ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

SCIENZE DELLA TERRA, DELLA VITA E DELL'AMBIENTE

Ciclo 38

Settore Concorsuale: 05/B1 - ZOOLOGIA E ANTROPOLOGIA

Settore Scientifico Disciplinare: BIO/08 - ANTROPOLOGIA

TRACING THE EVOLUTION OF HUMAN POLYGENIC ADAPTIVE TRAITS USING HIGH-RESOLUTION GENOMIC DATA

Presentata da: Giulia Ferraretti

Coordinatore Dottorato

Barbara Cavalazzi

Supervisore

Marco Sazzini

Co-supervisore

Stefania Sarno

Esame finale anno 2026

TABLE OF CONTENTS

ABSTRACT

It is well established that the ancestors of modern human populations have been exposed to multiple selective pressures during their migrations across the globe. This led different *Homo sapiens* populations to evolve local genetic adaptations, which enabled our species to thrive in diverse environmental settings. Nevertheless, the genetic bases of polygenic/complex adaptive traits, which are supposed to be responsible for the evolution of rapid adaptations, are still understudied due to the lack of proper methodological approaches. Here, we set up and fine-tuned a pipeline of analyses aimed at detecting genomic signatures ascribable to polygenic adaptations, which enabled the identification of functionally related gene networks pervasively targeted by the action of natural positive selection. The reliability of such integrated approach was tested on a large panel of whole genome sequence data belonging to several world-wide populations. Among the most compelling results, we detected polygenic selection signatures at loci modulating immune reactions of West Africans to endemic infections (e.g. *VTN*), and regulating cold-induced physiological responses in Eurasians (e.g. *THRB*) and Native Americans (e.g. *IRS1)*, which also potentially led to dis-adaptations in some ethnic groups form the latter genetic cluster due to recent environmental/dietary shifts. Moreover, polygenic bases of adaptations to hypobaric hypoxia were explored by detecting novel patterns of Denisovan adaptive introgression in a Tibetan population and genomic signatures of convergent evolution between Himalayan and Andean high-altitude populations involving pathways with pro-angiogenic/cardiovascular protective roles. Signatures of adaptive introgression were detected in the Siberian Yakut population as well and were found to have shaped their cold-induced adaptive responses mediated by insulin/lipid metabolisms. Overall, the present thesis proposes an alternative methodological workflow to investigate the evolution of human polygenic adaptations and provides new insights into the complex genetic architecture of adaptive traits evolved by our species to cope with a multitude of selective pressures.

1. INTRODUCTION

1.1 Investigating the genetic bases of human biological adaptations

The study of the genetic bases underpinning biological adaptations evolved by modern human populations constitutes a fascinating topic in the field of molecular anthropology and evolutionary genomics. Currently, it is well established that modern humans originated in Africa, in a time interval spanning from 300,000 (Stringer and Galway-Witham 2017) to 200,000 years ago (ya) (Stringer 2016), and subsequently migrated across the globe within a relatively short evolutionary timeframe (Oppenheimer 2012; Scheinfeldt and Tishkoff 2013; Willerslev and Meltzer 2021). During these migrations, the ancestors of non-African populations have had to progressively face a wide range of new selective pressures, including those induced by different climates, limited types of nutritional resources available, endemic pathogens, as well as in some cases extreme environmental conditions, such as those related to high-altitude (Scheinfeldt and Tishkoff 2013; Fan et al. 2016). Furthermore, multiple lines of evidence support extensive migrations of *Homo sapiens* across Africa as well (Rito et al. 2019; Bergström et al. 2021; Miller and Wang 2022), indicating a complex pattern of movements and interactions among early human populations. Given the vast diversity of biomes observable across the African continent, which encompass tropical, subtropical, arid, hyper-arid, and Mediterranean-type climates (Midgley and Bond 2015), we can hypothesize that also ancestral African populations have been exposed to a wide range of selective pressures throughout their evolutionary history, which led to the evolution of different local adaptations. To date, several studies successfully identified genes/genomic regions presenting signatures ascribable to the action of natural selection that likely concurred to the development of adaptive traits in modern humans (Figure 1.1).

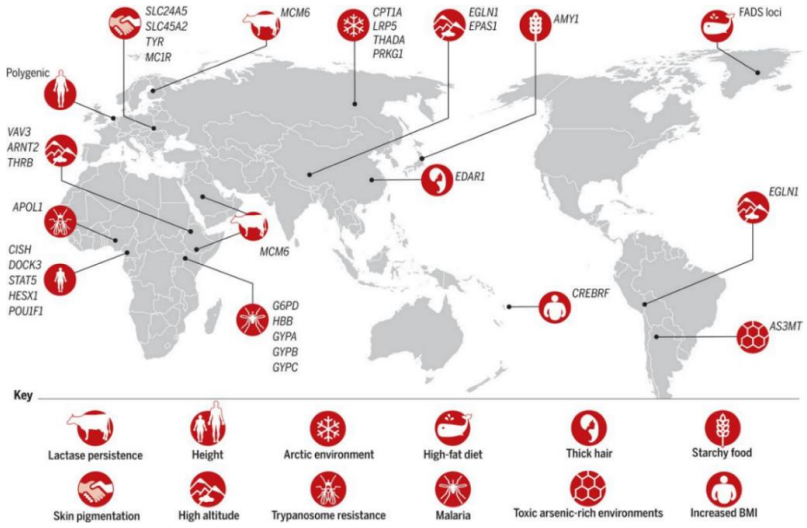


Figure 1.1. Map of the most well-known local adaptations evolved by modern human populations. Schematic representation of human local adaptations, each of them is labelled with the putative adaptive phenotype and/or the selective pressure that was hypothesized to have represented the evolutionary driving-force behind the signatures ascribable to the action of natural selection identified at the reported genes. The figure has been taken from Fan et al. (2016).

In this context, cardinal examples are represented by the footprints left by natural selection at loci linked to light skin pigmentation (Harding et al. 2000; Yang et al. 2018; Lamason et al.), pathogen-driven adaptations (Fumagalli et al. 2011; Couto-Silva et al. 2023), as well as the regulation of physiological and/or metabolic responses induced by climate (Hancock et al. 2011) and high-altitude (Bigham et al. 2010; Yi et al. 2010; Peng et al. 2011) (Figure 1.1). Notably, some of these traits (e.g. malaria resistance and metabolic adaptations to cold climates) have been also proposed to modulate convergent adaptations among modern human populations (Kwiatkowski 2005; Edwards et al. 2010; Balentine and Bolnick 2022), where convergent evolution refers to the independent emergence of similar biological traits in distinct species/populations as a result of similar selective pressures (Stern 2013). Nevertheless, most of these studies provided evidence concerning the occurrence of strong selective events (i.e. adaptations mediated by classical selective sweeps), which are evolutionary mechanisms that have been recently proposed to be relatively rare in our species (Hernandez et al. 2011).

1.2 Models describing the genomic footprints of natural selection

So far, multiple models capable of describing the patterns of genetic variability expected after the action of natural selection have been extensively developed. Primary, these models could be essentially divided in two groups: those accounting for classical selective sweeps (i.e. hard and soft selective sweeps) and those instead focused on the evolution of polygenic adaptations (Figure 1.2) (Scheinfeldt and Tishkoff 2013).

In detail, the hard selective sweep model describes adaptive events that took place after the occurrence of a single genetic mutation in a given population, which is highly beneficial and has a strong impact on a given phenotypic trait (Figure 1.2a) (Pritchard et al. 2010; Fu and Akey 2013; Scheinfeldt and Tishkoff 2013). The rapid rising in frequency of such beneficial mutation due to the action of positive natural selection comports a drastic reduction in the haplotypic variability surrounding the selected site due to the elevated linkage disequilibrium (LD) among adjacent single nucleotide variants (SNVs) (Figure 1.2a) (Pritchard et al. 2010; Fu and Akey 2013; Scheinfeldt and Tishkoff 2013). Due to these assumptions, hard selective sweeps are believed to have occurred relatively rarely throughout evolutionary history of human populations. Such selective events, in fact, require a substantial amount of time to evolve. Therefore, it is unlikely that multiple independent hard sweep events are responsible for driving complex adaptations to highly diverse environments in a species with a low mutation rate per generation and a low effective population size such as *Homo sapiens* (Terhorst et al. 2017). The soft selective sweep model assumes instead the existence of multiple positively selected variants/haplotypes, which could have been evolved neutrally for many generations and that, in the current environment, become beneficial (i.e. selection on standing genetic variation) (Figure 1.2b) (Pritchard et al. 2010; Fu and Akey 2013; Scheinfeldt and Tishkoff 2013). According to such a model, beneficial variants also rise quite rapidly in frequency into the population, comporting, however, a less marked loss in haplotype variability with respect to what assumed by the hard selective sweep model (Figure 1.2b). Given these

assumptions, the soft selective sweep model describes a more plausible pattern of adaptive evolution for populations belonging to our species.

Figure 1.2. Schematic representation of changes of haplotype variability according to multiple selection models. Each panel displays the shifts in frequency of variants over time during and after the action of natural selection. For all the panels, variants are displayed as circles on horizontal genomic segments (i.e. chromosomes) and advantageous mutations are represented with a star. **(a)** Illustration of the classical hard selective sweep model, according to which a novel adaptive variant arises in the population and rapidly increases in frequency until it approaches fixation. This process drastically reduces haplotype variability around the beneficial mutation due to the strong LD among nearby nucleotide sites and/or limited recombination over such a short timescale. **(b)** Selection on standing variation model, also known as soft selective sweep model, assumes that a single mutation (as reported in the example) and/or few more SNVs, already present in the gene pool of the population become advantageous in a new environment and increases in frequency over time until reaching moderate to elevated frequencies. Since such variants are carried by different haplotypes, genetic variation after soft selective sweep events is better preserved with respect to what happen when hard selective sweeps occur. **(c)** Selection on complex/polygenic traits involves multiple adaptive genetic loci on different chromosomes (i.e. segments reported with different colours) that lowly impact complex phenotypic traits as single units but that, collectively, are able to modulate polygenic adaptations. When such variants become advantageous, they increase in frequency as a set, leaving more subtle signatures of adaptation, which are characterized particularly by slight shifts in allele frequencies at multiple loci. The figure has been taken from Scheinfeldt & Tishkoff (2013).

Nevertheless, the scientific community largely agrees that adaptations occurred according to the polygenic model (Figure 1.2c) are increasingly supposed to represent the majority of the adaptive traits evolved by modern human populations to cope with a diversified range of selective pressures (Pritchard et al. 2010; Pritchard and Di Rienzo 2010; Barghi et al. 2020). Such a model, in fact, describes a more realistic theoretical framework, which is capable to explain also very rapid adaptations, that have plausibly played a crucial role during the expansion of our species around the globe (Pritchard et al. 2010). Particularly, the model of polygenic adaptation assumes the existence of several putatively adaptive variants, which can be located in diverse portions of a gene or even in different chromosomes, and that singularly have a low impact on phenotypic traits but, conjointly, can mediate the expression of complex biological traits (Figure 1.2c) (Pritchard et al. 2010; Pritchard and Di Rienzo 2010; Fu and Akey 2013; Scheinfeldt and Tishkoff 2013; Barghi et al. 2020). Furthermore, such combination of adaptive variants can differ among individuals from the same population, thus not necessarily implying great shifts in frequency of beneficial variants (Figure 1.2c) (Pritchard et al. 2010; Pritchard and Di Rienzo 2010; Fu and Akey 2013; Scheinfeldt and Tishkoff 2013; Barghi et al. 2020). Given all the features described above, polygenic adaptations are thought to be characterized by subtle genomic footprints left by the action of natural selection, which are extremely challenging to be detected (Pritchard et al. 2010; Scheinfeldt and Tishkoff 2013; Barghi et al. 2020), especially by adopting traditional selection scans alone (e.g. iHS, XP-EHH, PBS, nSL, DIND, etc.) (Sabeti et al. 2007; Yi et al. 2010; Ferrer-Admetlla et al. 2014).

Furthermore, a recent review by Barghi et al. (2020), introduced two important aspects that significantly contribute to describe the theoretical framework of polygenic adaptations. Particularly, such features are the *non-parallelism* and the *loci heterogeneity* due to events of redundancy. In detail, non-parallelism is defined as the acquisition of a certain adaptive phenotypes in populations subjected to similar selective pressures (i.e. replicate populations), which takes place throughout different alleles/loci (Barghi et al. 2020). This is a direct consequence of traits redundancy and loci heterogeneity. In fact, the redundancy is an evolutionary phenomenon according to which different combinations of alleles/loci produce the same phenotypic outcome (Nowak et al. 1997; Barghi et al. 2019; Barghi et al. 2020). In replicate populations, this causes polygenic adaptations to take place throughout heterogenous allele frequency shifts at the same and/or different loci (Figure 1.3) (Barghi et al. 2019; Barghi et al. 2020). Therefore, loci heterogeneity strictly depends on the original frequency of beneficial alleles in the populations, and hence by the populations' ancestral genomic backgrounds (Figure 1.3) (Barghi et al. 2019; Barghi et al. 2020). According to such a scenario, alleles with a starting higher frequency have a greater probability of reaching elevated frequency in the populations after the action of natural selection, potentially leading to a variation pattern resembling an incomplete selective sweep and/or soft selective sweep (i.e. selective sweep-like patterns) (Barghi et al. 2020). Rare and/or low frequency alleles will be instead likely characterized by more subtle frequency shifts, thus further contributing to frequency heterogeneity of loci involved in polygenic adaptations (Barghi et al. 2020).

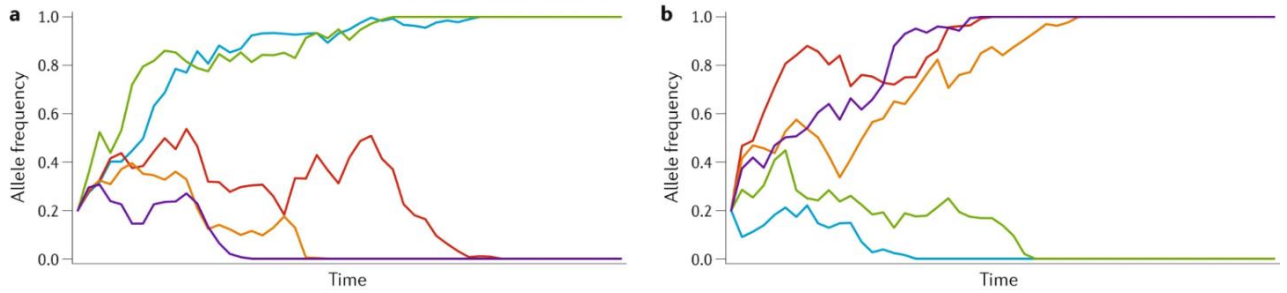


Figure 1.3. Non-parallelism, genetic redundancy and heterogeneity among loci are main characteristics of polygenic adaptations. For both panels, each line shows the trajectory of the frequency of each allele (y-axis) through time (x-axis), contributing to the selected complex trait in different simulated populations (a-b). Genetic redundancy results in non-parallel signatures of adaptation across replicate populations. Therefore, replicate populations reach the same trait optimum (i.e. evolve polygenic adaptation) using different alleles and/or genes, thus augmenting the heterogeneity of loci contributing to complex adaptations. The figure has been adapted from Barghi et al. (2020).

Therefore, based on these concepts, polygenic adaptation in replicate populations is achieved throughout the contribution of different genes/alleles that, nonetheless, regulate the same complex trait. Moreover, the same adaptive loci may contribute differently to complex adaptations due to the differences in the initial frequencies of the beneficial alleles in replicate populations.

