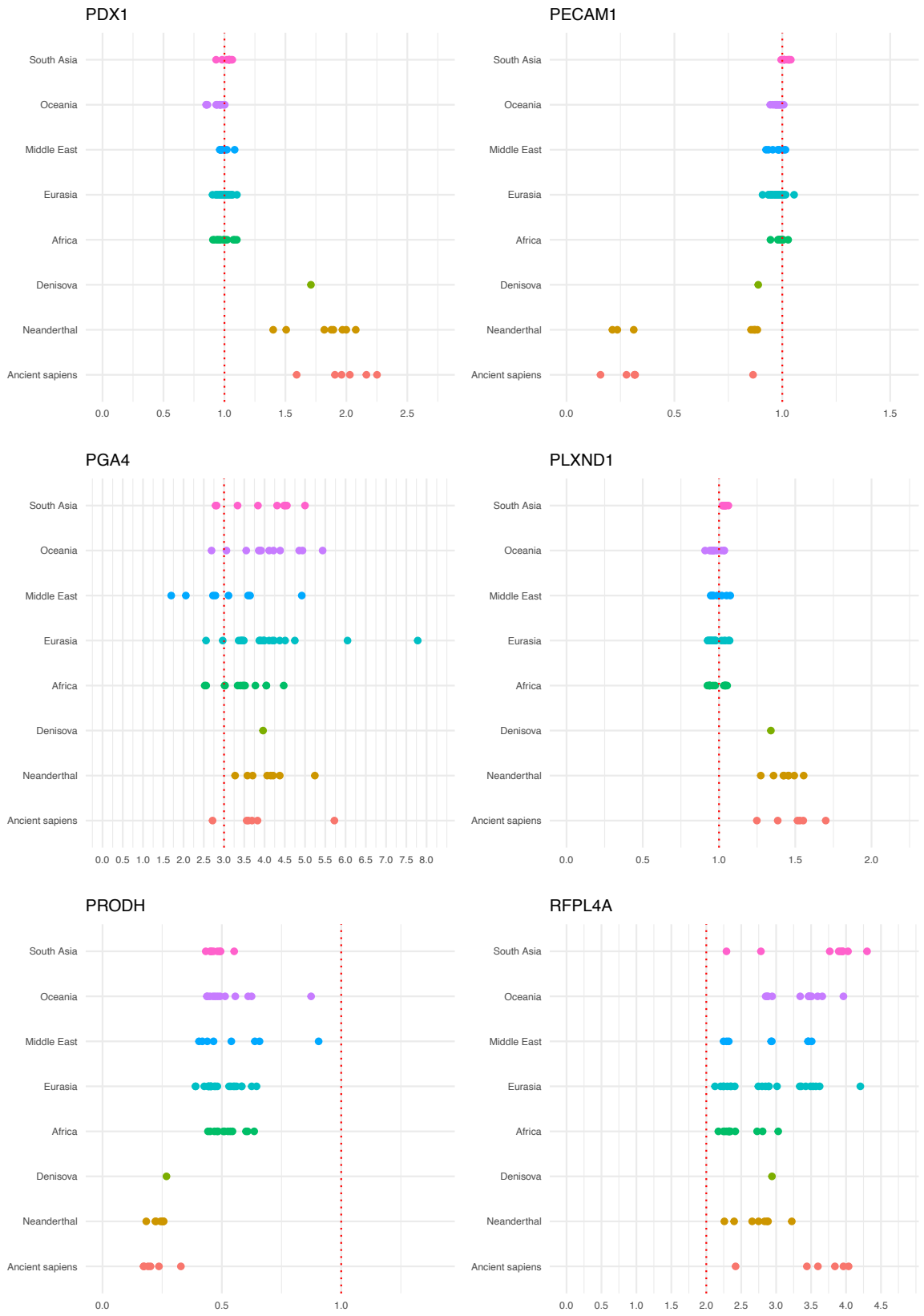


Figure S2 (continued).

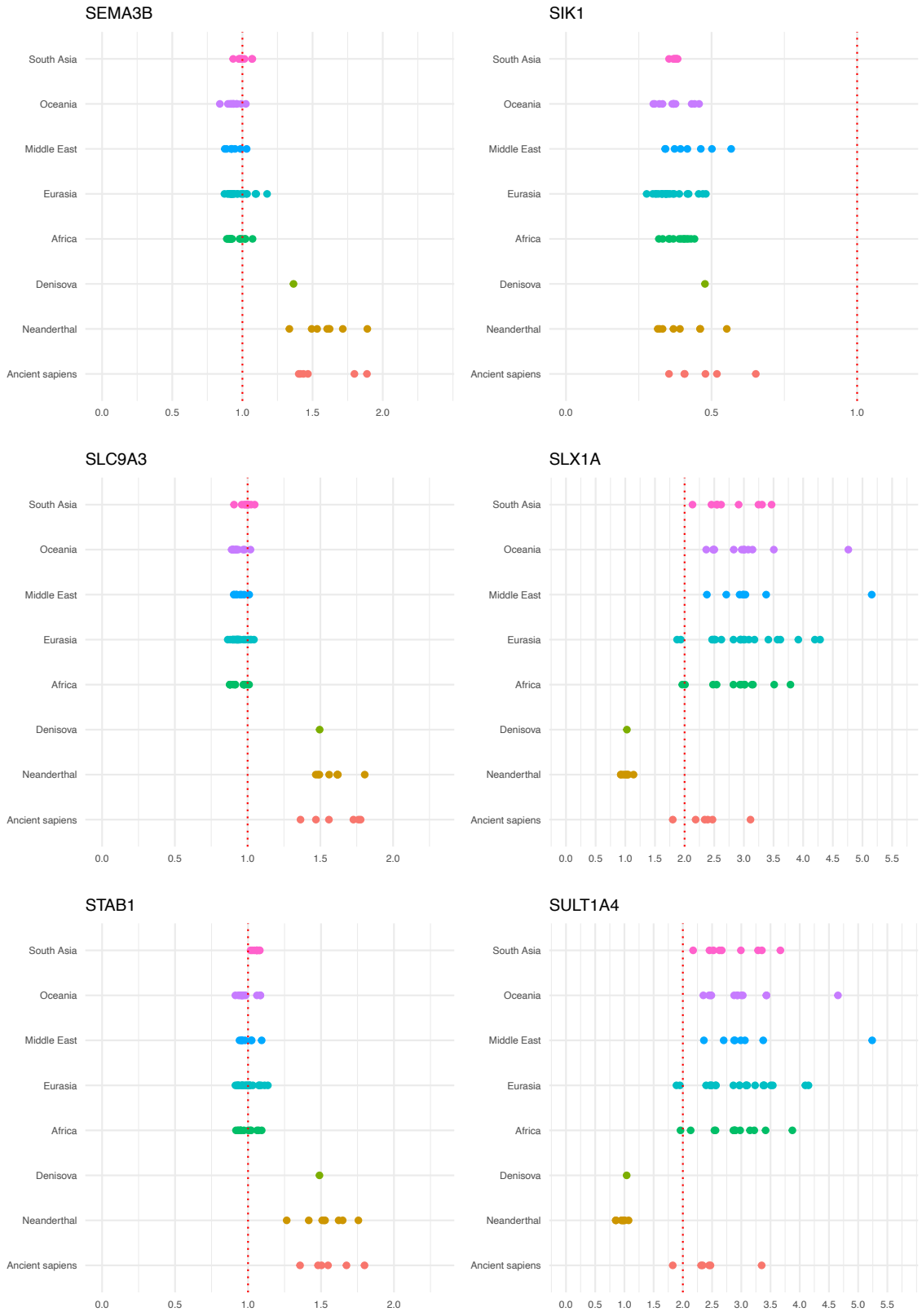


Figure S2 (continued).