1.3 Statistical approaches to investigate the evolution of polygenic adaptive traits

To date, several studies have been aimed at identifying SNVs associated to complex adaptive traits (Robinson et al. 2015; Bycroft et al. 2018; Adhikari et al. 2019; Ju and Mathieson 2021; Seo et al. 2022; He et al. 2023; Sohail et al. 2023) and/or entire genomic regions putatively involved in the evolution of polygenic adaptations (Hsieh et al. 2017; Gnecchi-Ruscone et al. 2018; Ojeda-Granados et al. 2022). However, the majority of them relied mainly on results obtained by means of genome-wide association studies (GWAS), which are known to have been designed according to the unrealistic common disease-common variant (CD-CV) hypothesis and have thus failed to describe an exhaustive picture of the genetic architecture of complex traits/diseases (Wang et al. 2005). Particularly, GWAS are a specific category of studies designed to detect significant associations between genetic markers and complex traits by comparing cohorts of cases (i.e. individuals presenting a given disease and/or trait) with defined control groups (Visscher et al. 2021). An example of GWAS output, particularly concerning the genetic variants found to be associated to light-skin pigmentation in Latin American populations, is displayed in Figure 1.4 (Adhikari et al. 2019).

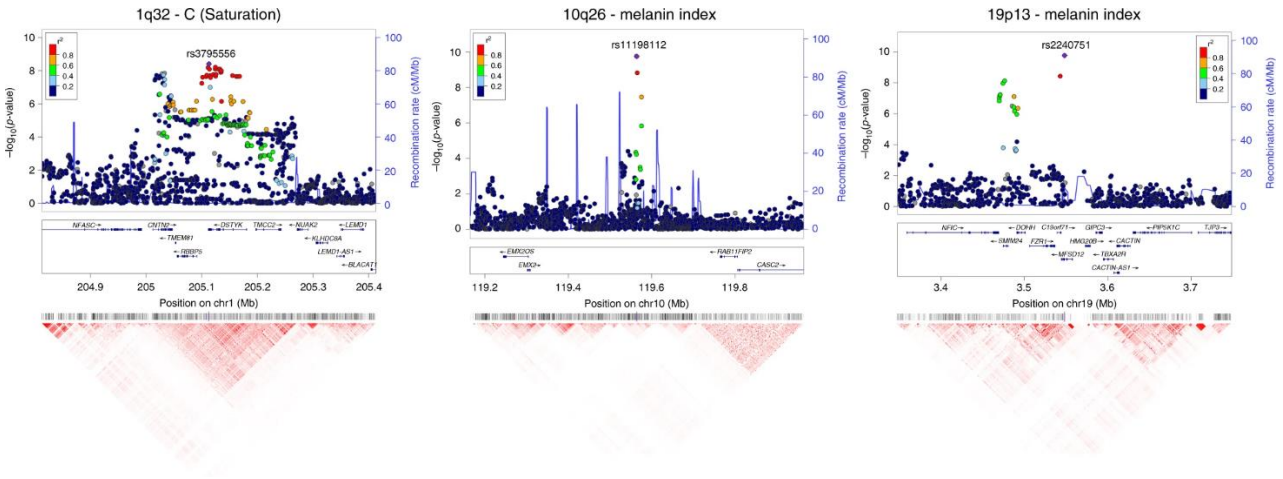


Figure 1.4. Genome-wide significant associations to pigmentation traits in Latin American populations. For all the plots, the x-axis reports the genomic location of the considered SNVs, while the y-axis on its left side shows their association P-value. The top of the plots indicates the traits for which the association was tested (i.e. skin pigmentation measured using reflectometry to evaluate the melanin index and skin saturation). In each chromosomal interval, index SNVs (i.e. genetic variants showing the highest P-value for the association to a given trait in a certain genomic region) are highlighted with a purple diamond. The colours for other SNVs represent the strength of LD between that variant and the index SNV. Local recombination rate between SNVs is shown as a continuous blue line (i.e. right y-axis). Genes in each region, as well as their intron/exon structure, are shown in the middle of each panel. At the bottom, a pairwise LD heatmap across all SNVs in a region is displayed (i.e. with red colours indicating $r^2 = 1$ and white colours indicating $r^2 = 0$). The figure has been adapted from Adhikari et al. (2019).

As briefly mentioned before, GWAS are based on the CD-CV hypothesis, which posits that the relatively high prevalence of complex common diseases is due to the high frequency of the related risk variants in the population (Schork et al. 2009). This comports that disease susceptibility resulted from the conjunctive action of several common variants, and that, unrelated affected individuals share a significant proportion of such risk alleles (Schork et al. 2009). Therefore, GWAS should be capable of identifying genetic markers at high frequency in cases and that significantly differ in terms of frequency with respect to control groups. Accordingly, adopting the GWAS framework to investigate the evolution of polygenic adaptations could represent a strategy capable of detecting common variants (i.e. SNVs putatively favoured by the action of natural selection) among populations known to be biologically adapted to a given environment (e.g. high-altitude human groups from the Himalayas) and/or characterized by known adaptive traits (e.g. light-skin pigmentation in non-African populations). Nevertheless, the validity of the common disease-common variant hypothesis in fully explaining the genetic susceptibility to complex diseases has been substantially criticized (Wang et al. 2005; Schork et al. 2009). Therefore, applying the GWAS framework to the study of polygenic adaptations introduces several limitations and potential conceptual biases in the reconstruction of the adaptive history of human populations.

Specifically, one of the major limitations of GWAS-based approaches is the need to adopt a high level of statistical significance to account for the bias related to multiple testing procedures (Tam et al. 2019). Usually, this is accomplished thought the Bonferroni correction to maintain the genome-wide false-positive rate at 5%,

which however results in a reduced power to detect all the possible associations (Tam et al. 2019). Moreover, focusing only on highly significant associations means that only adaptive SNVs with a large effect on the considered trait, and thus having been targeted by strong natural selection, could be detected. A simple way to overcome such limitation is to increase the sample size of both case and control cohorts (Tam et al. 2019). However, assembling large cohorts, particularly when composed of individuals with largely shared genetic backgrounds, is often challenging. This, in fact, leads to another important concern in GWAS, which is represented by the potential for spurious associations arising due to population stratification rather than according to true genetic risk differences between cases and controls (McClellan and King 2010; Tam et al. 2019; Barghi et al. 2020). In fact, such false-positive signals often result from major differences in terms of ancestry between and/or within the studied cohorts (McClellan and King 2010; Tam et al. 2019; Barghi et al. 2020).

As concerning the conceptual limitations, GWAS identify variants that affect a trait irrespective of whether they are beneficial or deleterious (Barghi et al. 2020). Moreover, to result significantly associated to a trait, genetic variants need to present great shifts in frequency in the cases with respect to the controls, thus not likely resembling the nature of alleles that are expected to play a relevant role in the evolution of polygenic adaptations (Scheinfeldt and Tishkoff 2013). Finally, GWAS are not free from *a priori* assumptions since, in order to be tested, a complex trait has to be firstly defined and characterized in the studied cohorts of individuals, thus imposing important constraints to the range of adaptive traits that could be potentially investigated.

Another type of statistical approaches which is, however, strictly related to GWAS, is that based on the calculation of polygenic scores (PGS) for a complex trait. PGS is the results of a weighted measure according to which all the SNVs found to be associated to a given trait according to GWAS results, even though they do not pass the stringent significance threshold corrected for multiple testing, are considered (Barghi et al. 2020; Lewis and Vassos 2020). Specifically, unlike GWAS, PGS-based studies do not treat each SNV equally, but they weight each variant according to the amount of phenotypic variance it explains (Barghi et al. 2020; Ma and Zhou 2021). This approach can potentially account for studying a larger fraction of phenotypic variation that is often missed when considering only GWAS significant markers (Barghi et al. 2020; Ma and Zhou 2021). Nevertheless, given the strong dependence of polygenic scores on GWAS results, they are basically characterized by the same conceptual limitations (Lewis and Vassos 2020; Ma and Zhou 2021).

An alternative, and quite independent, strategy to investigate the genetic bases of polygenic adaptations is instead constituted by statical approaches aimed at detecting networks of genes significantly enriched in signatures ascribable to the action of natural selection (Figure 1.5) (Daub et al. 2013; Gouy et al. 2017). Remarkably, such methods do not require large sample sizes to obtain significant results, thus avoiding the need for adopting stringent P-value thresholds and the bias potentially introduced by population stratification. However, these approaches present two main limitations. Firstly, the lack of a direct link between putative adaptive loci and phenotypes comports some uncertainty about the nature of the selected traits (Barghi et al. 2020). Secondly, there is ambiguity in classifying the detected signatures as a single selective event acting on

a set of genes regulating a specific trait, as assumed under the model of polygenic adaptation, or as multiple independent sweep-like selective events acting on a set of correlated traits (Barghi et al. 2020). Despite these limitations, the genomic signatures detected by these methods represent some of the best proxies available of selective events occurred under a reliable approximation of a model of polygenic adaptation and are notably less affected by false-positive associations caused by the genetic substructure of the investigated populations (Barghi et al. 2020).

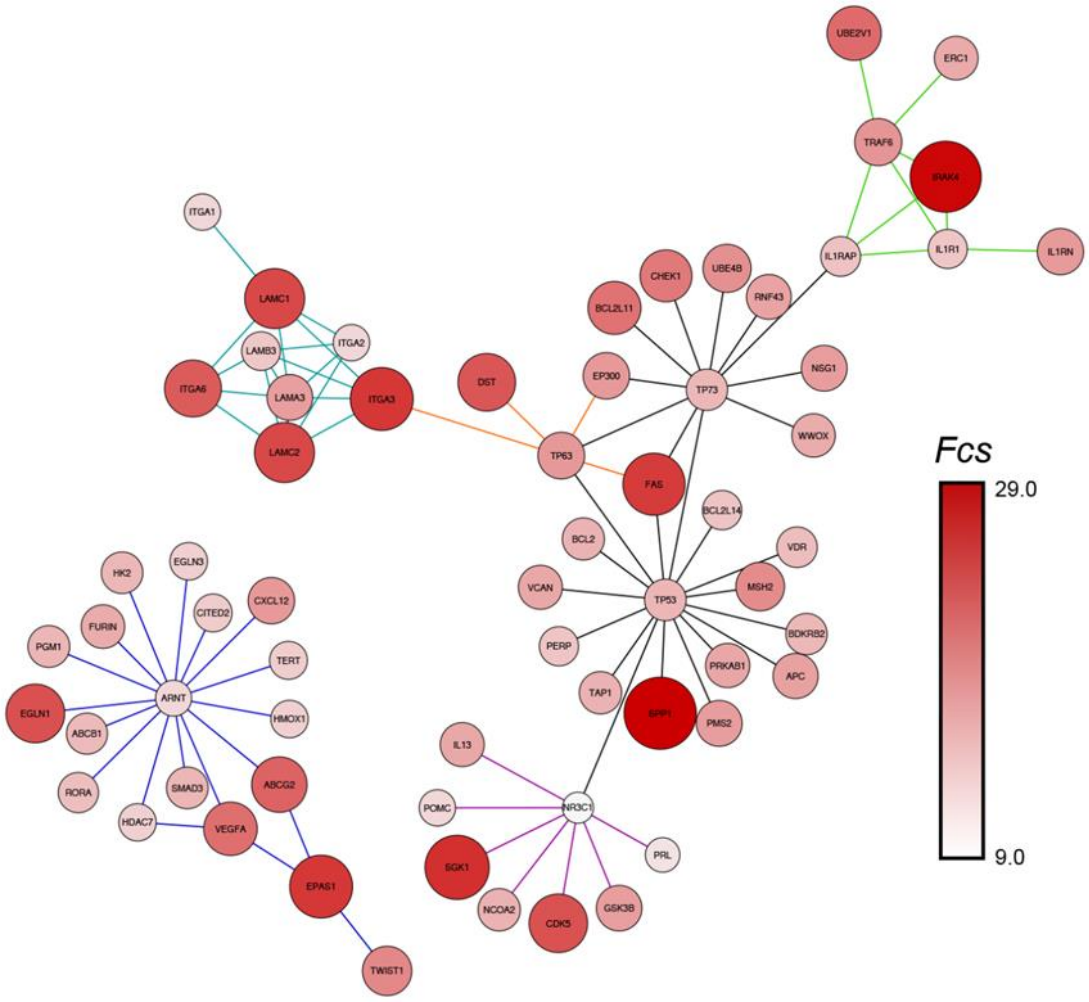


Figure 1.5. Networks of genes showing selection signatures likely involved in the evolution of polygenic adaptations in high-altitude Sherpa populations. The reported networks resulted significantly impacted by the action of natural selection and were identified at first by calculating the Fisher combined score (Fcs) to account for results from different selection scans (Deschamps et al. 2016) and secondly by using genome-wide distribution of this score to inform the *signet* network-based method (Gouy et al. 2017). Such a combined approach was aimed at detecting genomic signatures of polygenic adaptations. Genes participating to the *Hypoxia inducible factor* (HIF) 1α and 2α pathways are connected by blue lines. Loci within the pro-angiogenetic *integrin* pathways are connected by aquamarine lines. Genes belonging to the *Glucocorticoid receptor regulatory* and to the *Interleukin-1* signalling pathways are connected by purple and green lines, respectively. Members of the *P63* pathway are connected by orange lines, while those belonging to both *P53* and *P73* regulatory pathways are connected by black lines. Both the colour intensity and the size of the circles are proportional to the Fcs score calculated for each gene, which specifically ranged from ~9.0 to ~29.0. The figure has been taken from Gnecchi-Ruscone et al. (2018).

1.4 Inbreeding with archaic hominins: the role of adaptive introgression in shaping biological adaptations of modern humans

Admixture with extinct hominin species constitutes another important evolutionary aspect that has potentially influenced patterns of adaptive variability (included those of complex adaptive traits) of modern human populations (Sachdeva and Barton 2018; Barghi et al. 2020). In fact, such a process has been proposed to affect loci heterogeneity within the framework of polygenic adaptations, particularly resulting in large heterogeneous allele frequency shifts of introgressed haplotypes (Sachdeva and Barton 2018).

To date, a variety of studies have identified traces of inbreeding events between our species, Neanderthals, and Denisovans by analysing whole genome sequence/genome-wide data from modern human populations and/or by comparing modern and archaic genomes (Green et al. 2010; Marnetto and Huerta-Sánchez 2017; Bergström et al. 2020). Notably, some of these putative introgressed genomic segments were also proved to have been subjected to the action of natural selection in present-day populations, thereby illustrating events of adaptive introgression (Huerta-Sánchez et al. 2014; Gittelman et al. 2016; Racimo et al. 2017; Gouy and Excoffier 2020). Particularly, such findings concern genes linked to high-altitude adaptations (Huerta-Sánchez et al. 2014), immune-related loci, as well as genes regulating skin pigmentation (Gittelman et al. 2016).

Among the loci whose adaptive evolution was significantly influenced by archaic introgression events, *EPAS1* represents a cardinal example, having likely resulted from the admixture between the Denisovan species and the ancestors of present-day Tibetan highlanders (Figure 1.6) (Huerta-Sánchez et al. 2024). In fact, multiple independent studies have identified signatures of both archaic introgression (Huerta-Sánchez et al. 2014; Zhang et al. 2021) and positive selection (Beall 2007; Bigham et al. 2010; Yi et al. 2010; Peng et al. 2011; Xu et al. 2011) at the *EPAS1* genomic variation observed in high-altitude populations of Tibetan ancestry, with all signals appearing to be concentrated within a specific 32.7 kb region of the gene (Figure 1.6).

Nevertheless, these studies were again aimed at identifying classical hard selective sweeps, for which the putative adaptive variants/haplotypes are found in archaic hominin species as well, thus potentially identifying introduced genomic segments in the gene pool of *Homo sapiens* populations that have been then subjected to selective sweeps. However, these lines of evidence do not account at all for complex adaptive traits and thus explain just a limited portion of the adaptive introgression events evolved by our species, suggesting that additional biological functions and/or loci may have been affected by introgression and then involved in the evolution also of well-characterized adaptive phenotypes. In this context, few recent studies have developed sophisticated statistical approaches capable to test more realistic models of adaptive introgression, in which both admixture and selection events are evaluated simultaneously (Setter et al. 2020; Gower et al. 2021; Zhang et al. 2023). For instance, *MaLAdapt* (Zhang et al. 2023) and *VolcanoFinder* (Setter et al. 2020) statistics allow to test for the weak action of natural selection on putative introgressed standing genetic variation, thus allowing for the potential identification of the genetic determinants of polygenic adaptive introgression events. Therefore, the utilization of such approaches to investigate signatures of polygenic adaptive introgression in specific case-studies could enable to expand our knowledge concerning such a complex evolutionary process, which is, so far, still relatively understudied.