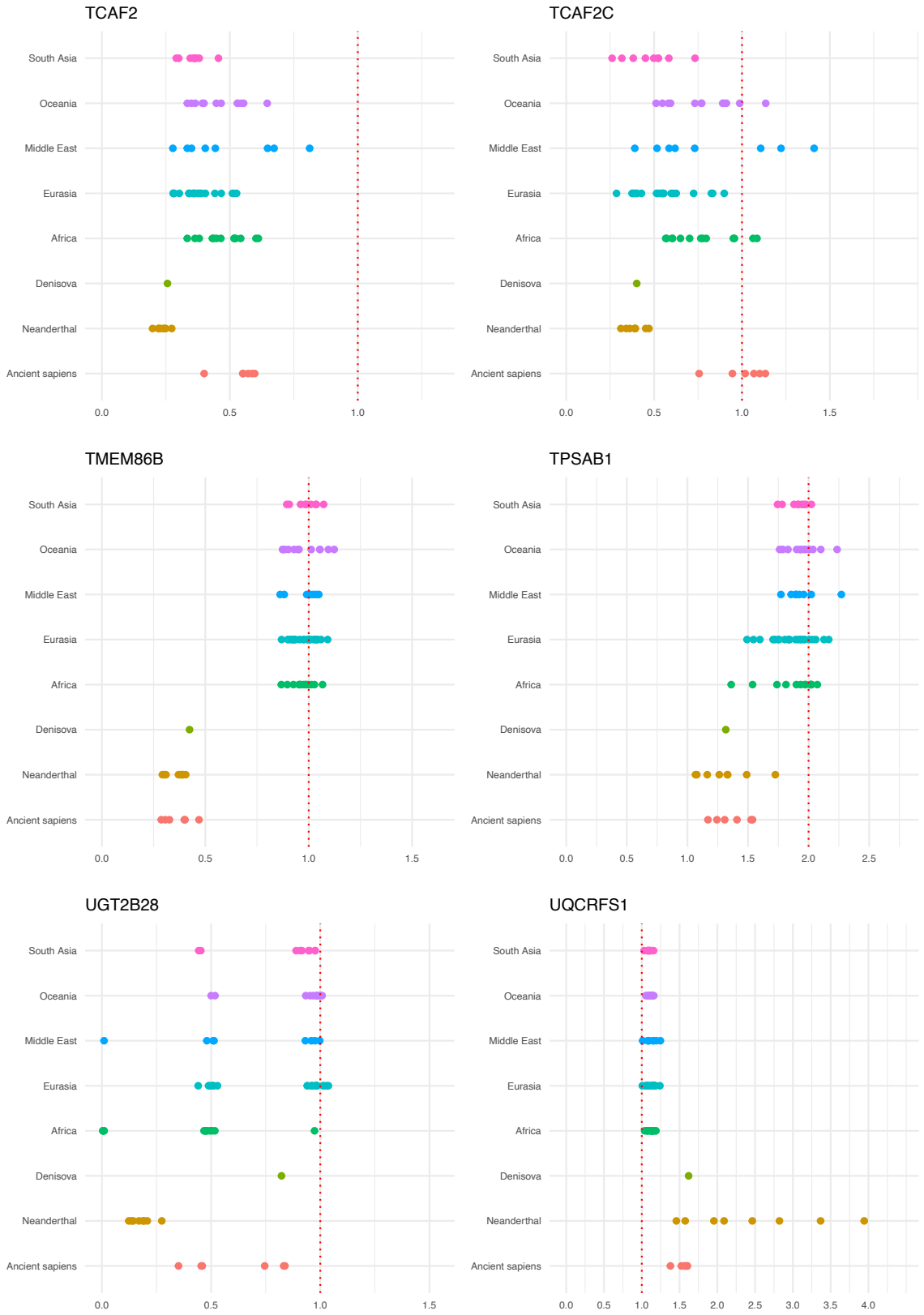


Figure S2 (continued).

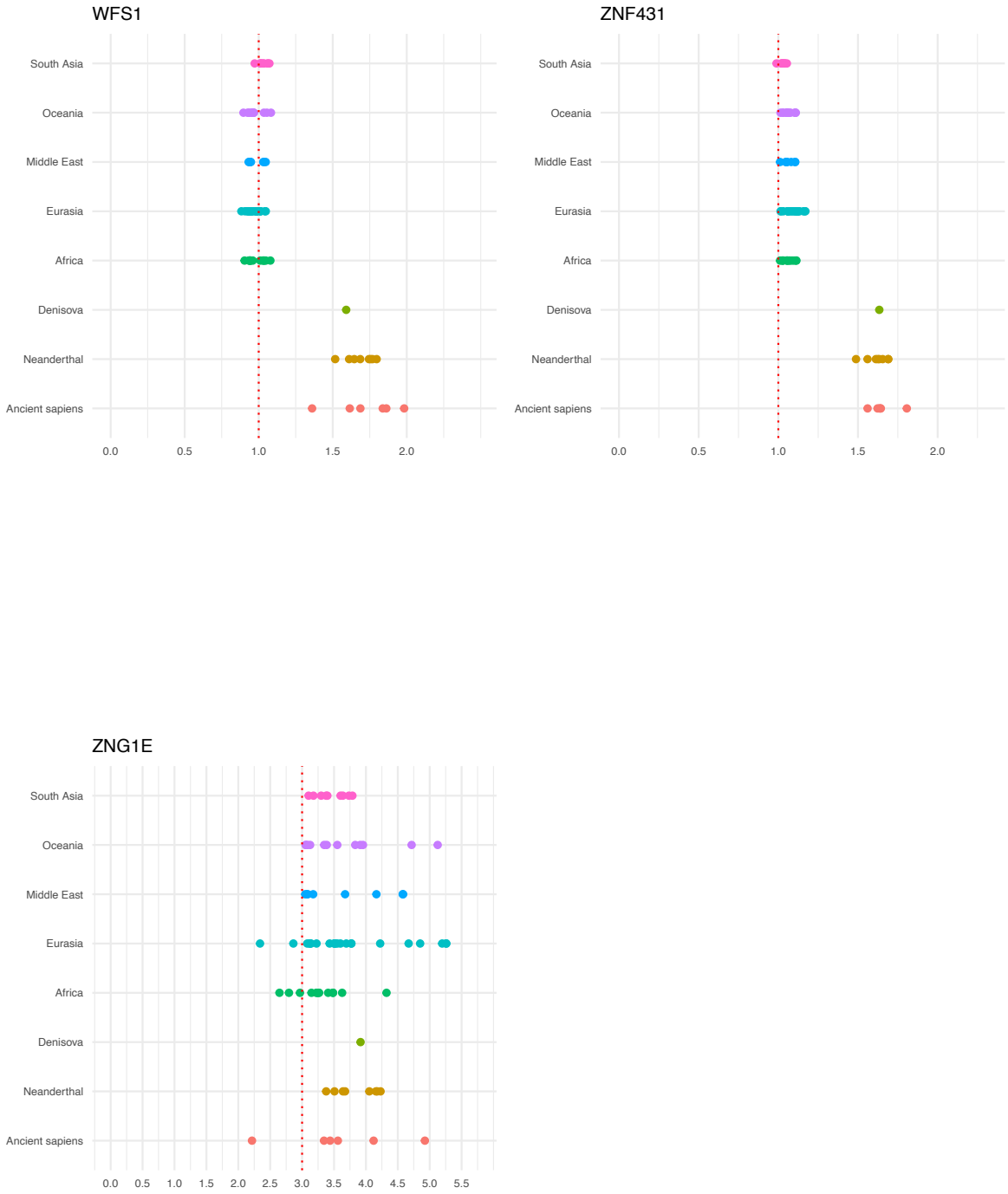


Figure S2 (continued).

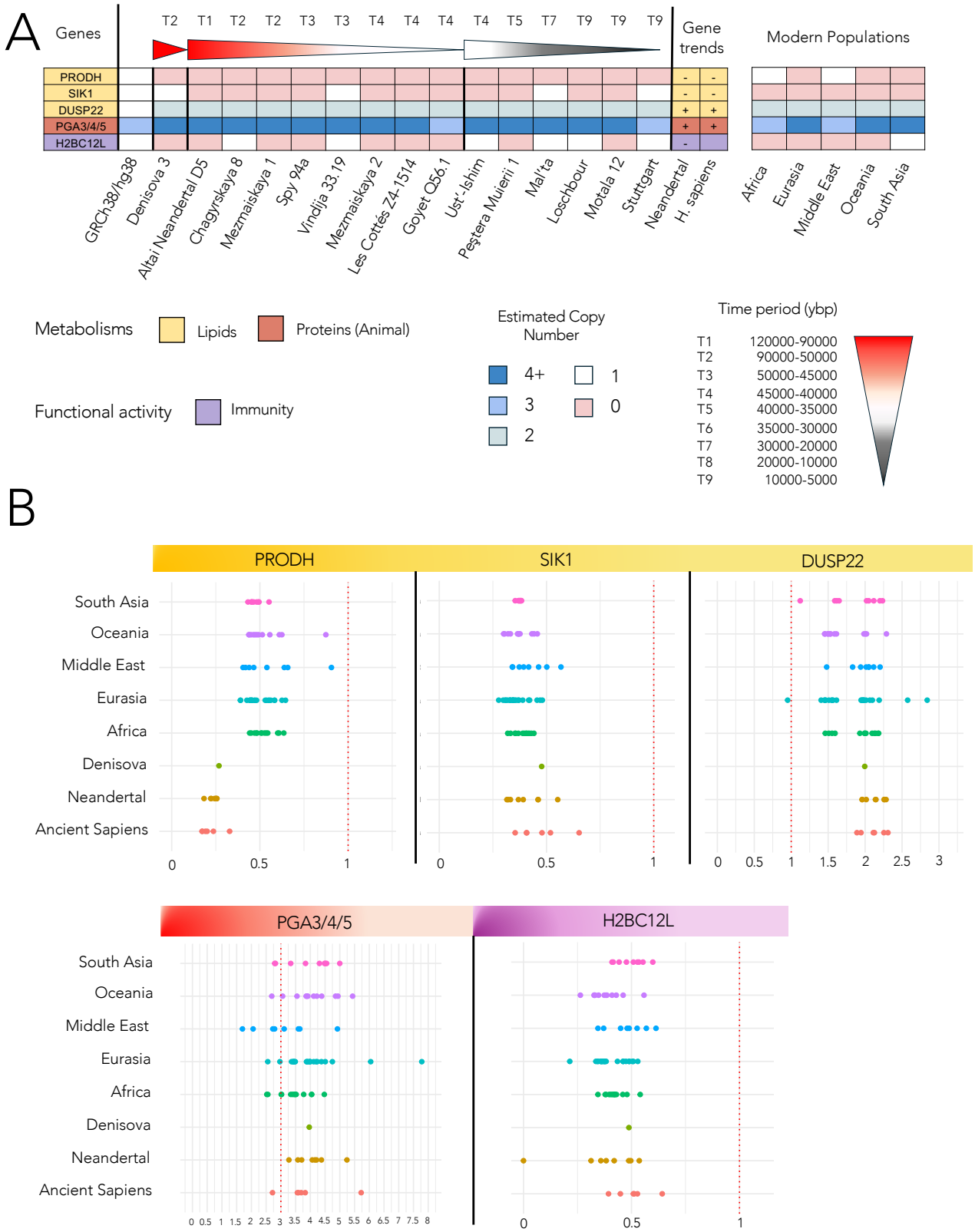


Figure S3. Genes filtered by the analysis of CNV based on CNV distributions on modern populations. Five genes have been filtered out, *PRODH, SIK1, DUSP22, PGA3/4/5* and *H2BC12L*. **A.** Compare with Tables 1 and 2. **B.** Compare with Figure 3.

C

		Africa													Eurasia																					
Genes	GRCh38/hg38	Gambian1	Gambian2	Mandenka1	Mandenka2	Mandenka3	Somali1	BantuKenya1	BantuKenya2	BantuKenya3	BantuTswana1	BantuTswana2	KhomaniSan1	KhomaniSan2	Icelandic1	Icelandic2	Saami1	Saami2	Mansi1	Mansi2	Tubalar1	Tubalar2	Altaian1	Mongola1	Mongola2	Oroqen1	Oroqen2	Oroqen3	Even1	Even2	Even3	Yakut1	Yakut2	EskimoChaplin1	EskimoNaukan1	EskimoNaukan2
DUSP22	1	2	1.51	2.14	2.18	2.12	1.93	2.18	1.59	2.09	1.99	1.55	1.46	2.01	1.95	1.51	1.45	1.97	0.95	1.97	1.98	1.47	1.4	1.99	1.56	2.1	2.19	2.06	2.84	1.95	1.46	2.58	1.61	1.55	1.57	2.01
PRODH	1	0.47	0.51	0.6	0.53	0.48	0.64	0.54	0.44	0.61	0.51	0.48	0.54	0.45	0.39	0.58	0.39	0.45	0.44	0.46	0.44	0.46	0.58	0.65	0.62	0.48	0.55	0.43	0.54	0.47	0.45	0.55	0.56	0.47	0.45	0.53
SIK1	1	0.35	0.39	0.41	0.42	0.41	0.44	0.4	0.37	0.4	0.32	0.36	0.33	0.43	0.34	0.47	0.28	0.33	0.32	0.35	0.31	0.33	0.46	0.42	0.48	0.37	0.42	0.39	0.36	0.35	0.31	0.37	0.34	0.33	0.34	0.3
PGA4	3	4.04	4.05	4.47	3.49	3.77	3.46	3.4	3.34	2.56	3.52	3.02	2.52	3.01	2.56	4.19	3.97	2.97	3.87	3.44	3.98	3.49	4.75	7.78	4.24	3.92	4.11	3.36	3.41	3.43	4	4.51	3.47	4.38	6.05	3.88
H2BC12L	1	0.34	0.4	0.42	0.48	0.54	0.43	0.4	0.42	0.41	0.38	0.38	0.42	0.46	0.49	0.51	0.38	0.35	0.21	0.34	0.36	0.34	0.47	0.5	0.38	0.46	0.53	0.49	0.38	0.37	0.37	0.36	0.44	0.36	0.37	0.46