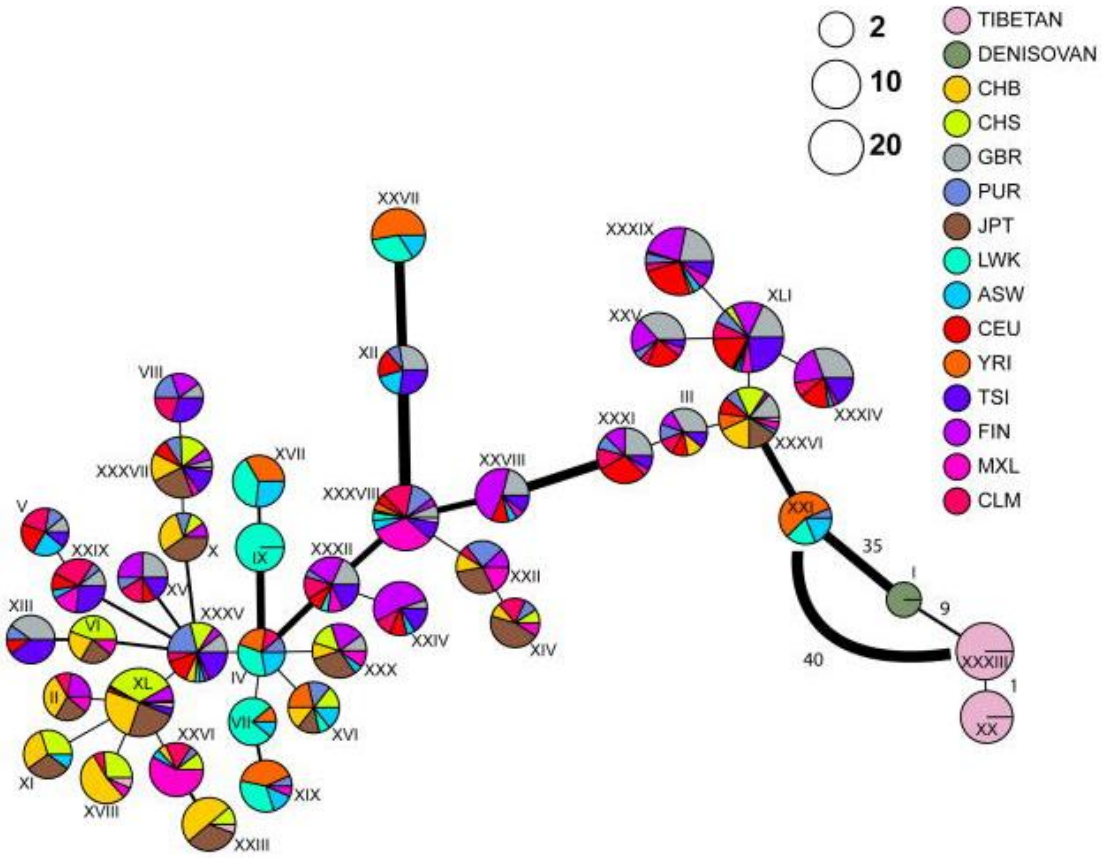


Figure 1.6. Haplotype structure at the *EPAS1* putative adaptive introgressed genomic region in high-altitude Tibetan and worldwide human populations. The haplotypes were defined from all the SNVs present in both the sequenced 1000 Genomes and Tibetan samples within a specific region (i.e. spanning 32.7kb) of the *EPAS1* gene, which was pinpointed as carrying candidate adaptive variants in high-altitude Tibetans that are also observed in the Denisovan genome. The Denisovan haplotypes (forest green) were added to the set of modern haplotypes in order to evaluate the number of pairwise differences among them. Each pie chart represents one unique haplotype, labelled with Roman numbers, and the radius of the pie chart is proportional to the frequency of that haplotype. The sections in the pie provide the breakdown of the haplotype representation among populations. The width of the edges connecting the pie charts is proportional to the number of pairwise differences between the joined haplotypes. The legend shows all the possible haplotypes among the analysed populations (see Auton et al. 2010 for label information). The numbers next to the edge in the bottom right are the number of pairwise differences between the corresponding haplotypes. The figure also shows an edge connecting the Tibetan haplotype XXXIII and its closest non-Denisovan haplotype (XXI) to indicate its divergence from the other modern human groups. The figure has been taken from Huerta-Sánchez et al. (2014).

2. AIMS OF THE STUDY

During its evolutionary history, populations of the *Homo sapiens* species progressively occupied a variety of geographical areas, thus having to survive and reproduce in a multitude of environments characterized by remarkably different selective pressures, such as endemic pathogens, specific food sources and extreme climatic conditions (Fumagalli et al. 2011; Hancock et al. 2011; Ojeda-Granados et al. 2022). Polygenic adaptations (rather than traditional models describing the action of natural selection) are increasingly supposed to represent the main mechanisms having mediated the evolution of adaptive traits in the ancestors of modern human populations, enabling their rapid biological adaptation to such a diversified range of environments (Pritchard and Di Rienzo 2010; Barghi et al. 2020). However, evidence concerning this type of adaptations are still limited and mainly results from genome-wide association studies, which can be underpowered in cohorts with reduced sample size and/or biased by internal genetic substructure (i.e. population stratification) and/or by incorrect matching between case and control groups of appreciably different ancestry (Tam et al. 2019). Such a scarcity of evidence is also due to the lack of proper statistical methodologies aimed at testing a realistic approximation of the model of polygenic adaptation. Concomitantly, inbreeding with archaic hominin species was proved to constitute an important evolutionary process having influenced the adaptive variability of modern human populations (Laurent et al. 2011; Vernot and Akey 2014; Huerta-Sánchez et al. 2014; Prüfer et al. 2014; Sankararaman et al. 2014; Gittelman et al. 2016; Dannemann and Racimo 2018; Enard and Petrov 2018; Racimo et al. 2018), thus further complicating the reconstruction of a comprehensive picture able to pinpoint the genetic bases of complex adaptations evolved by our species.

To overcome these limitations, the present project primarily aimed at developing a pipeline of analyses capable of detecting genomic signatures left by the action of natural selection and potentially ascribable to the model of polygenic adaptation. Furthermore, the project attempted to test the reliability of such a methodological workflow on whole genome sequence (WGS) data belonging to several populations roughly representative of the world-wide human biodiversity. Briefly, rather than relying on a unique approach, the assembled pipeline was designed to result from the combination of three different methods (Gouy et al. 2017; Mughal and DeGiorgio 2019; Harris and DeGiorgio 2020), which are overall capable to identify genomic patterns ascribable to both strong and moderate-to-weak selective events. Moreover, the adopted approach was conceived to enable the evaluation of different aspects of the genomic variation resembling a pervasive action of natural selection on functionally related loci, thus conferring consistency to the results validated by all the steps of the pipeline.

In addition, the present project secondly aimed at investigating archaic adaptive introgression events, by examining them from a polygenic point of view (i.e. by focusing on the evolution of complex adaptive traits). For this purpose, two specific human groups were considered as valuable case studies: high-altitude populations of Tibetan ancestry and Yakut people from Northern Siberia. Particularly, the study of these populations was thought to be especially relevant in the contest of the investigation of adaptive introgression

mechanisms because: i) these human groups are still currently subjected to severe selective pressures (i.e. hypobaric hypoxia and sub-arctic climate, respectively) and ii) previous studies reported consistent evidence of gene flow occurred between the ancestors of these populations and the Denisovan species (Huerta-Sánchez et al. 2014; Bergström et al. 2020; Zhang et al. 2021). Based on these considerations, it is likely to hypothesize that at least a portion of their genomic variability impacted by archaic introgression was involved also in the evolution of complex adaptations to high altitude/extreme cold environments.

3. MATERIALS & METHODS

3.1 Dataset composition and curation

The publicly available high-coverage Whole Genome Sequence (WGS) data analysed in the present study were collected from different repositories and merged into single datasets by performing a series of Quality Check (QC) operations aimed at testing for data consistency and avoiding the introduction of potential bias due to low-quality variant calling.

In detail, the considered WGS data mainly included those generated in the studies by Auton et al. (2015) within the framework of the 1000 Genomes Project, by Bergström et al. (2020), which focused on the samples belonging to the Human Genome Diversity Panel (HGDP), and by Sazzini et al. (2020), which sequenced individuals from the Italian peninsula. Overall, these studies provided high-resolution genomic data representative of several human groups of Northern and Sub-Saharan African, Continental and Southern/Mediterranean European, as well as Central and Northern/Eastern Asian ancestries. These data also included very few samples belonging to Oceanian and non-admixed Native American populations. That being so, to improve the representativeness of this latter group of populations, we also took advantage from WGS data for indigenous populations from Mexico (Jiménez-Kaufmann et al. 2022), which have been made available thanks to the collaboration with Dr. Andrés Moreno Estrada, director of the Human Population Genomics Lab of the LANGEBIO - Centro de Investigación y de Estudios Avanzados (Irapuato, México), where I conducted my PhD period abroad. In particular, these WGS data completed those for samples of Native and admixed populations from Mexico, Peru, Colombia and Puerto Rico included in the HGDP and in 1000 Genomes Project panels.

An additional dataset made up of genome-wide (i.e. DNA microarray-based) data for indigenous populations from South America was also assembled in order to contextualize genomic variation of high-altitude Andean Aymara WGS generated by (Lindo et al. 2018) and to select low-altitude control populations to perform the analyses described in Ferraretti et al. (2025) (see Supplementary Table S1 reported in Ferraretti et al. 2025 for details). In particular, this study - whose results will be briefly discussed in the present thesis - was aimed at investigating the genetic bases of complex adaptive traits evolved by high-altitude Andean groups in response to the stress induced by hypobaric hypoxia and at identifying convergent adaptations shared with Himalayan populations. For this purpose, a further ad-hoc dataset suitable for the genetic contextualization of Tibetan WGS data retrieved from Jeong et al. (2018) was assembled. Such a dataset enabled to carry out the analyses described in Ferraretti et al. (2024) (see Supplementary Table S1 reported in Ferraretti et al. 2024 for details), which were aimed at disclosing the impact of introgression of Denisovan alleles in modulating the adaptive responses evolved by high-altitude populations from the Himalayas.

Overall, the obtained extended dataset included 3,430 WGS for individuals belonging to 73 world-wide human populations and was then subdivided to obtain different sub-datasets representative of the main geographical macro-areas considered and of specific case-studies, such as those focused on high-altitude groups from

Andean and Himalayan regions. Once assembled, all these sub-datasets were separately subjected again to QC procedures specifically designed for the types of analyses to be performed. In detail, we prepared the datasets to be subjected to genotype-based population structure analyses, such as Principal Components Analyses (PCA) and ADMIXTURE, by retaining those SNVs that: i) presented less than 5% of missing data, ii) respected the Hardy-Weinberg Equilibrium (HWE) according to the stringent p-value threshold obtained after Bonferroni correction for multiple testing, iii) presented a Minor Allele Frequency (MAF) greater than 1%, and iv) were in low Linkage Disequilibrium (LD) between each other (i.e. showing a r^2 score minor or equal to 0.2). We also excluded from subsequent analyses those individuals that presented more than 5% of missing data and identity-by-descent (IBD) kinship scores (PiHat) greater than 0.27 after estimation of the degree of recent shared ancestry for each pair of subjects based on identity-by-state (IBS) calculations, to remove closely related individuals to the second degree, as suggested for datasets including populations with relatively low effective population size and appreciable inbreeding levels (Ojeda-Granados et al. 2022).

In parallel, to carry out haplotype-based population structure analyses by means of the ChromoPainter and fineSTRUCTURE algorithms, we phased the following datasets: i) the extended merged dataset obtained by the combination of samples from the studies by Auton et al. (2015), Bergström et al. (2020) and Sazzini et al. (2020), and ii) the sub-dataset assembled specifically for the investigation of adaptive evolution of populations of Native American ancestry. For the phasing procedure, differently than for genotype-based analyses, were retained low-frequency variants (i.e. showing MAF $\leq$ 1%) and SNVs in LD (i.e. showing a r^2 score above 0.2). The extended merged dataset was subdivided in three sub-datasets to obtain phased copies of WGS data representative of African, European and North/East Asian geographical macro-areas. Schematic representation of these methodological steps is reported in Figure 3.1. The phasing procedure was performed using the SHAPEIT software v2.r904 (Delaneau et al. 2013), as well as HapMap phase 3 recombination maps and the 1000 Genomes Project dataset as a reference panel (Auton et al. 2015).

Furthermore, in order to avoid loss of data (in terms of numbers of available SNVs) due to the described merging procedure, phased WGS data for filtered individuals belonging to the same population and who resulted genetically homogeneous according to haplotype-based structure analyses were directly extracted from their respective original datasets and then submitted to subsequent analyses (e.g. SMC++ and selection scans). On these data we finally applied the following QCs to remove SNVs: i) with more than 5% of missing data, ii) showing significant deviation from the HWE, iii) presenting ambiguous A/T or G/C alleles, iv) with mapping quality $\leq$ 20, and v) with values of sequence depth < 30.

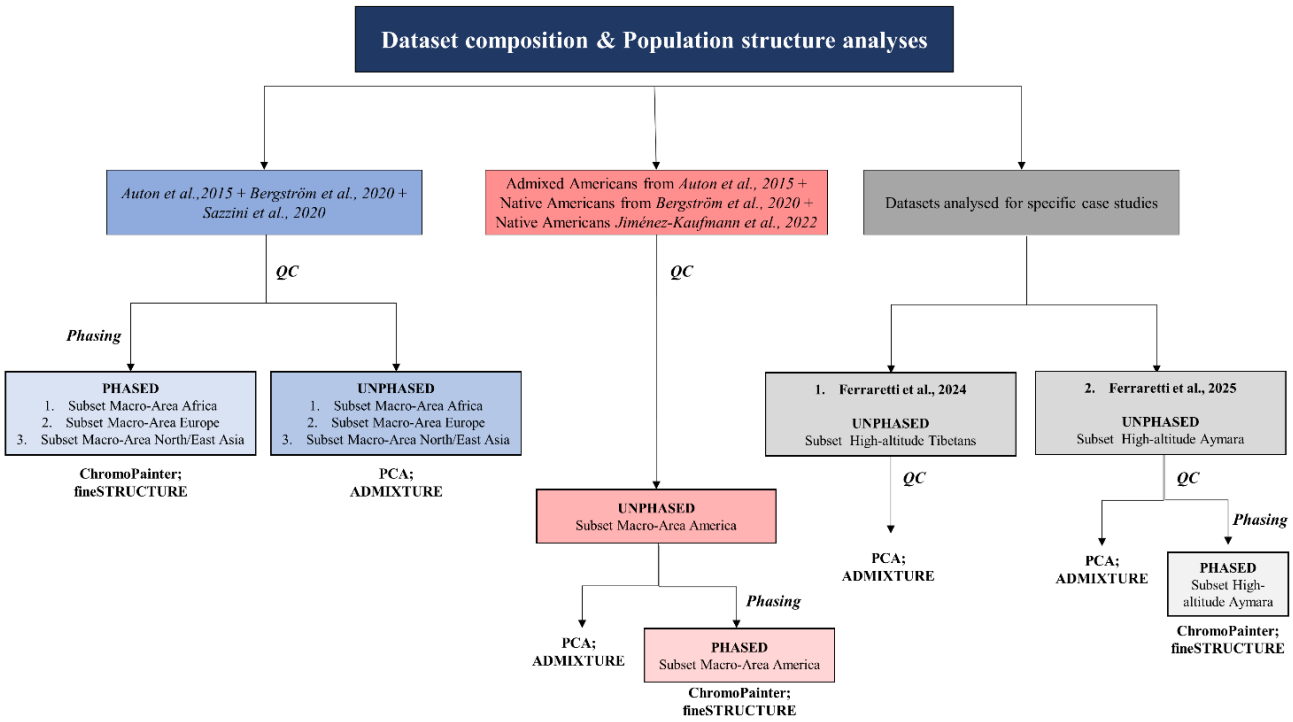


Figure 3.1. Overview of the composition of the datasets and of the population structure analyses performed. Scheme summarizing i) the reference studies used to collect WGS data, ii) the sub-datasets specifically assembled for each of the considered geographical macro-area and iii) the population structure analyses carried out on both the unphased and phased copies of the sub-datasets. From left to right: samples generated in the studies by Auton et al. (2015), Bergström et al. (2020), and Sazzini et al. (2020) were merged and subjected to QC procedures. Such extended dataset was then subdivided to obtain the unphased copies of the sub-datasets assembled for the African, European and North/East Asian geographical macro-areas, which were submitted to genotype-based population structure analyses. In parallel, the extended dataset was phased and subdivided to obtain the phased copies of the African, European and North/East Asian datasets, on which ChromoPainter and fineSTRUCTURE fine-scale population structure analyses were performed. Admixed American samples from Auton et. al. (2015), as well as WGS data for Native American populations from Bergström et al. (2020), were merged with genomic data generated by Jiménez-Kaufmann et al. (2022), which contained 50 WGS for individuals of indigenous un-admixed Mexican populations. Such a dataset was subjected to QC, PCA and ADMIXTURE analyses, as well as to phasing procedures and haplotype-based structure analyses. Finally, the datasets on which the studies by Ferraretti et al. (2024) and Ferraretti et al. (2025) were conducted were used to contextualize the genomic variation of Tibetan (Jeong et al. 2018) and Aymara (Lindo et al. 2018) high-altitude populations.