		Middle East								Oceania												South Asia								
Genes	GRCh38/hg38	Palestinian1	Palestinian2	Palestinian3	Iranian1	Iranian2	Balochi1	Balochi2	Balochi3	Dusun1	Dusun2	Igorot1	Igorot2	PapuanHighlands1	PapuanHighlands2	PapuanHighlands3	PapuanSepik1	PapuanSepik2	PapuanSepik3	Australian1	Australian2	Xibo1	Xibo2	Xibo3	Lahu1	Lahu2	Lahu3	Cambodian1	Cambodian2	Cambodian3
DUSP22	1	2.05	1.48	2.05	1.94	2.01	2.12	2.2	1.83	1.46	1.98	1.5	1.45	1.53	1.53	1.5	1.61	2	2.02	1.59	2.29	2.24	1.62	2.12	2.05	2.02	1.12	1.59	2.2	1.65
PRODH	1	0.4	0.64	0.91	0.42	0.66	0.47	0.44	0.54	0.46	0.45	0.44	0.47	0.49	0.51	0.62	0.56	0.61	0.48	0.44	0.87	0.43	0.46	0.45	0.46	0.49	0.45	0.49	0.55	0.48
SIK1	1	0.34	0.5	0.57	0.34	0.46	0.39	0.37	0.42	0.38	0.37	0.32	0.3	0.37	0.3	0.46	0.43	0.44	0.33	0.36	0.37	0.35	0.37	0.38	0.37	0.38	0.37	0.38	0.38	0.38
PGA4	3	3.64	2.79	3.59	2.06	4.92	1.7	2.73	3.11	3.88	3.92	4.39	5.43	3.55	4.11	4.86	3.86	4.94	3.06	2.69	4.22	3.84	2.79	3.33	5	4.31	4.55	2.82	4.48	4.51
H2BC12L	1	0.37	0.53	0.61	0.34	0.57	0.48	0.49	0.45	0.37	0.41	0.33	0.35	0.43	0.33	0.56	0.39	0.46	0.33	0.38	0.26	0.44	0.48	0.41	0.42	0.53	0.53	0.6	0.55	0.51

D

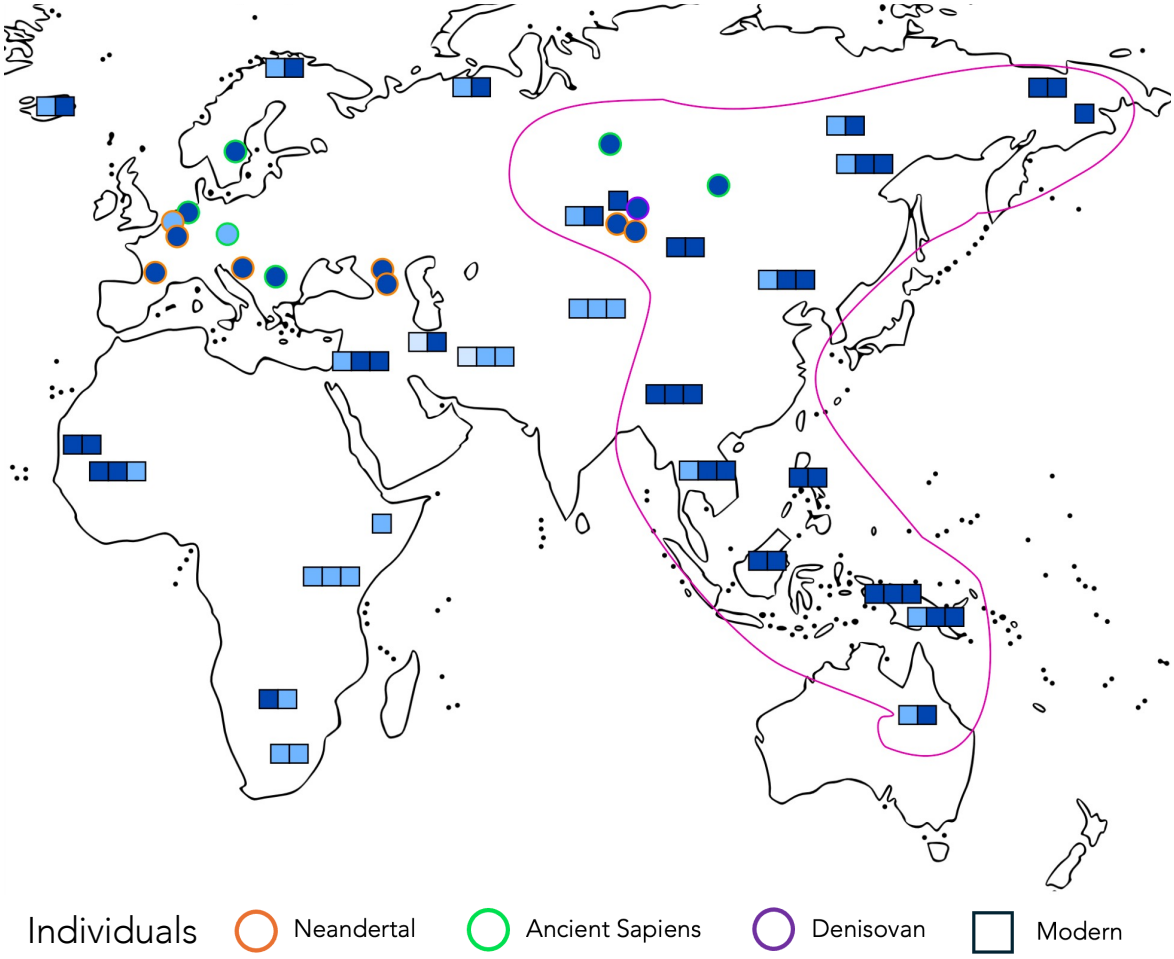


Figure S3 (continued). **Genes filtered by the analysis of CNV based on CNV distributions on modern populations**. **C.** CNV estimations of the five genes *PRODH, SIK1, DUSP22, PGA3/4/5* and *H2BC12L*. **D.** Visualisation of *PGA3/4/5* haploid CNVs (colors as shown in **A** and **C**) for both modern (square) and ancient (circle) individuals. Different populations are identified by colored borders: *Hn* in orange, *Hs* in green, *Den* in purple, and modern individuals in black. Most modern genomes from the Eastern regions of Eurasia, the South Asia and Oceania (pink shape) exhibit high CNVs, similar to ancient individuals in Eurasia, while modern genomes in the Middle East, and in Africa show lower CNVs.