3.2 Genotype- and haplotyped-based population structure analyses

The unphased datasets were subjected to genotype-based PCA and ADMIXTURE analyses to check for data consistency after the merging procedure and to investigate broad patterns of population genetic structure. Specifically, PCA was carried out by employing the *smartpca* function implemented in the EIGENSOFT package (Patterson et al. 2006), while ADMIXTURE analysis was run using the ADMIXTURE software version 1.3.0 (Alexander et al. 2009) by considering a progressively increasing number of hypothetic ancestry

components (Ks) ranging from 2 to 13 and by choosing the configuration that best fits to the data based on the lowest 5-fold cross-validation (CV) error associated to each K.

Haplotype-based ChromoPainter and fineSTRUCTURE analyses were then performed to explore patterns of fine scale population structure and to identify genetically homogeneous populations to be submitted to downstream analyses. In detail, the ChromoPainter pipeline consists of two main steps (i.e. *estimate* and *run* phases). As concerns the *estimate* step, the phased datasets were used to infer the mutation rate (mu) and the effective population size (Ne) parameters by using the *–in* and *–iM* flags implemented in the ChromoPainterv2 software (Lawson et al. 2012) based on a subset of four representative chromosomes (i.e. 4, 10, 15, and 22) and five representative individuals for each population. Subsequentially, during the *run* step, the average Ne and mu values calculated for each individual and for each chromosome were used to process the data of all the 22 chromosomes included in each phased dataset.

More in detail, for processing the datasets of African, European and North/East Asian macro-areas the average mu and Ne values estimated on the extended merged dataset were used. On the other hand, for processing the Native American dataset *ad-hoc* average estimates were calculated on all the available samples, as well as on 10 individuals belonging to the Yoruba, CEPH and Han Chinese populations in order to account the African, European and East Asian ancestry components potentially present in the genomes of admixed Americans. Subsequentially, the *run* step was carried out on the same individuals, with each group in turn being considered as *donor* and *recipient* populations by using the *0 0 -a 0 0* option. The obtained ChromoPainter co-ancestry matrices were then submitted to the fineSTRUCTURE algorithm version 2.1.3 with the aim of identifying population clusters characterized by high internal genetic homogeneity by relying on haplotype sharing patterns between pairs of individuals.

In detail, the estimation of the fineSTRUCTURE algorithm were corrected on the base of the *c* value previously calculated with ChomoPainter. Moreover, the flags *-x*, *-y* and *-z* were respectively set at 3,000,000, 2,000,000 and 10,000 for the primary step of the analysis. Finally, for the *run* step required for the estimation of the tree topology, the *–x* and *–t* flags were set at 1,000,000 iterations.

3.3 Inferring effective population size fluctuations in genetically homogeneous groups

To infer Ne history in genetically homogeneous populations, their respective datasets were analysed with the SMC++ algorithm (Terhorst et al. 2017). Among the methods that rely on Sequentially Markovian Coalescent (SMC) approaches for estimating the demographic history of populations, SMC++ is indeed preferred to process large datasets since it is capable to analyse hundreds unphased genomes simultaneously (Terhorst et al. 2017). Initially, each genomic dataset was converted to the required format using the *vcf2smc* function by incorporating six different distinguished lineages following indications by Bergström et al. (2020).

Furthermore, the negative mask files relative to the 1000 Genomes and HGDP panels were used to inform the SMC++ algorithm about the location of chromosomal intervals that were not covered in the input files, thus avoiding the introduction of inferential bias due to fictitious long run of homozygosity.

The mask files were obtained by complementing the positive masks available at:

- https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/working/20160622_genome_mask_GRCh38/StrictMask/
- https://ngs.sanger.ac.uk/production/hgdp/hgdp_wgs.20190516/accessibility-mask/

Subsequently, the *estimate* function was applied to fit a population size history to each dataset, by considering a mutation rate of 1.25 x 10^{-8} per site per generation (Terhorst et al. 2017).

Finally, the analysis spanned a period between 1,000,000 and 1,000 years ago, corresponding to 34,000 and 34 generations respectively, assuming a generation time of 29 years (Bergström et al. 2020).

3.4 Integrating multiple methods to detect genomic signatures of adaptive evolution

To detect genomic signatures ascribable to a wide range of selective events, an integrated pipeline of analyses was assembled by relying on likelihood-, network-, and Machine Learning (ML)-based methods (Figure 3.2).

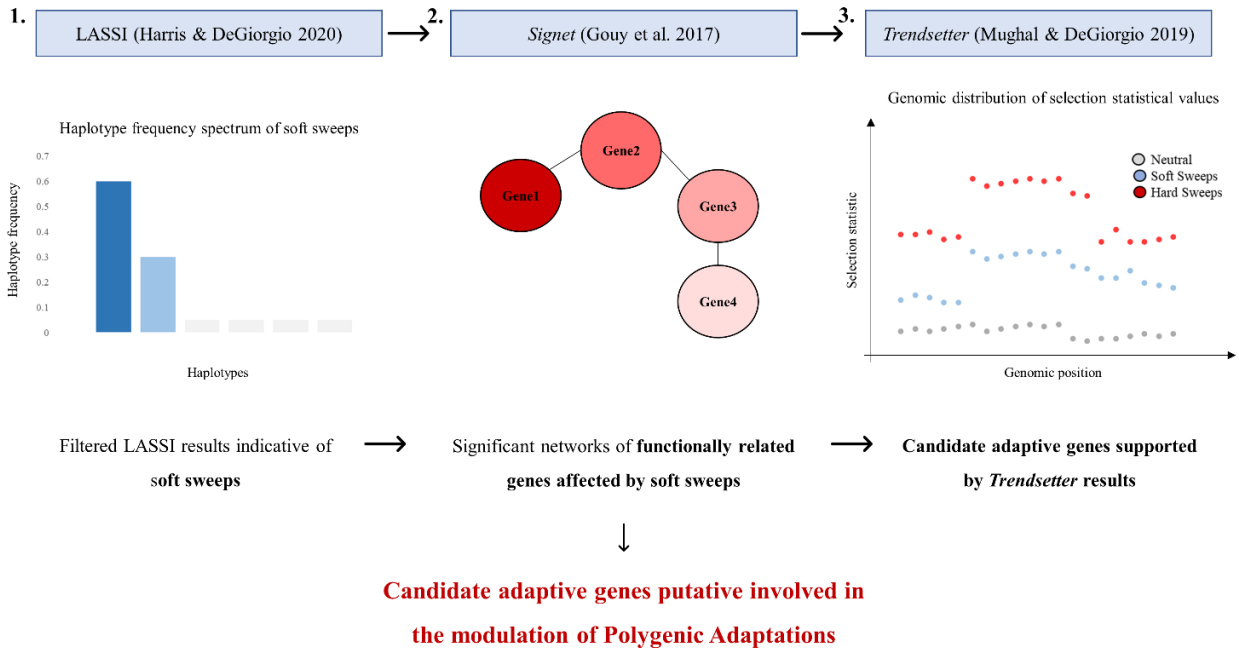


Figure 3.2. Integrated pipeline of analyses aimed at detecting genomic signatures ascribable to selective events under the polygenic adaptation model. The LASSI method developed by Harris & DeGiorgio (2020) was primarily run on phased data representative of each of the genetically homogeneous populations identified by population structure analyses. The obtained results were filtered in function of the *m* parameter in order to exclude likelihood values informative of genomic regions affected by hard selective sweeps and were subsequentially submitted to the *signet* analysis developed by Gouy et al. (2017). The *signet* method was thus used to point out biological functions modulated by groups of functionally related genes previously supposed to have been targeted by weak-to-moderate natural selection according to the LASSI analysis. The *Trendsetter* approach (Mughal and DeGiorgio 2019) was finally used to validate candidate adaptive genes pointed out by *signet* analysis or to detect additional adaptive loci contributing to the same/related biological functions. Overall, the described pipeline of analyses was assembled to pinpoint the possible genetic determinants of complex adaptive traits involved in the modulation of polygenic adaptations.

These approaches were combined to pinpoint the genetic determinants of biological traits plausibly impacted by the action of positive natural selection, including complex ones whose adaptive evolution was enabled by natural selection having simultaneously targeted many loci contributing to the modulation of the same biological function (Pritchard et al. 2010; Pritchard and Di Rienzo 2010; Barghi et al. 2020).
The results obtained have been also contextualized from a functional perspective, first through the implementation of STRING enrichment analyses and then through a meticulous literature search regarding the functional roles of single genes putatively involved in polygenic adaptations.

3.4.1 Likelihood Approach for Selective Sweep Inference - LASSI

Firstly, it was implemented the Likelihood Approach for Selective Sweep Inference (LASSI) developed by Harris & DeGiorgio (2020), which can distinguish between strong and weak footprints left by natural selection (i.e. hard and soft selective sweeps) on the base of the investigation of the haplotype frequency spectrum calculated for sliding genomic windows. This approach was applied on the phased data obtained for each of the identified genetically homogeneous populations. The schematic representation of haplotype frequency spectra evaluated by the LASSI statistic is reported in Figure 3.3.
Briefly, the rationale of the LASSI approach is the identification of the modification of the haplotype frequency spectrum of a given genomic region resulted after the action of natural selection on it (Harris and DeGiorgio 2020; Ferraretti et al. 2024). In detail, by considering the hard selective sweep model, according to which in the population arises a beneficial mutation with very strong impact on a given phenotypic trait, the expected haplotype frequency spectrum of such a genomic region will be characterized by a large decrease in haplotype variation. This determines that just a single haplotype carrying the putative adaptive variant and with an extremely elevated frequency is detectable in the population (i.e. sweeping haplotype) (Figure 3.3). Accordingly, the other haplotypes whether they exist, are expected to be found at very low frequency in the studied population. On the other hand, when considering the soft selective sweep model, which assumes that natural selection acts on a few new variants on a single gene and with low functional impact or on standing (i.e. pre-existing) genetic variation, again, at a single gene, the resulted haplotype frequency spectrum will be characterized by the existence of two or few more haplotypes that reach moderate frequency in the population (Harris and DeGiorgio 2020) (Figure 3.3).
In opposite, the haplotype frequency spectrum expected under neutrality presents multiple haplotypes at low frequency in the population. In particular, a single reference neutral spectrum can be obtained for each population by averaging the frequency values of each haplotype class obtained for the spectra calculated across the entire genome (Harris and DeGiorgio 2020) (Figure 3.3). Finally, through the comparisons between the neutral reference spectra, a likelihood statistic (i.e. LASSI *T* statistic) is calculated for each empiric spectrum, where positive values support the hypothesis of adaptive evolution for a given genomic window (Harris and DeGiorgio 2020).

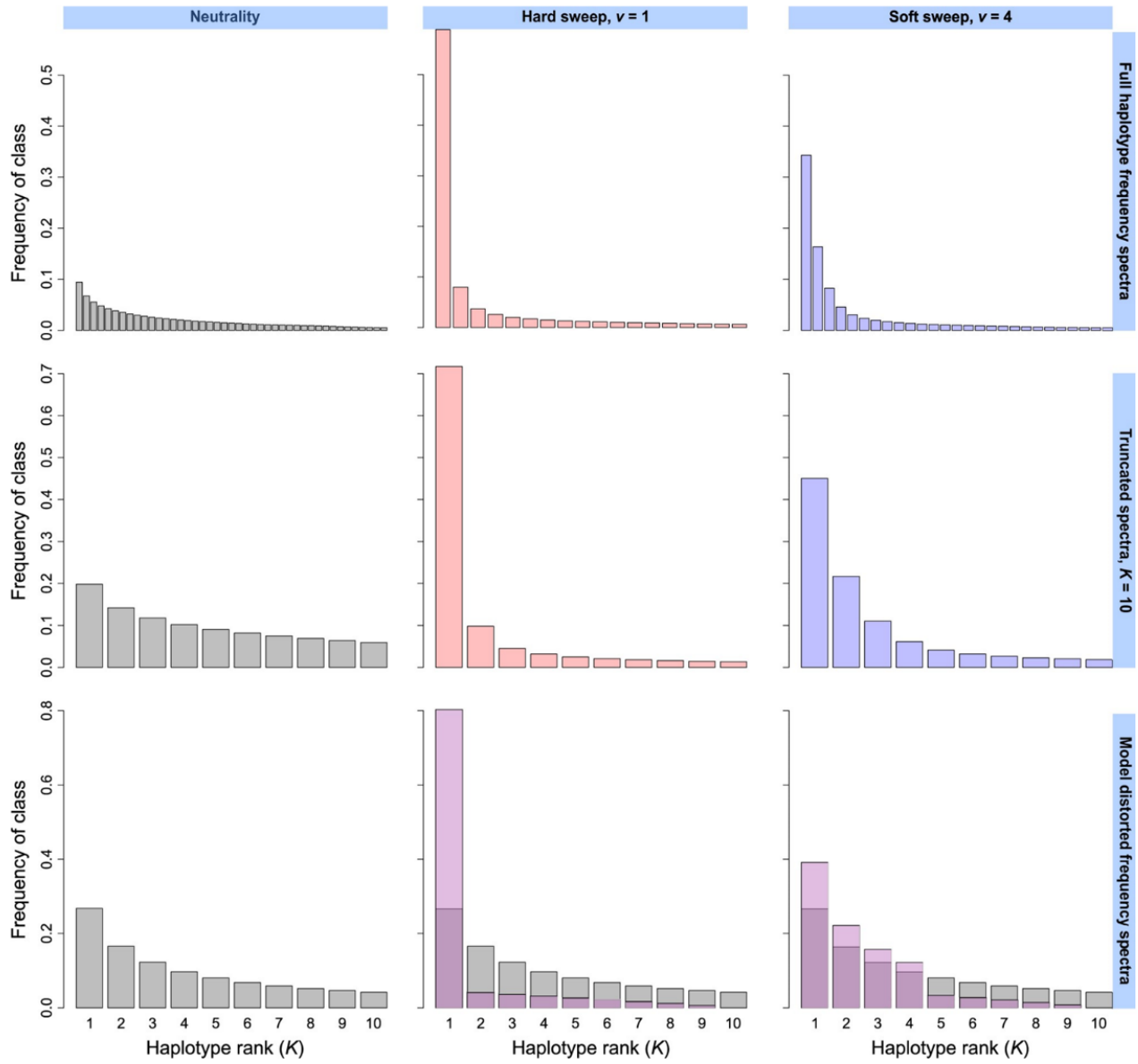


Figure 3.3. Schematic representation of the haplotype variation patterns tested by the LASSI approach. The figure reports simulated haplotype frequency spectra for genomic windows characterized by neutral evolution or adaptive evolution according to the hard sweep and soft sweep models. The number of simulated sweeping haplotypes (i.e. v) utilized for distinguishing between hard and soft sweeps is reported on the top of the figure. Under neutrality (i.e. top grey), the many sampled haplotypes inferred in the analysed genomic window exist at low frequency. In contrast, spectra associated to selective sweeps are characterized by high-frequency haplotypes and fewer total haplotypes (i.e. top light pink and lilac). The middle panel reports the truncated spectra generated in order to congruently compare the diverse haplotype spectra obtained along the genome. Overall, haplotype spectra preserve their shape relative to the untruncated spectra above. The bottom panel finally reports the distorted neutral spectrum computed from sampled haplotypes to yield spectra corresponding to alternative models (i.e. purple). In such spectra, the mass of non-sweeping classes is transferred to sweeping classes, resembling the expected pattern under a true selection event. The figure has been taken from Harris & DeGiorgio (2020).