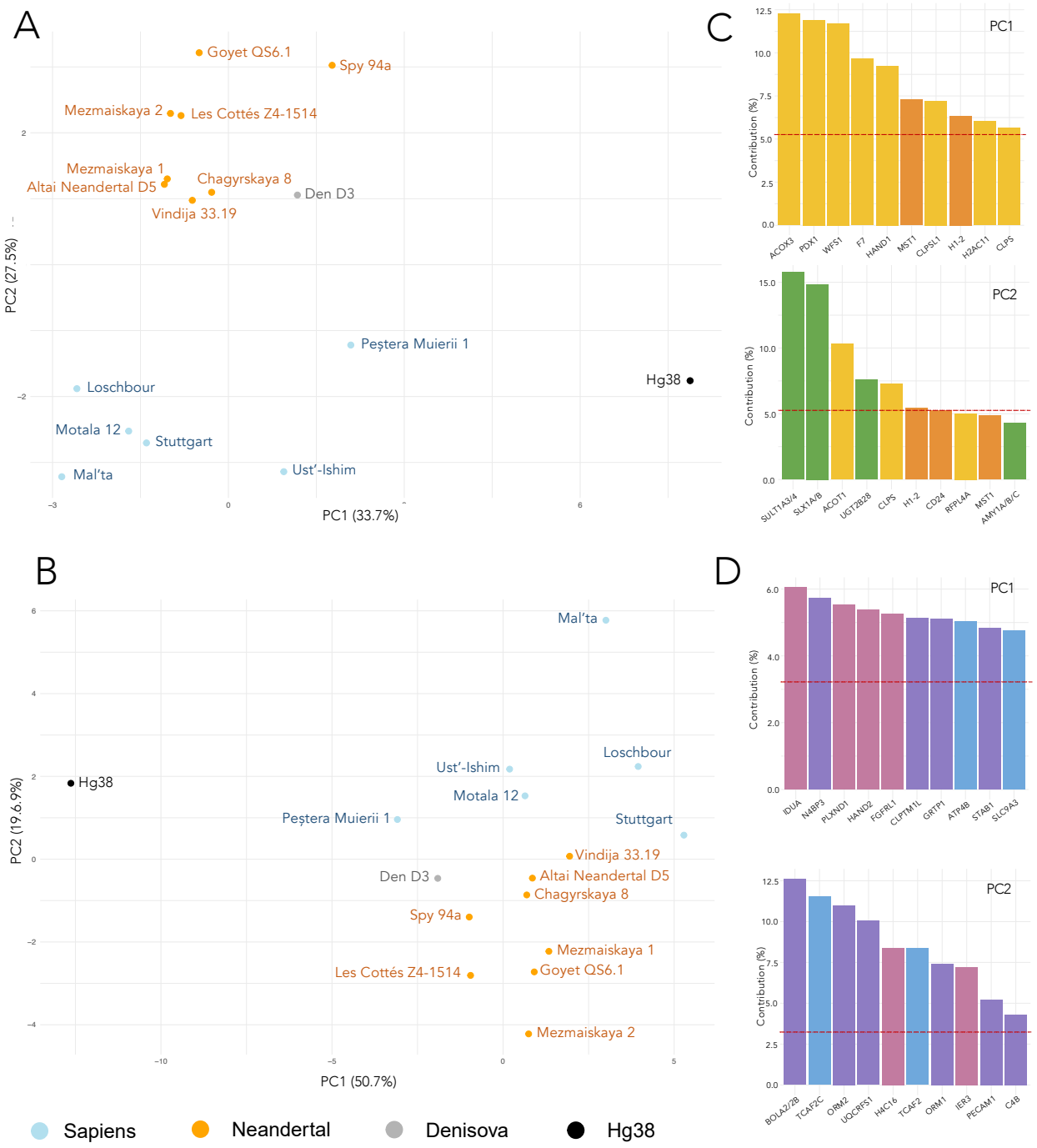


Figure S4. Comparative CNV analysis of *Hn*, *Den* and *Hs* living in the European belt with the reference modern human genome. **A.** Fifteen *Hn*, *Den* and *Hs* individuals and the reference modern genome hg38 are represented in a three-dimensional space obtained by PCA from the 19-dimensional space defined from differential CNV values of the 10 diet-related genes. View of the 2D space where each individual (points) is colored with respect to human populations: *Hn* (orange), *Den* (grey), and *Hs* (light blue). The modern reference genome GRGh38/hg38 is also plotted (blue). PC1 explains the 32.7% of the variance and PC2 the 28.9%. **B.** Same analysis as in A but based on the differential CNV values of the 31 accessory genes. PC1 explains the 49.0% of the variance and PC2 the 20.7%. **C.** Contributions of factors for the definition of each principal component of the 2D space, obtained by PCA for the 19 genes analysis in **A**. Colored bars indicate important contributions from CNV in genes acting in the three overarching metabolic processes (colors as in **Figure 2F**). **D.** Contributions of factors for the definition of each principal component of the 2D space, obtained by PCA for the 31 genes analysis in **B**. See legend in **C**.

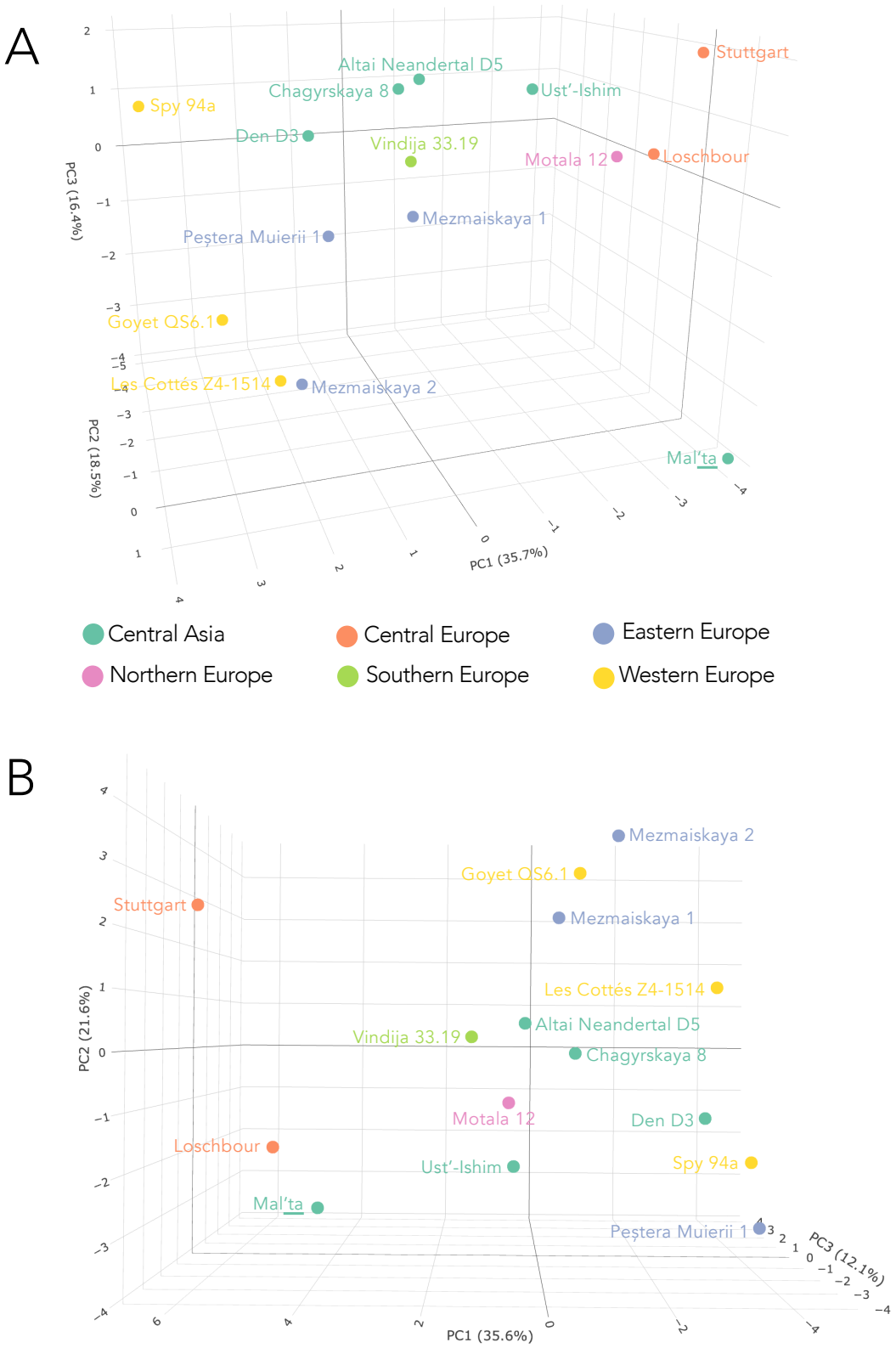


Figure S5. Comparative CNV analysis of *Hn*, *Den* and *Hs* living in the European belt: geographical regions. Complement to **Figure 4**. **A.** Fifteen *Hn*, *Hs* and *Den* individuals are represented in a three-dimensional space obtained by PCA from the 19-dimensional space defined from diet-related gene CNV (see **Supplementary Text 2**). Individuals are colored with respect to geographical localisation. **B.** As in A but with the 15 individuals plot in the three-dimensional space obtained by PCA from the 31-dimensional space defined from accessory gene CNV.

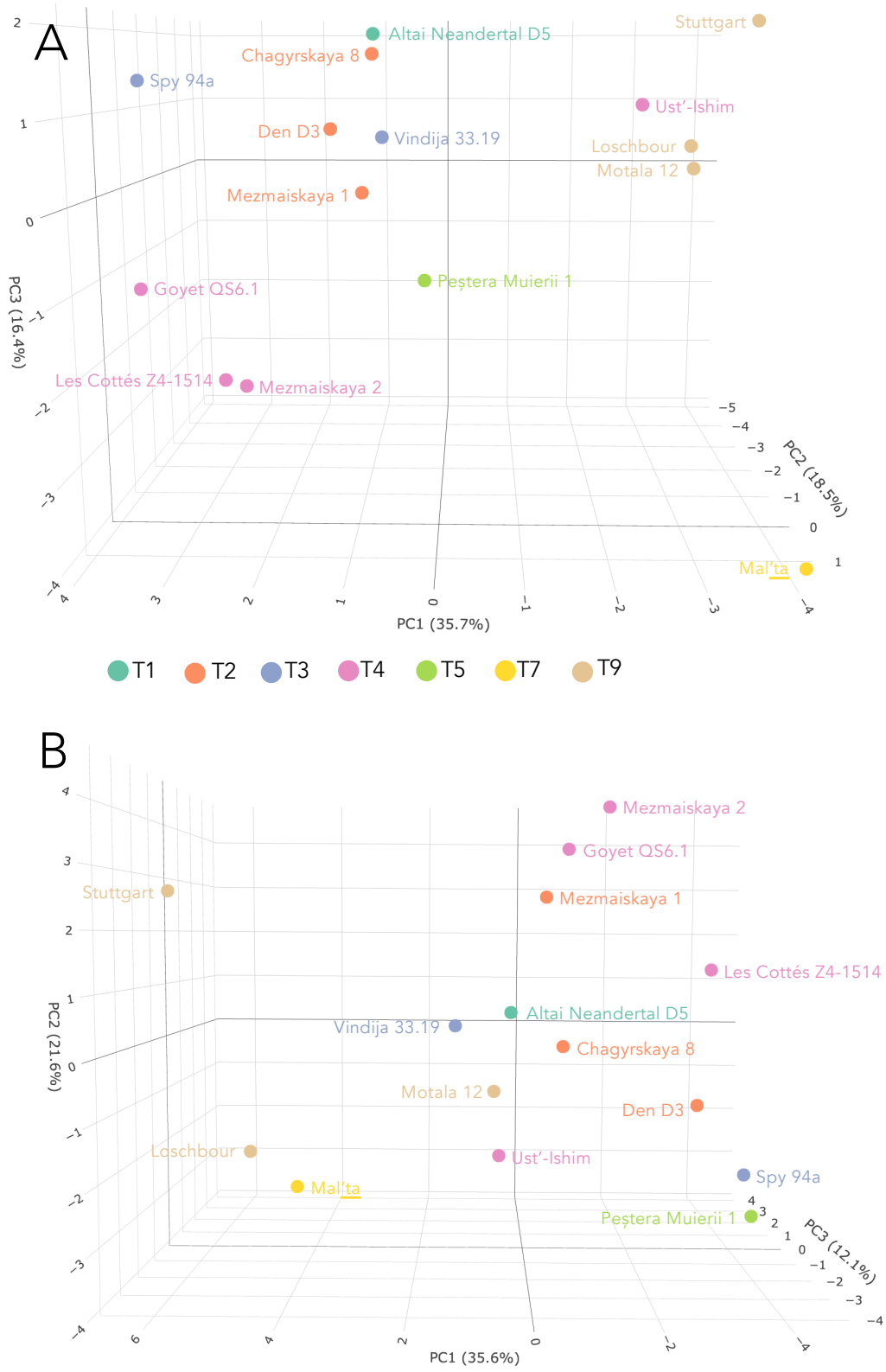


Figure S6. Comparative CNV analysis of *Hn*, *Den* and *Hs* living in the European belt: time period. Complement to **Figure 4**. **A.** Fifteen *Hn*, *Hs* and *Den* individuals are represented in a three-dimensional space obtained by PCA from the 19-dimensional space defined from diet-related gene CNV (see **Supplementary Text 2**). Individuals are colored with respect to time periods. **B.** As in A but with the 15 individuals plot in the three-dimensional space obtained by PCA from the 31-dimensional space defined from accessory gene CNV.