The LASSI method was chosen among several statistical approaches because: i) it was demonstrated to have an improved power with respect to traditional selection scans (e.g. H12 and nSL) in the identification of selective events under a vast range of selection coefficients and demographic models (Harris and DeGiorgio 2020); ii) it can be used to easily distinguish hard selective sweeps from other selective events by filtering results in function of the *m* parameter (i.e. the number of inferred sweeping haplotypes that are supposed to carry adaptive variants, which is ≥ 2 for soft selective sweeps). Moreover, the LASSI method requires a certain number of variants to define the length of the windows for which the haplotype spectra are reconstructed. To avoid setting a random number for this user-defined parameter, the *--blocks* function implemented in the PLINK v1.90b5.2 package (Chang et al. 2015) was run on the unphased datasets to infer haplotype blocks and the number of SNVs included in each of them. This procedure was performed for each homogenous population and the average number of SNVs calculated by taking into account all the inferred haplotype blocks specific of each human group was used to divide the genome into the sliding windows considered by the LASSI algorithm.

The overall distribution obtained for the LASSI statistic was then filtered according to the *m* parameter (i.e. by considering $m \geq 2$) to exclude genomic windows affected by hard selective sweeps (by supposing they are not involved in the evolution of complex adaptive traits, whose identification was the main goal of the present study) and by retaining chromosomal intervals showing the greatest likelihood values associated to each gene. This filtering step indeed allowed to obtain a statistical distribution of values informative for signatures of weak-to-moderate natural selection, which constitute those more likely involved in the modulation of polygenic adaptations. Such distribution was then used to build the required input data for running the *signet* algorithm, which was chosen to investigate the potential interaction of multiple adaptive loci overall contributing to polygenic adaptations (Gouy et al. 2017).

3.4.2 *Signet* gene-networks reconstruction

The *signet* method is a statistical approach that can be used to detect genomic signatures ascribable to selective events occurred under the model of polygenic adaptation by pinpointing networks of genes contributing to the same functional pathway (i.e. likely modulating the same biological function) and whose variation patterns have been collectively shaped by the action of natural selection (Gouy et al. 2017) (Figure 3.4).

Thanks to the implementation of such an approach on the genomic distributions obtained by means of LASSI and indicative of weak-to-moderate selective events, significant *signet* networks have been considered as groups of genes simultaneously targeted by natural selection in the considered population, thus having potentially contributed to the evolution of complex adaptive traits.

In detail, the *signet* method crosschecks information contained in the input list (i.e. the greatest likelihood score computed for each gene identified by LASSI has putatively targeted by weak-to-moderate natural selection) with those collected in reference databases as regards the composition of annotated functional pathways, such as the Kyoto Encyclopaedia of Genes and Genomes (KEGG), using a simulated annealing algorithm approach

to define High Scoring Subnetworks (HSS) within each biological pathway (Gouy and Excoffier 2020; Ferraretti et al. 2025).

Basically, the HSS are networks made up of genes capable to maximize the *signet* output normalized score associated to each of these networks. To calculate such a score, the algorithm starts building random networks and iteratively modifies them by adding or removing one gene at a time, then recalculating the normalized score after each change, and repeating the process until it identifies the networks with the highest normalized scores. Finally, to test if scores of the estimated HSS are significantly larger than what would be expected by chance, a null distribution of HSS scores is generated and compared to the empirical *signet* values obtained for each network.

Gene networks showing p-values < 0.05 were retained as significant groups of candidate adaptive loci contributing to the same functional pathway according to information stored in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Furthermore, barely and/or not significant results (i.e. p-value < 0.06) were also briefly described in the Results/Discussion sections, but only when the same networks resulted significant for other populations showing related ancestry with that under investigation.

The *signet* algorithm was run using the R version 3.6.3 and by performing 20,000 iterations to generate the null distributions of the highest scoring subnetwork (Gnecchi-Ruscone et al. 2018; Ferraretti et al. 2025).

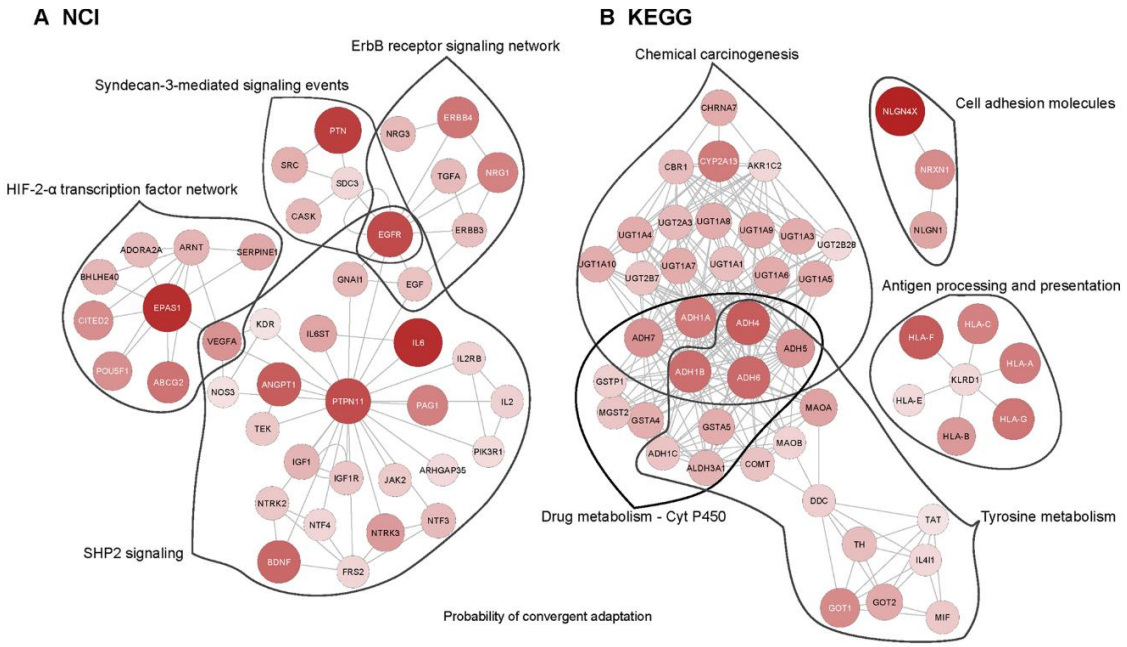


Figure 3.4. Standard representation of significant gene networks identified with the *signet* approach. Example of partially overlapped significant gene networks included in related functional pathways annotated in National Cancer Institute (NCI) (A) and KEGG (B) databases. The color intensity and size of the circles (i.e. gene/nodes) are proportional to the gene score (i.e. the statistical values associated to each gene in the input list). The figure has been adapted from Gouy et al. (2017).

3.4.3 *Trendsetter* supervised multinomial regression classifier

The last step of the implemented pipeline consisted in the utilization of the supervised multinomial regression classifier *Trendsetter*, which evaluates the distribution of diverse selection/summary statistics across genomic windows to predict signatures ascribable to both strong and weak natural selection (Mughal and DeGiorgio 2019) (Figure 3.5). In detail, these statistics comprehend r^2 (Hill and Robertson 1968) and π (Tajima 1983) that are capable to interrogate different aspects of genetic variation, such as patterns of LD decay among adjacent SNPs and alterations in the site frequency spectrum, respectively (Mughal and DeGiorgio 2019). Moreover, the *Trendsetter* method evaluates the genomic distributions of the number of haplotypes (Nhaps), as well as of the H1, H12, and H2/H1 statistics (Garud et al. 2015), which have been demonstrated to exhibit high power to detect both hard and soft selective sweeps and to collectively enable accurate distinction between them (Mughal and DeGiorgio 2019).

Such an approach was used to further validate at single gene level the selection signals identified with the combination of LASSI and *signet* methods and/or to point out additional putative adaptive loci which are proposed to participate to the same or related biological processes.

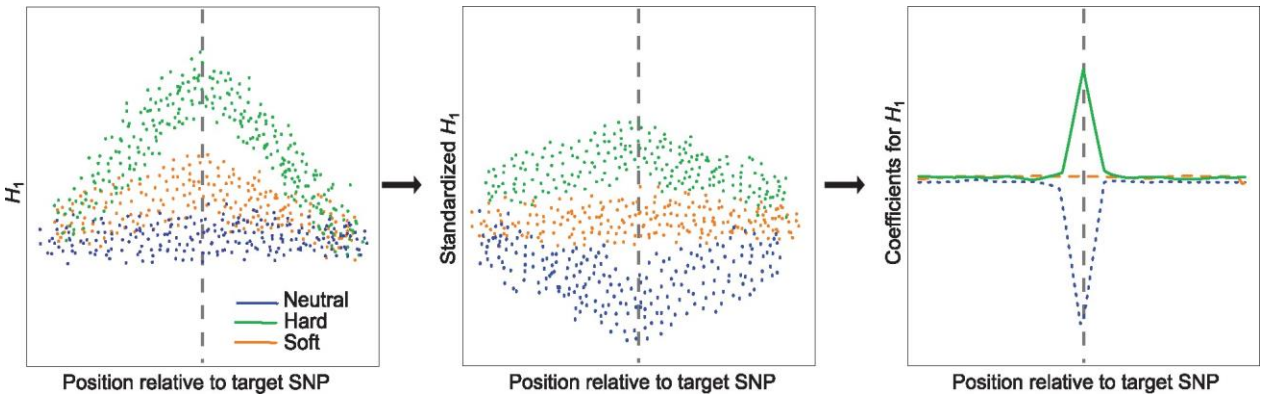


Figure 3.5. Schematic steps followed by *Trendsetter* method to train the machine-learning classifier. Each summary statistic (e.g. the expected haplotype homozygosity H1) is computed across genomic windows for a set of simulations representative of neutral, hard selective sweep, and soft selective sweep scenarios (left panel). The distributions of statistical values are then standardized at each genomic position, so that different summary statistics become comparable (middle panel). For instance, the standardized H1 yields strong negative values for simulations representative of neutral evolution and positive values for simulations informative of hard selective sweep occurred near a target SNP. Accordingly, simulations representative of soft selective sweeps exhibit intermediate values between the neutral and hard selective sweep scenarios. The model then performs trend filtering on the spatial distribution obtained for each summary statistic and for each class. This leads to the definition of curves describing the spatial distribution of summary statistics around a target SNP and for each evolutionary scenario (right panel). In such a case, the curve reported for the H1 statistic dramatically reduces for the neutral class near the centre of the genomic window, while is elevated near the target SNP for the hard selective sweep class, resembling the standardized distribution pattern. These curves are finally used to train the classifier algorithm. The figure has been adapted from Mughal & DeGiorgio (2019).

This methodological choice was primarily guided by evidence supporting enhanced power of the *Trendsetter* algorithm in the identification of selective events with respect to LASSI and traditional selection statistics (Harris and DeGiorgio 2020). Furthermore, such a method is capable to test more complex evolutionary scenarios with respect to most approaches used to investigate the action of natural selection because it considers also the *Ne* fluctuations of the population under investigation and the resulting modifications of genomic variation patterns due to population contraction/expansion when inferring adaptive processes. In fact, the *Trendsetter* pipeline consists of a training phase in which the algorithm learns how to distinguish the distributions assumed by selection/summary statistics by exploiting simulated genomic data. Particularly, simulated data incudes three classes of evolutionary scenarios: hard selective sweeps, soft selective sweeps and genomic regions with neutral evolution (see the following section for further details). Therefore, *Trendsetter* classifiers were trained using three sets of simulations, which were built for each population and for each possible predictable class, and comprising 5,000 simulations each (Mughal and DeGiorgio 2019). Moreover, three additional test sets were generated, each including 1,000 simulations for each class, in order to build calibration curves and to evaluate the accuracy of the trained classifiers by using the *calibration* and *metrics* packages implemented in the Scikit-learn Python library (Pedregosa et al. 2011). Finally, the trained algorithms were run on the real datasets assembled for the studied populations to classify genomic regions as affected by hard selective sweeps, soft selective sweeps or having undergone neutral evolution.

Due to the prolonged computational time required for this machine learning-based work-flow, the described integrated pipeline made up of LASSI, *signet* and *Trendsetter* approaches was implemented by performing also the *Trendsetter* validation step for human groups belonging to the Yoruba, Han Chinese, Tibetan, Yakut, Russian, Maya, Nahua and Pima populations, as well as for CEPH individuals, by considering these human groups as a reliable proxy of genomic variation observable at the diverse geographical macro-areas considered. The first two steps of the pipeline (i.e. LASSI + *signet*) were instead performed on all the genetically homogeneous populations/clusters identified by means of genotype- and haplotype-based structure analyses (see Figure 4.4 in the Results section). Furthermore, genomic windows presenting values of the LASSI statistic falling in the top 1% and/or top 5% of the likelihood distributions obtained for Tibetan, Maya and Sardinian populations were also explored. This was done because: 1) as concerns results pertaining to Tibetans, this further approach enabled to deepen finings previously published in the study by Ferraretti et al. (2024), and 2) to further investigate results potentially linked to selective pressures that are known to have acted specifically on Sardinian and Mayan populations, such as endemic malaria (Kosoy et al. 2011) and *Trypanosoma cruzi* infection (Gamboa-León et al. 2014).

3.5 Generating simulated datasets to train the *Trendsetter* classifiers

Simulated genomic data used to train the *Trendsetter* classifiers were generated with the software *discoal* (Kern and Schrider 2016) by considering *Ne* values estimated with the SMC++ method. In detail, for each possible evolutionary scenario, the number of haplotypes for each population (i.e. Yoruba, Han Chinese, Yakut, Russian

and CEPH) with a chosen length of 1.1 Mb were simulated (Mughal and DeGiorgio 2019). For human groups subjected to prolonged isolation and characterized by very low effective population size, such as Native Americans and Tibetans, the length of simulated haplotypes was augmented in order to obtain the minimum number of SNVs required by *Trendsetter* to analyse simulated genomic windows. In particular, the haplotype length was set at 3.3 Mb and at 2.2 Mb for Native American and Tibetan populations, respectively. Furthermore, according to Mughal and DeGiorgio (2019), the mutation rate (*theta*) was set at 1.25×10^{-8} per site per generation, and the population recombination rate (*rho*) was set at 1×10^{-8}. The values of both *theta* and *rho* were scaled by the length of the simulated haplotype and by the current effective population size (N_0). In addition to these basic parameters, all the simulated scenarios required population size changes to be specified. To do so, the *-en* flags were used to simulate instantaneous population size changes according to the obtained SMC++ results.

Generating simulations informative of soft and hard selective sweeps also demand the definition of specific parameters able to distinguish adaptive from neutral evolution. In detail, soft and hard selective sweeps were modelled as stochastic events, with beneficial mutations introduced at the centre of the simulated region and assuming they have been subjected to the action of natural selection with a per-generation selection coefficient (i.e. *s*) randomly drawn from a uniform distribution within the range [0.005 - 0.5] (Mughal and DeGiorgio 2019).