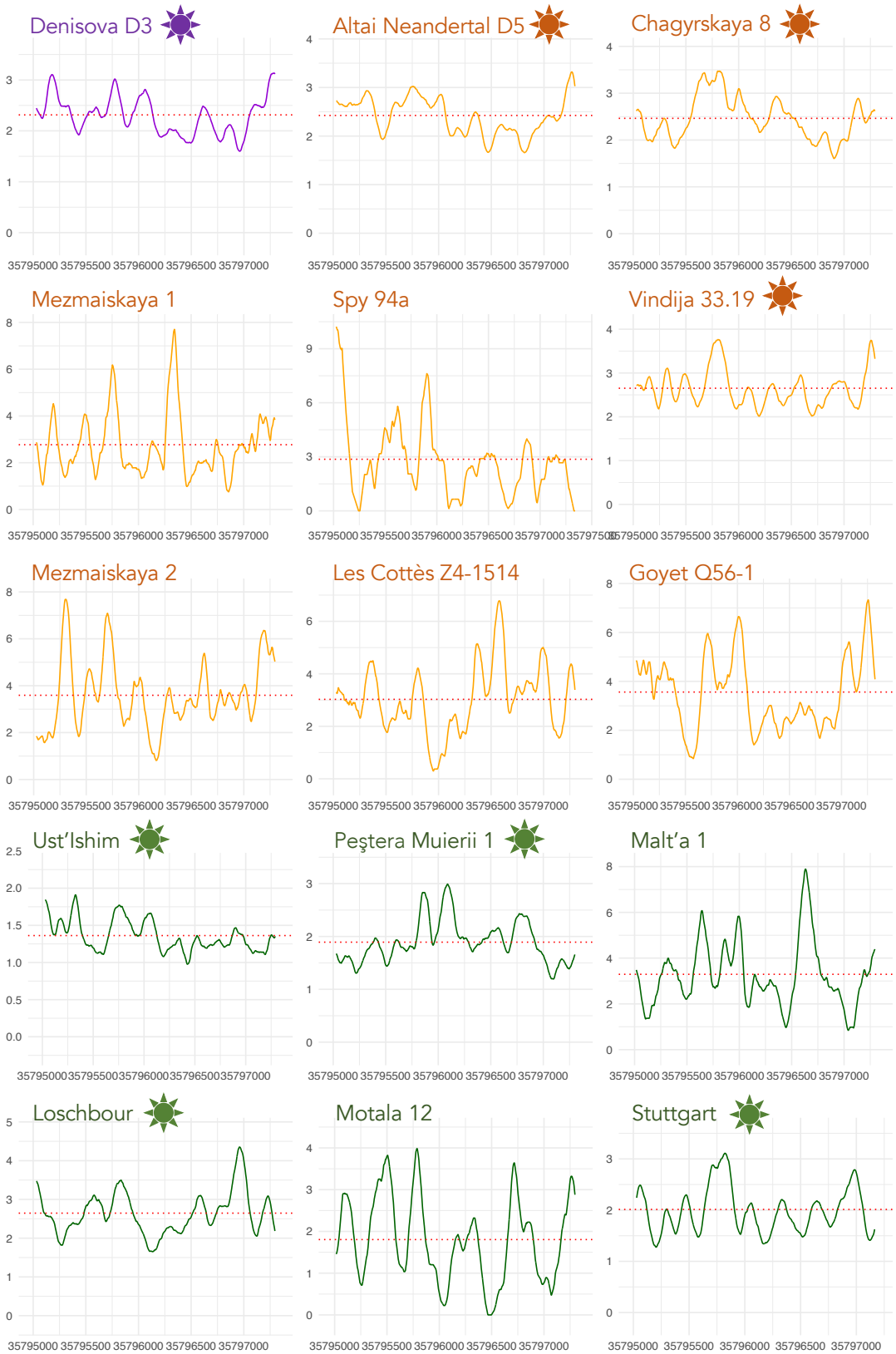


Figure S7. Positional coverage profiles from the read mapping of the CLPS gene. Positional coverage profiles for the CLPS gene across the 15 ancient genomes analyzed in this study, representing Denisovan (purple), Neandertal (orange), and Sapiens (green) populations. The profiles are computed using a sliding window of 100 bp across the gene length. The red dotted line indicates the CNV estimation of CLPS in each genome. The x-axis represents the positions of the CLPS gene on the chromosome, while the y-axis shows the corresponding positional coverage. The star symbol next to the individual's name indicates a high coverage genome (**Table S1**).

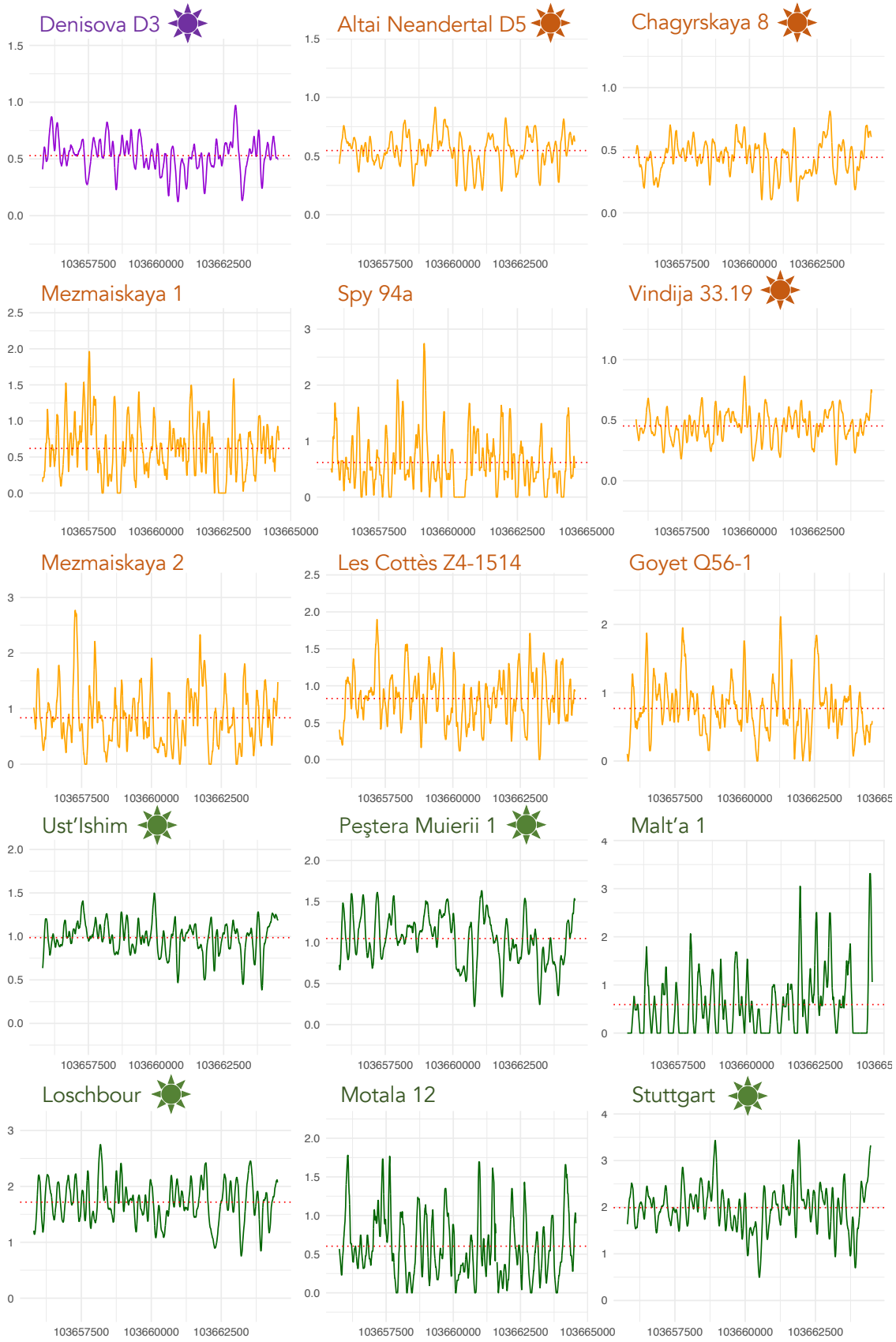


Figure S8. Positional coverage profiles from the read mapping of the the copies of the AMY1A gene. See legend of Figure S7.