Furthermore, for all selection simulations, the timing of fixation of the adaptive allele tau (i.e. *T*) was randomly selected from a uniform distribution spanning from 0 (i.e. present) to 1,200 generations before the present for Han Chinese, Russian and CEPH populations, analogously to what was previously set for Eurasian human groups (Mughal and DeGiorgio 2019). In particular, such boundary should allow to detect biological adaptations developed by human populations after out of Africa migrations of *H. sapiens*.

For the Yoruba population, the lower limit of *T* was set by considering the split time estimation between West (i.e. Yoruba) and Central-East (i.e. Mbuti) African populations calculated in the study by (Bergström et al. 2020), which is equal to 69k years ago. The motivation that guided such a choice was to consider a plausibly interval of time indicative of the occupation of West Africa by modern human populations, allowing consequentially to identify complex biological adaptations specifically evolved in response to the selective pressures imposed by equatorial-type rainforest environments. According to such a rationale, for the Yakut, Native American and Tibetan populations the inferior tau limit was set respectively at 1,034, 507 and 172 generations in the past, which represent the periods indicative of the first traces of modern humans' presence in the Yakutia region (Pitulko et al. 2004), the American continent colonization by *H. sapiens* (Becerra-Valdivia and Higham 2020), and the human permanent settlement on the Tibetan Plateau (Li et al. 2019). Finally, to simulate datasets informative of soft selective sweeps from standing genetic variation, the initial frequency of the beneficial allele was randomly sampled from a uniform distribution within the interval [0.01 - 0.10] (Mughal and DeGiorgio 2019).

4. RESULTS

4.1 Composition of the datasets after QC procedures

The composition of each dataset in terms of included populations after the QC procedures implemented for each of the considered macro-areas is summarized in Supplementary Table S1. These subsets included samples characterized for the following numbers of SNVs, resulting from the merging of the overall dataset.

- African macro-area subset: 4,628,037 SNVs after QC performed prior to the phasing procedure and 477,038 SNVs after QC filtering for genotype-based population structure analyses.
- European macro-area subset: 5,112,378 SNVs after QC performed prior to the phasing procedure and 398,194 SNVs after QC filtering for genotype-based population structure analyses.
- North/East Asian macro-area subset: 5,056,113 SNVs after QC performed prior to the phasing procedure and 345,021 SNVs after QC filtering for genotype-based population structure analyses.
- American macro-area subset: 4,384,696 SNVs after QC performed prior to the phasing procedure and 294,443 SNVs after QC filters for genotype-based population cluster analyses.

For further details concerning the composition of the datasets including high-altitude populations of Tibetan and Aymara ancestries, please see Supplementary Table S1 from Ferraretti et al. (2024) and Supplementary Table S1 from Ferraretti et al. (2025), respectively.

4.2 Genotype- and haplotype-based population structure analyses

Genotype-based PCA was firstly performed separately on datasets assembled for each macro-area and allowed us to demonstrate the reliability of the performed merging procedures (in terms of representativeness of the data), as well as to preliminarily investigate the genomic variability of the human populations under investigation (Figure 4.1 and Supplementary Figure S1).

In fact, all the obtained PCA plots highlighted that individuals belonging to a given population clustered near to each other, forming in most of the cases easy-recognizable groups (Figure 4.1 and Supplementary Figure S1). Furthermore, human populations supposed to present largely shared evolutionary histories were found to be localized in proximity in the PCA genetic space (Figure 4.1 and Supplementary Figure S1).

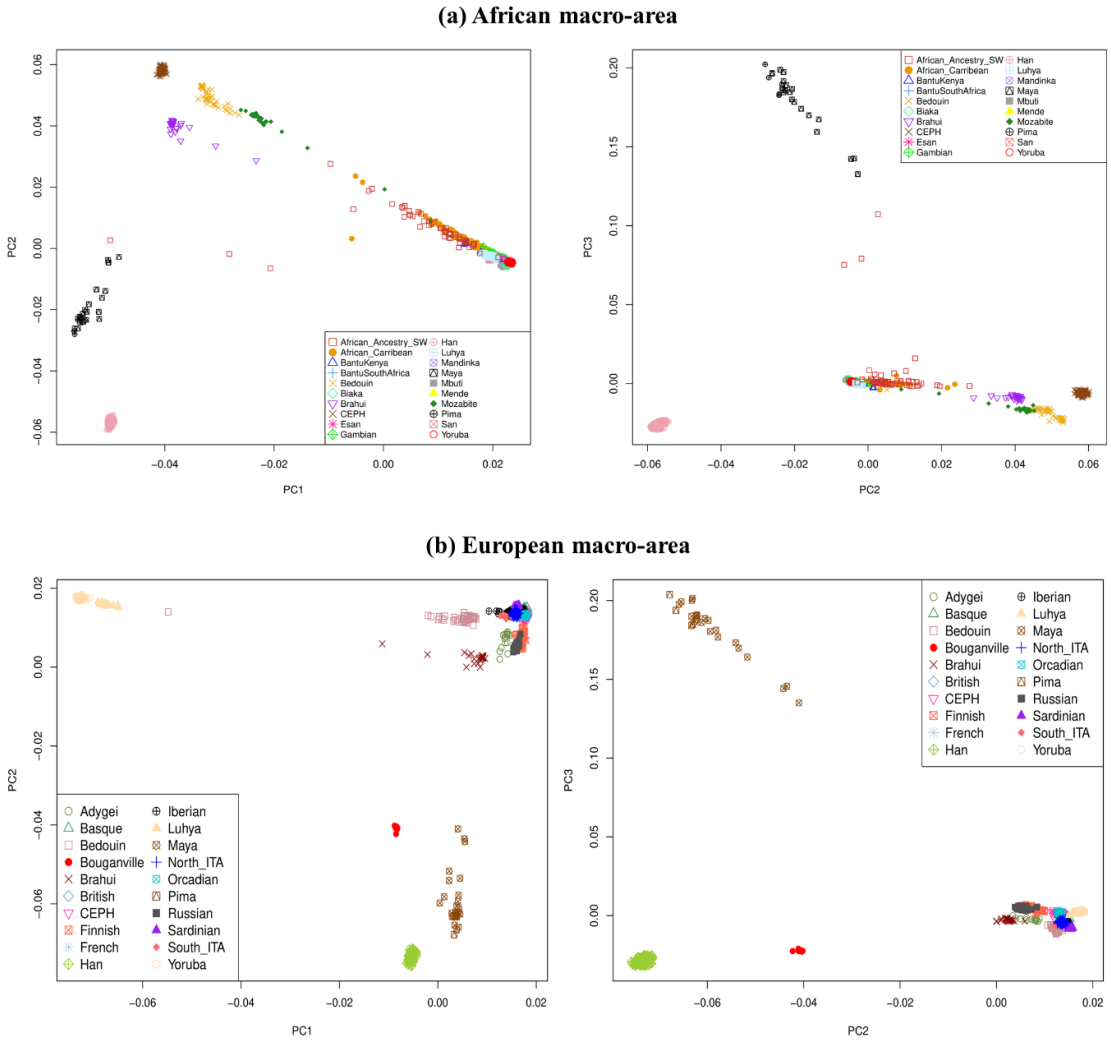


Figure 4.1. PCA plots obtained for (a) African and (b) European macro-areas. For all the plots in the figure, from left to right, are displayed the intersections between PC1 vs PC2 and PC2 vs PC3. **(a)** PC1 describes the differences between the investigated African populations and the "reference" non-African populations included in the dataset for the sake of comparison (i.e. from top to bottom CEPH, Brahui, Maya, Pima and Han Chinese human groups). In the PC1 vs PC2 plot, diverse groups of samples representative of different African populations are well recognizable. In particular, from left to right are displayed the Mozabite, Luhya, Mbuti, Easan, Baika, Gambian, and Yoruba populations. However, some individuals that can be defined as genetic outliers with respect to their population of origin are also visible (e.g. four samples belonging to the Mozabite population) since they appreciably deviated from the primary group of samples from that population. Accordingly, samples labelled as African_Ancestry_SW (i.e. red square) and African Caribbean (i.e. orange circle) included genomic data for individuals that cannot be assigned to a certain ethnic group, which indeed presented internal genetic substructure. PC2 summarizes the genetic differences among the non-African human groups, while PC3 describes the genetic isolation and low level of genomic variability of the Maya and Pima indigenous American populations. Finally, samples belonging to the Bedouin population were found to localize between CEPH and North African Mozabite populations. **(b)** PC1 illustrates the gradient of variation between the considered non-African populations and "reference" African groups included for the sake of comparison (i.e. Luhya and Yoruba). PC2 instead portrays the variability observed between European populations and both East Asian and Native American "reference" groups. On the top right of the plot are localized the European samples, which in order are those representative of Sardinian, Basque, Iberian, North Italian, South Italian, Orcadian, Finnish and Russian populations. Again, in this plot the Bedouins were found to cluster between African and European populations. Finally, PC3 depicts the variability observed between "reference" Native American groups and all the other populations considered.

In line with the picture emerged from PCA, ADMIXTURE analysis also showed the aggregation of single individuals belonging to the same population into genetically homogenous groups based on the inferred profiles of ancestry components (Figure 4.2 and Figure 4.3, Supplementary Figures S2, S3 and S4). The obtained cross-validation (CV) errors associated to each K are displayed in Supplementary Figure S5. For instance, Figure 4.2 and Figure 4.3 report two examples of the results obtained. In particular, Figure 4.2 displays the ADMIXTURE profiles of individuals included in the dataset representative of the North/East Asian macro-area. These results especially contributed to contextualize the genomic variability of the Yakut ethnic group, which has been extensively examined in the study by Ferraretti et al. (under review), which has been drafted during the last year of the PhD Programme and that was focused on the investigation of complex adaptations to extreme cold environments evolved by high-latitude Eurasian populations. The related manuscript is attached to the present thesis in the appendix section. Moreover, results obtained in the context of such a study were further detailed in section 4.5.3.3 of the thesis.

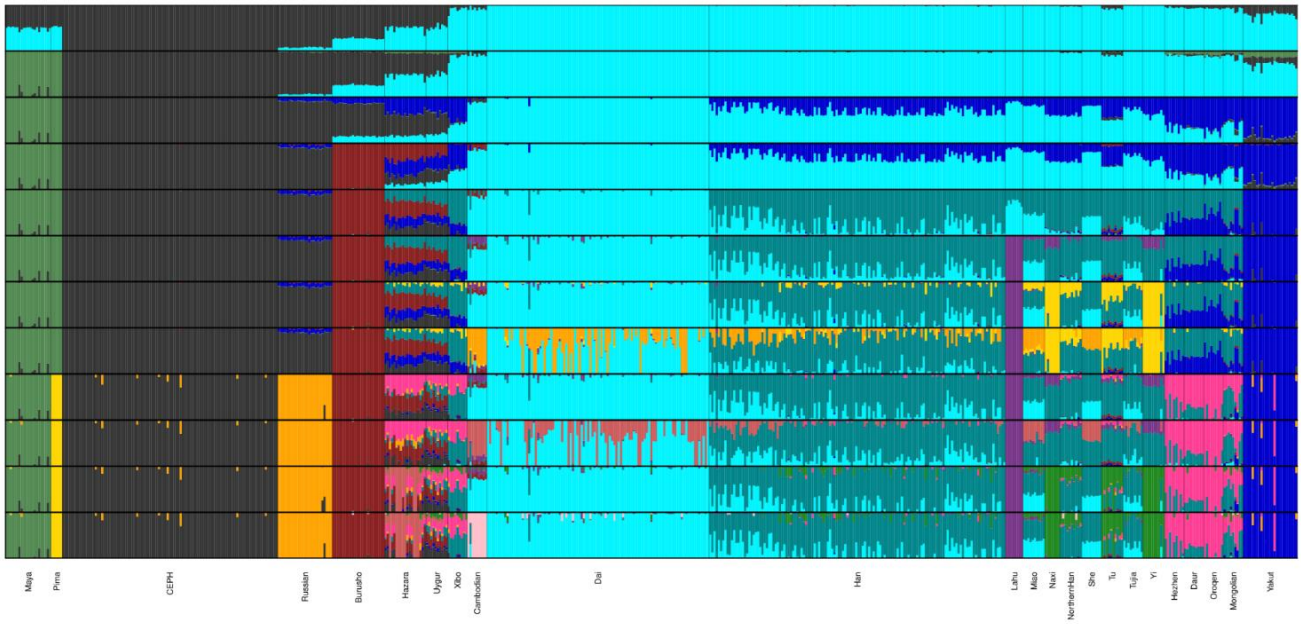


Figure 4.2. ADMIXTURE profiles inferred for the individuals representative of the North/East Asian macro-area. The plot shows ADMIXTURE estimates for the inferred ancestry components for K values ranging from 2 to 7. Each row corresponds to the number of Ks, while the coloured columns identify single individuals belonging to a certain population. The labels of each population are reported at the bottom of the plot. The K=4 configuration was identified as the optimal configuration for data of the North/East Asian macro-area based on the minimization of the associated CV error. Particularly, Yakut samples from North Siberia were found to be characterized by the predominant blue ancestry component, which reaches appreciable proportions also in most of the populations from East Asia (e.g. Xibo), Central Asia (e.g. Burusho), and Russia. The ancestry component reaching the greatest frequency in East Asian populations (i.e. light-blue) was instead detectable in minimal proportions in the Yakuts. Additionally, the typical European (i.e. grey) and Native American (i.e. green) components turned out to be recognizable in the Yakut ethnic group as well. Except for few outliers, ADMIXTURE profiles of single individuals turned out to be congruent with those of their specific populations, suggesting the overall genetic homogeneity of the human populations investigated.

Figure 4.3 instead displays another example of the results obtained by means of the ADMIXTURE analysis. In detail, the figure visualizes the averaged proportions of the ancestry components inferred for the best ADMIXTURE configuration (i.e. K=7) obtained for data representative of the American macro-area, especially for the considered Native American populations along with the geographic location of the performed sampling.

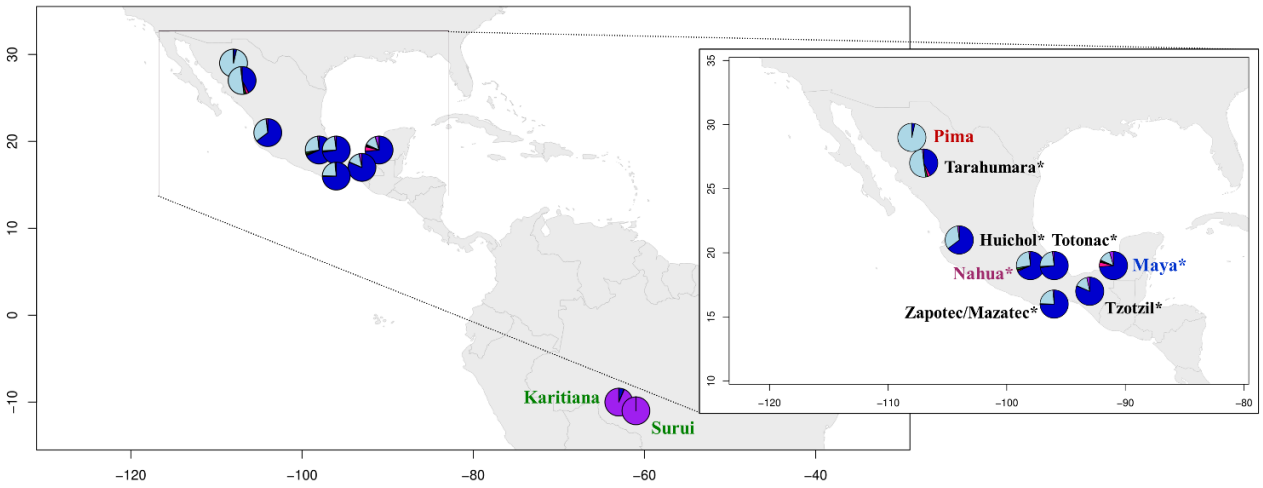


Figure 4.3. ADMIXTURE results for the individuals representative of the American macro-area. Map displaying the geographical location and the ADMIXTURE profiles (i.e. pie charts) of indigenous populations from Central and South America. Population samples labelled with the asterisk have been retrieved from the Mexican Biobank dataset. Specifically, the plot reports averaged ancestry proportions obtained for the best configuration (K=7) according to the minimum value of the CV error. The ADMIXTURE profiles clearly enable to distinguish the Mexican indigenous populations (which are overall characterized by greater proportions of the blue ancestry component) from those of South America, which present almost fixed proportions of the purple ancestry component, while the blue one is still recognizable at very low frequencies. The Maya (blue), Nahua (pink) and Pima (red) human groups constituted the populations presenting an adequate sample size and/or inbreeding levels to be further analysed. The Surui and Karitiana populations - which included individuals characterized by P-HAT greater than 0.27- were not included in the datasets subjected to the developed pipeline of analyses described in section 3.4 in order to avoid the introduction of bias in the obtained results due to elevated levels of inbreeding.

Finally, to assess the genetic homogeneity of each population with an even greater level of resolution, we investigated it by estimating the degree of shared haplotype chunks/blocks between pairs of individuals, as enabled by the Chromopainter and fineSTRUCTURE approaches (Supplementary Figures S6, S8, S9 and S10). Overall, according to these results we obtained a total of 24 filtered populations presenting a suitable sample size to be subjected to further analyses. Specifically, eight of these populations originated from Africa, eight from Europe, four from North and East Asia, and four from Central and South America (Figure 4.4, Supplementary Figures S6, S8, S9 and S10).

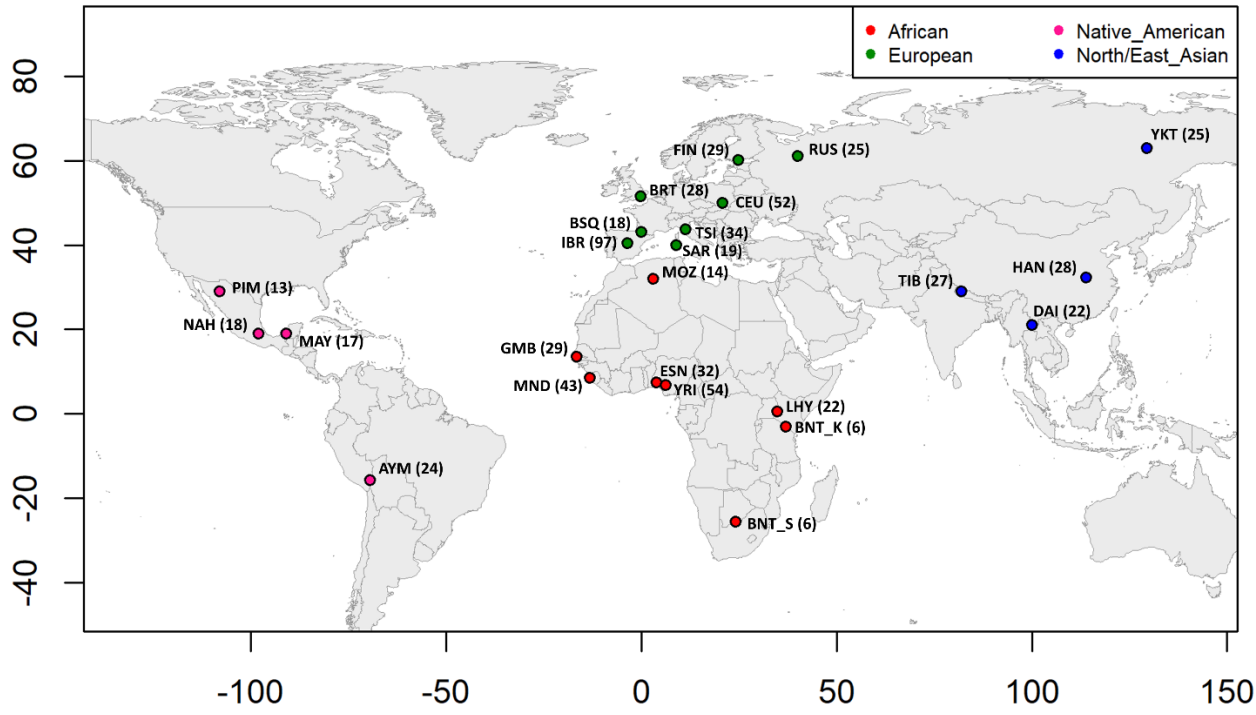


Figure 4.4. Map showing the geographic locations of the populations filtered according to the results from Chromopainter/fineSTRUCTURE clustering analysis. Labels: High-altitude Aymara (AYM); Bantu Kenya (BNT_K); Bantu South Africa (BNT_S); British (BRT); Basque (BSQ); CEPH individuals (CEU); Dai Chinese (DAI); Esan (ESN); Finnish (FIN); Gambian (GMB); Han Chinese (HAN); Iberian (IBR); Luhya (LHY); Maya (MAY); Mende (MND); Mozabite (MOZ); Nahua (NAH); Pima (PIM); Russian (RUS); Sardinian (SAR); Tibetans (TIB); Tuscan (TSI); Yakut (YKT); Yoruba (YRI). For the African macro-area (i.e. red) a total of eight populations were obtained of which one from North, four from West, two from East and one from South Africa. The individuals belonging to BNT_S and BNT_K populations where unified based on their shared ethnolinguistic origins and according to Chromopainter/fineSTRUCTURE results reported in Supplementary Figures S6 and S7. The dataset representative of the European macro-area includes eight populations: Northern Europeans (Finnish and Russian), the British, continental Europeans (CEU), North Italians (Tuscans), Sardinians, and Western Europeans (Basques and Iberians). The populations belonging to the North/East Asian and American macro-areas comprehend three human groups living in extreme environments (i.e. the Yakuts from sub-arctic climates, high-altitude Tibetans and the high-altitude Aymara population). These macro-areas also include two East Asian populations (i.e. Han and Dai Chinese) and three groups of indigenous populations from Mexico (i.e. Pima, Maya and Nahua), respectively.

Particularly, some of the groups finally analysed to infer their adaptive evolution resulted from the consolidation of different population clusters pointed out by ChromoPainter/fineSTRUCTURE analyses, which included individuals from the same population and/or from the same ethnolinguistic group (Figure 4.5 and Supplementary Figure S7). Specifically, haplotype variation within these populations was further investigated through PCA by assessing genetic differences based on the number of shared haplotype chunks between individuals. This analysis, based on ChromoPainter results, supported the merging of some Iberian, Tuscan and Bantu clusters into single analytical units (Figure 4.5 and Supplementary Figure S7).

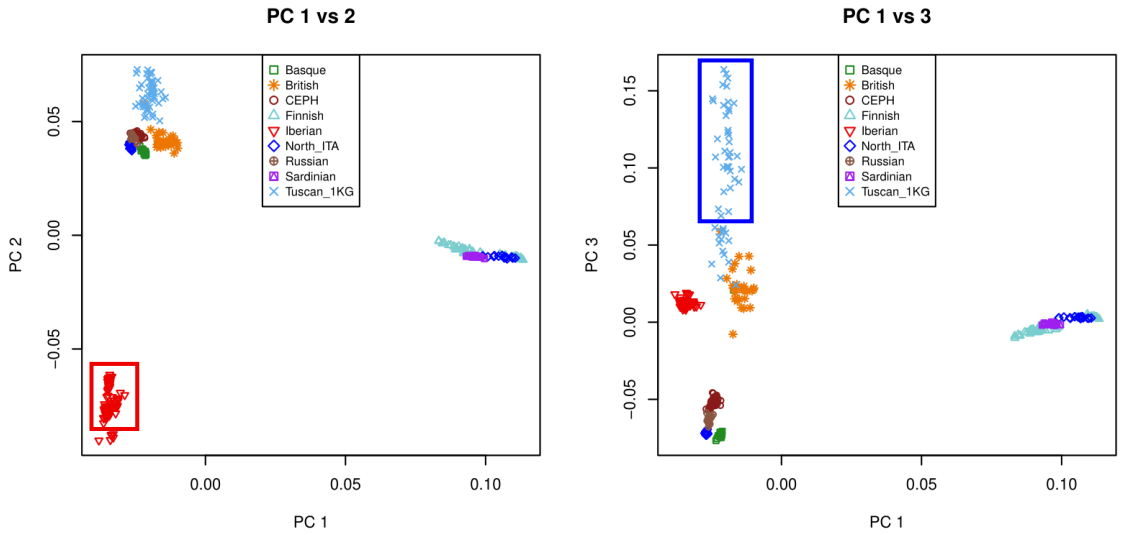


Figure 4.5. PCA evaluating the levels of shared haplotype chunks among European populations. PC1 visualizes the separation between Finnish, Sardinian and part of the North Italian samples (Sazzini et al. 2020) from the rest of the considered human groups. PC2 describes the differences among Iberians and the macro cluster localized at the top left of the plot, including British, Tuscan, North Italian, Russian, Basque and CEPH individuals. PC3 shows the gradient of variation observable across Tuscan individuals according to which 16 of these samples were found to localize near to the British population. The red and blue rectangles highlight the final set of individuals included in the Iberian and Tuscan clusters, respectively, which have been retained according to these results and that have been reported in Figure 4.4.

Furthermore, the clusters obtained for both Han and Dai Chinese populations, which included genomic data generated in the context of two different reference studies (Auton et al. 2015; Bergström et al. 2020), were further filtered in order to avoid the introduction of potential bias in results from SMC++ analysis due to the improper utilization of specific defined mask files. This procedure resulted in the identification of the final Han and Dai Chinese filtered groups as including 33 and 22 samples, which have been respectively retrieved from the HGDP and 1000 Genomes panels (Figure 4.4). Collectively, the majority of the datasets reported in Figure 4.4 were subjected to SMC++ analyses (by using the unphased copy of the dataset when available), while all of them were submitted to the LASSI-*signet* pipeline of analyses. More in detail, the SMC++ method, which constitutes a primary step required for the fulfilment of both *discoal* and *Trendsetter* pipeline, was performed on the populations reported in Figure 4.6 because of the long computational times required for the overall pipeline and especially for the generation of population-specific genomic simulations. In fact, to evaluate *Trendsetter*'s reliability across human populations with diverse ancestries, while also respecting the time constraints of the PhD, we applied this pipeline to at least one representative population from each geographic macro-area (Table 1) rather than on all the available populations. The number of retained SNVs in both the phased and unphased copies of the datasets obtained for each filtered population is listed in Table1.

Table 1. SNVs retained after QC procedures for the datasets assembled for each filtered population. Columns: **(1)** Population analysed; the number of SNVs included in **(2)** phased and **(3)** unphased datasets after QC filtering procedures. The number of SNVs included in the unphased copy of the datasets is not reported for the WGS generated in the study by Auton et al. (2015) since the authors released only the phased copy of the complete reference panel. This information is not described also for high-altitude Aymara samples since the corresponding unphased dataset was not utilized for genotype-based population structure analyses and neither for implementing the SMC++ method. Genetically homogenous populations belonging to Native Mexican human groups (i.e. Maya, Nahua and Pima) are characterized by the same number of SNVs since these subsets were extracted from both the phased and unphased merged datasets representative of the American macro-area. The populations reported in bold are those subjected to the *Trendsetter* pipeline while all the human groups listed were submitted to LASSI-*signet* analyses. Particularly, the phased copies of the datasets were used to perform LASSI-*signet* as well as *Trendsetter* pipelines.

Population	SNVs (Phased Dataset)	SNVs (Unphased Dataset)
Bantu	9,913,946	9,941,097
Esan	12,698,558	.
Gambian	12,608,797	.
Luhya	11,542,269	.
Mende	14,399,195	.
Mozabite	7,426,081	7,446,628
Yoruba	14,491,062	.
CEPH	8,966,260	.
Basque	6,264,285	6,276,401
British	7,655,754	.
Finnish	7,467,975	.
Iberian	11,145,731	.
Russian	7,482,199	7,510,297
Sardinian	6,596,032	6,610,269
Tuscan	8,383,990	.
Dai	6,803,291	.
Han	6,386,345	7,188,871
Tibetan	6,921,628	7,968,372
Yakut	6,691,957	6,715,004
Aymara	6,920,659	.
Maya	4,040,155	4,040,155
Nahua	4,040,155	4,040,155
Pima	4,040,155	4,040,155

4.3 Effective population size inference in genetically homogeneous populations

Almost the entire set of genetically homogenous populations identified by fine-scale ChromoPainter/fineSTRUCTURE clustering analysis were then subjected to the SMC++ method to infer Ne fluctuations, as required for the generation of genomic simulations. Overall, all the human groups analysed showed comparable Ne values around 20,000 estimated individuals until 100 thousand years ago (kya). From this period, the effective population size trends started to diverge among populations (Figure 4.6).

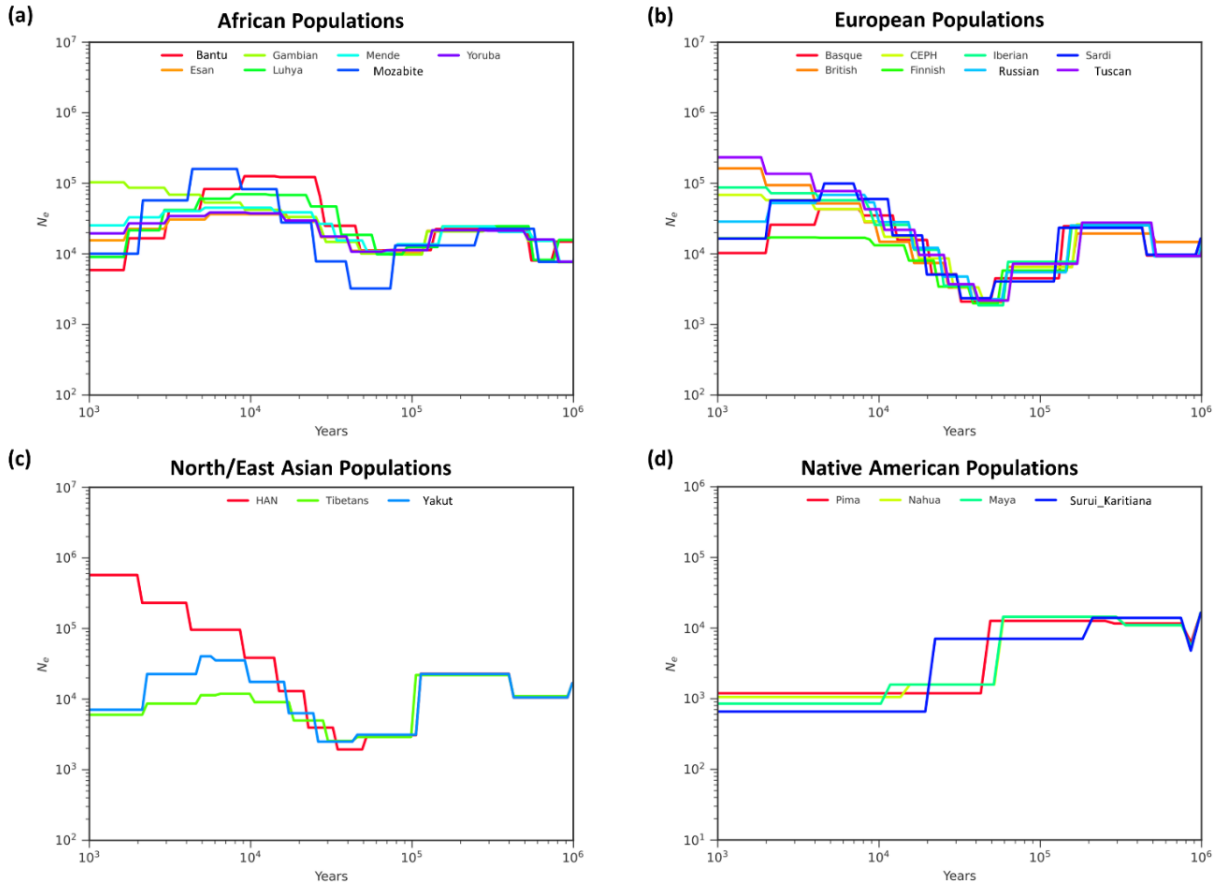


Figure 4.6. Ne fluctuations inferred for African (a), European (b), North/Est Asian (c) and Native American (d) populations. (a) African human groups (except the Mozabite population) display bland decrease in Ne around 100-60 kya and slow population expansion since approximately 50 kya and peaking at 10 kya. During this period, Ne values are of about 100,000 estimated individuals. In more recent times, almost all African populations were found to be characterized by moderate decline in Ne, which finally presented values ranging from 6k to 20k estimated individuals. On the other hand, the Luhya presented Ne values equal to 100k around 1 kya. **(b-c)** Eurasian populations were found to be collectively characterized by consistent decrease in Ne from 100 up to 30 kya and by rapid population expansion. The latter process is severely blunted in the Yakut, high-altitude Tibetan, Sardinian, Finnish and Basque relatively isolated human groups. **(d)** Ne trends obtained for indigenous populations from Central and South America reflected the low level of genetic diversity previously advanced for these human groups. In fact, starting from 10 kya Ne values for these populations settled down at 1,000 estimated individuals.