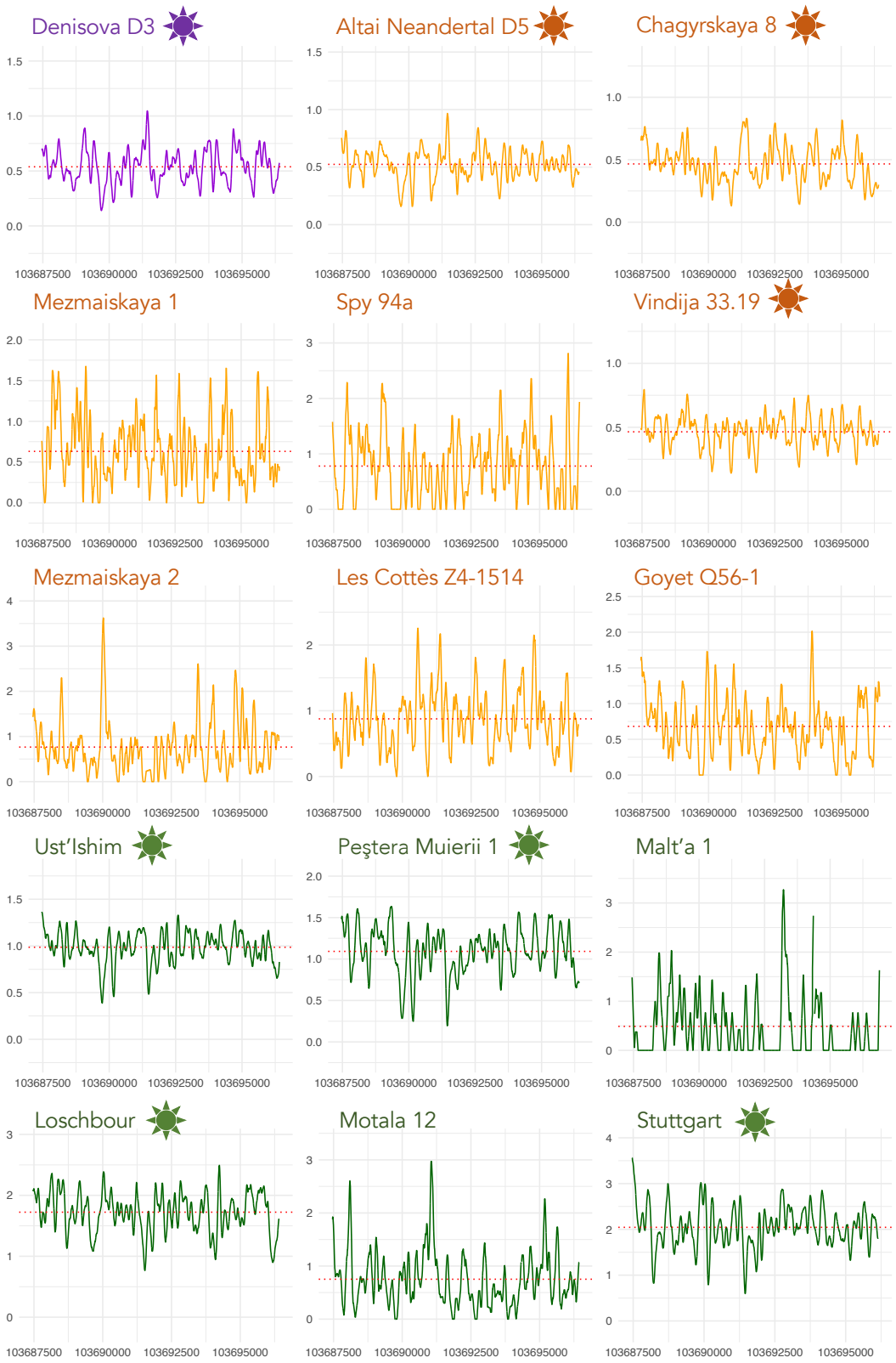


Figure S8 (continued) Positional coverage profiles from the read mapping of the the copies of the AMY1B gene. See legend of Figure S7.

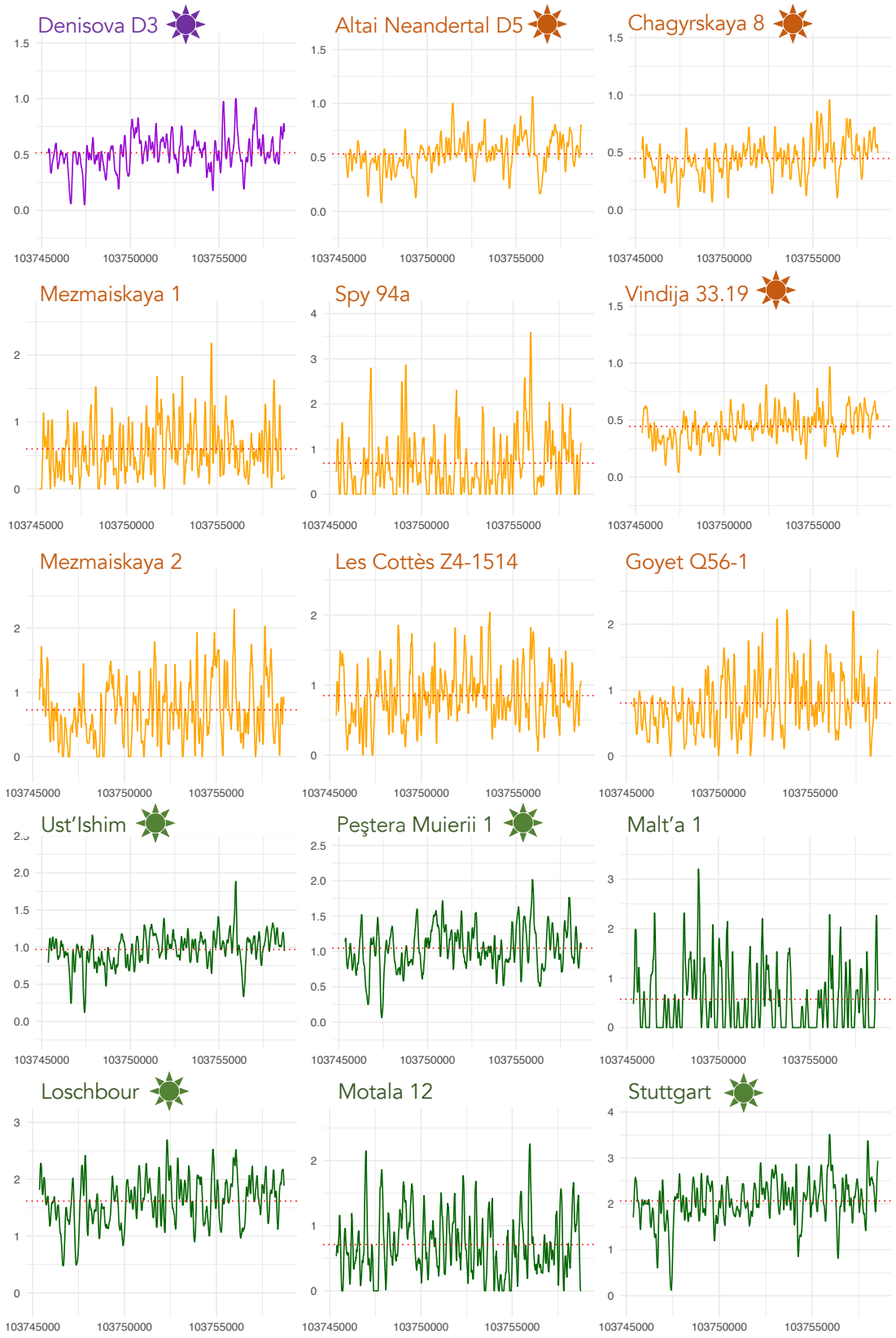


Figure S8 (continued) Positional coverage profiles from the read mapping of the the copies of the AMY1C gene. See legend of Figure S7.