In particular, the African groups were found to have experienced relatively mild bottlenecks followed by Ne growth, which reached its peak at 10 kya, as previously attested (Terhorst et al. 2017) (Figure 4.6a). Among these populations, only the Mozabite presented a more consistent decrease in Ne (at 3,000 estimated individuals) starting from 70 to 40 kya, possibly due to the reduced sample size of this specific subset (i.e. 14 individuals) with respect to the other African groups considered (Terhorst et al. 2017) (Figure 4.6a).

Human populations from European and North/East Asian macro-areas were instead found to be characterized by a more significant decline of Ne, which dropped at 2,000 estimated individuals around 60 kya and then started to recover only around 30 kya (Figure 4.6b and 4.6c), reflecting the loss of genetic diversity during the Out of Africa migrations. Furthermore, the Eurasian populations, except for the particularly isolated groups of high-altitude Tibetans, Yakuts, Basque, Finnish and Sardinians, resulted to be collectively characterized by recent and rapid demographic expansions starting approximately 10-15 kya (Figure 4.6b and 4.6c).

Finally, Native American populations presented Ne estimations that are concordant with those obtained for Eurasian groups until ~60 kya. However, Ne values inferred for these populations did not support demographic expansion after this period, attesting the great level of isolation and progressive reduction in genetic variability of indigenous populations from Central and South America (Figure 4.6d). Overall, the described SMC++ estimates were in line with those proposed in previous studies (Terhorst et al. 2017; Bergström et al. 2020), thus corroborating the consistency of the obtained results.

4.4 Evaluating the reliability of the trained multinomial classifiers

The inferred Ne values (Terhorst et al. 2017) were used to generate genomic simulations (Kern and Schrider 2016) for the filtered populations subjected to the *Trendsetter* pipeline. In particular, a total of 5,000 genomic simulations for each of the three evolutionary scenarios (i.e. classes) tested by the *Trendsetter* approach were generated for the Yoruba, CEPH, Russian, Han Chinese, high-altitude Tibetan, Yakut, Maya, Pima and Nahua populations and used for the *training* step of the related pipeline of analyses. Particularly, these human groups were selected to represent each geographical macro-area, allowing for the evaluation of *Trendsetter* reliability across populations with diverse ancestries. The reduction in the number of analysed populations by using the *Trendsetter* pipeline was due to the long computational times required particularly for generating population-specific genomic simulations.

Furthermore, validation sets containing 1,000 additional simulations for each tested class were obtained for the same populations and by setting the same parameters used to simulate the training subsets. Specifically, these additional sets of simulations were used to evaluate the accuracy and reliability of the trained algorithms, as well as to visualize calibration/reliability curves of the estimated output probabilities. Overall, reliability plots showed the good calibration achieved for the predicted output probabilities according to what expected for linear multinomial regression classifiers (Figure 4.7, Supplementary Figures S11, S12 and S13) (Mughal and DeGiorgio 2019).

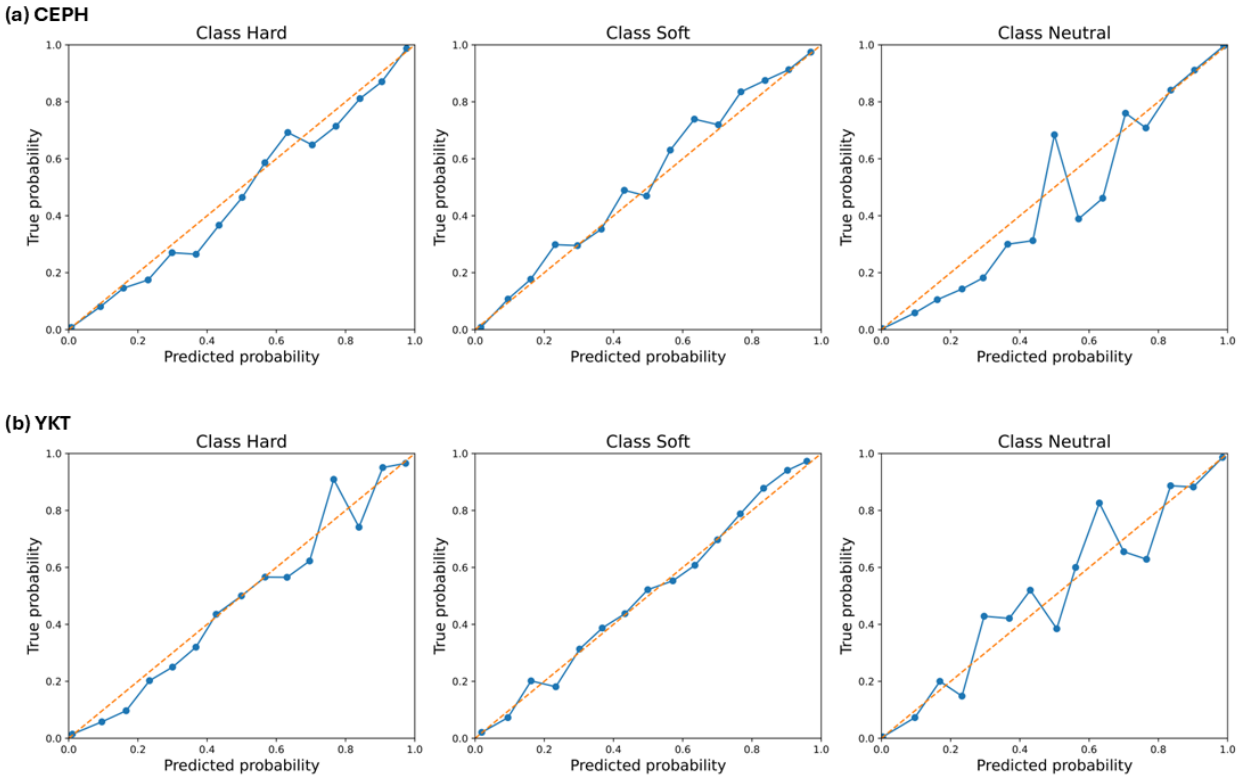


Figure 4.7. *Trendsetter* calibration plots obtained for CEPH (a) and Yakut (b) populations. Calibration/reliability curves of *Trendsetter* inferred probabilities resulted after the application of the trained models to the validation sets built for both the CEPH and Yakut groups. The curves corresponding to each tested class (from left to right: adaptive evolution mediated by hard selective sweeps, adaptive evolution mediated by soft selective sweeps and neutral evolutionary scenarios) well fit with the diagonal thus indicating that no further calibration was needed. The curves were built by dividing the values of the class frequency/predicted *Trendsetter* probabilities in 15 bins.

Moreover, accuracy values were found to be collectively elevated for most of the models, attesting particularly the capacity of the trained classifiers in distinguishing neutrally evolving genomic regions from those having experienced adaptive evolution (Figure 4.8, Supplementary Figure S14 and Supplementary Table S2). In detail, such values varied among 0.95 (i.e. YRI and TIB populations) and 0.69 (i.e. Maya) (Supplementary Table S2). Values of precision, recall and f1-score metrices obtained for all the populations analysed were overall elevated in all the tested classes as well, with greater values being obtained for the neutral one (Supplementary Table S3). In detail, these latter values ranged from 0.8 to 0.98, supporting greater efficiency of the trained *Trendsetter* classifiers in distinguishing genomic regions that evolved neutrally from those subjected to selective pressures rather than distinguishing between hard and soft selective sweep classes (Figure 4.8, Supplementary Figure S14 and Supplementary Table S3).

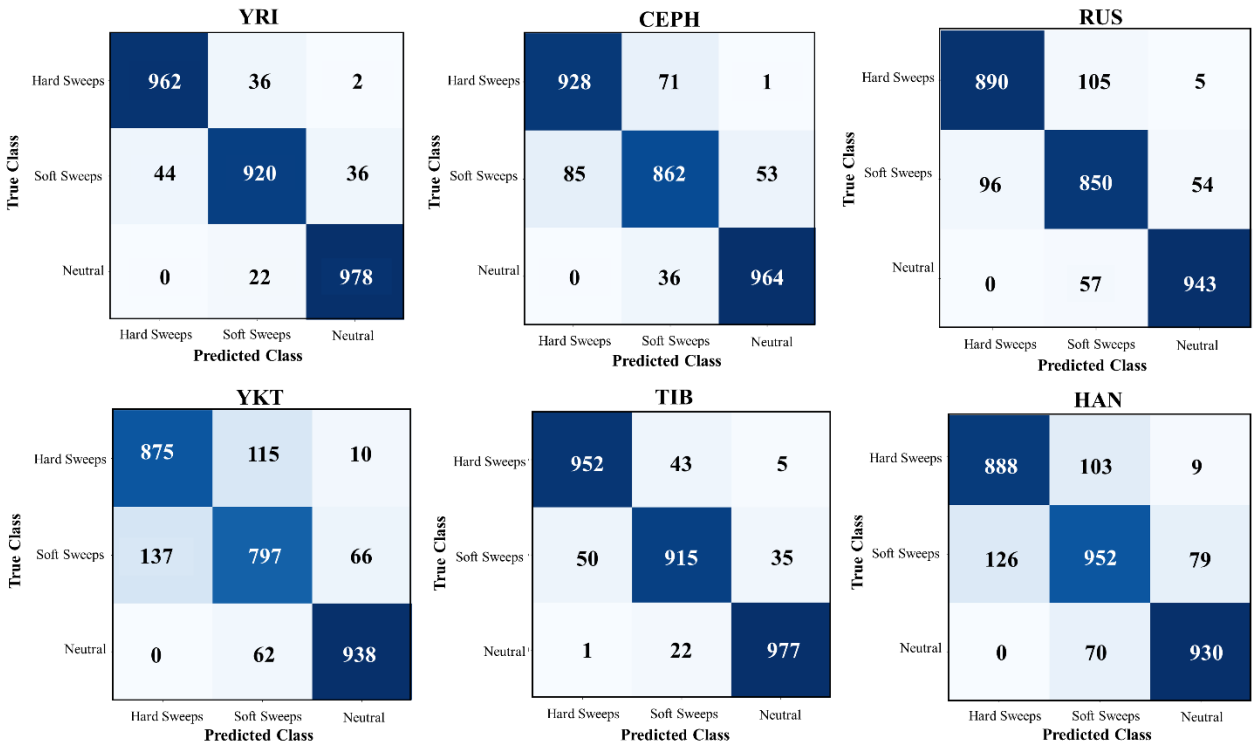


Figure 4.8. Confusion matrices displaying the reliability of the trained *Trendsetter* classifiers obtained for African, European and North/East Asian populations. Confusion matrices were built by relaying on results obtained after the application of the trained *Trendsetter* classifiers on the validation sets of genomic simulations. Along the diagonal are reported the numbers of correct classifications assigned by the algorithm, while in the remaining rows/columns are shown the misclassifications. The top of the figure displays the confusion matrices resulted for Yoruba (YRI), CEPH and Russian (RUS) individuals, while the bottom of the figure reports those resulted for Yakut (YKT), high-altitude Tibetan (TIB) and Han Chinese (HAN) populations.

4.5 Genomic signatures of polygenic adaptations in world-wide human populations

The subsequent sections of the thesis report the principal findings emerged from the analyses performed to identify genomic regions whose variation patterns have been remarkably shaped by the action of natural selection according to a model of polygenic adaptation, and that turned out to have adaptively evolved specifically in a given human group and/or in a set of populations with shared ancestry. These results, which will be presented for each of the considered geographical macro-areas, pertain firstly to gene networks identified through the combined LASSI-*signet* approach as pervasively targeted by positive selection, and secondly, to the outcomes derived from the *Trendsetter* method used as validation analysis. Collectively, biological functions putatively implicated in the modulation of polygenic adaptations are listed in Supplementary Table S4 along with relative P-values and gene-networks. Moreover, the main findings published in the studies by Ferraretti et al. (2024) and Ferraretti et al. (2025) have been integrated in sections 4.5.3.2.1 and 4.5.4.2, respectively.

4.5.1 Adaptive evolution of populations from the African macro-area

4.5.1.1 Results from LASSI-signet analyses

As regards West African populations (i.e. ESN, GMB, MND and YRI) and the cluster of Bantu groups, combination of the LASSI-*signet* methods allowed the identification of putative adaptive networks of genes belonging to the *ECM-receptor interaction* and *Focal adhesion* pathways (Supplementary Table S4a). Although these pathways regulate broad cellular/physiological functions such as cell adhesion, migration, differentiation, proliferation, and apoptosis (Kanehisa and Goto 2000), the genes comprised in the reconstructed networks are known to be involved more specifically in the modulation of immune response to diverse infections, such as Malaria (i.e. *VTN*, *THBS2*, *THBS3*, *THBS4*, and *FN1* genes) (Roberts et al. 1985; Eda and Sherman 2004; Tougan et al. 2018), Amoebiasis (e.g. *VCL* and *FN1*) (Váazquez et al. 1995; Espinosa-Cantellano et al. 2000), and Salmonella infection (e.g. *FLNB*, *FLNC*) (Kanehisa and Goto 2000) (Figure 4.9a). Particularly, the *FLNB* and *FLNC* genes were included in the *Focal adhesion* network obtained for the Bantu populations, while *VCL* resulted a putative adaptive gene in both Bantu and Mende groups (Supplementary Table S4a). Furthermore, signatures of positive selection at the *VTN* and *THBS2*/*3*/*4* genes were detected in Yoruba, Gambian, Mende and Esan, while those concerning *FN1* were obtained for Mende, Gambian and Bantu groups (Supplementary Table S4a). In line with this finding, we also identified a significant network belonging to the *JAK-STAT signalling* pathway in both Gambian and Yoruba populations, which particularly included multiple immune-related genes (i.e. codifying for interleukins and related receptors and/or involved in the activation of cytokine responses), such as *IL15*, *IL19*, *IL22RA2* and *IFNA6* (Figure 4.9b, Supplementary Table S4a).

A significant gene-network belonging to the *Tuberculosis* pathway and including several *HLA* genes (i.e. *HLA-DPA1*, *HLA-DPB1*, *HLA-DRB1*, *HLA-DRB5*) was also found for the Yoruba population (Figure 4.9c). Finally, the *CD36* gene, which has been previously proposed to be involved in the modulation of Malaria infection (Omi et al. 2003; Sinha et al. 2008; Cabrera et al. 2014), was found to participate to the significant gene-network obtained for Gambians and belonging to the *AMPK signalling* pathway (Supplementary Table S4a).

Conversely, signatures of polygenic adaptations in North African Mozabite were instead related to sugar metabolism, as suggested by the significant gene-network belonging to the *Starch and sucrose metabolism* pathway (Supplementary Table S4a). In detail, this network includes genes such as *PYGL*, *GYS2*, *SI* and *MGAM***,** which codify for specific enzymes implicated in glycogen new synthesis and degradation, as well as in starch digestion (Sayers et al. 2022) (Supplementary Figure S15).

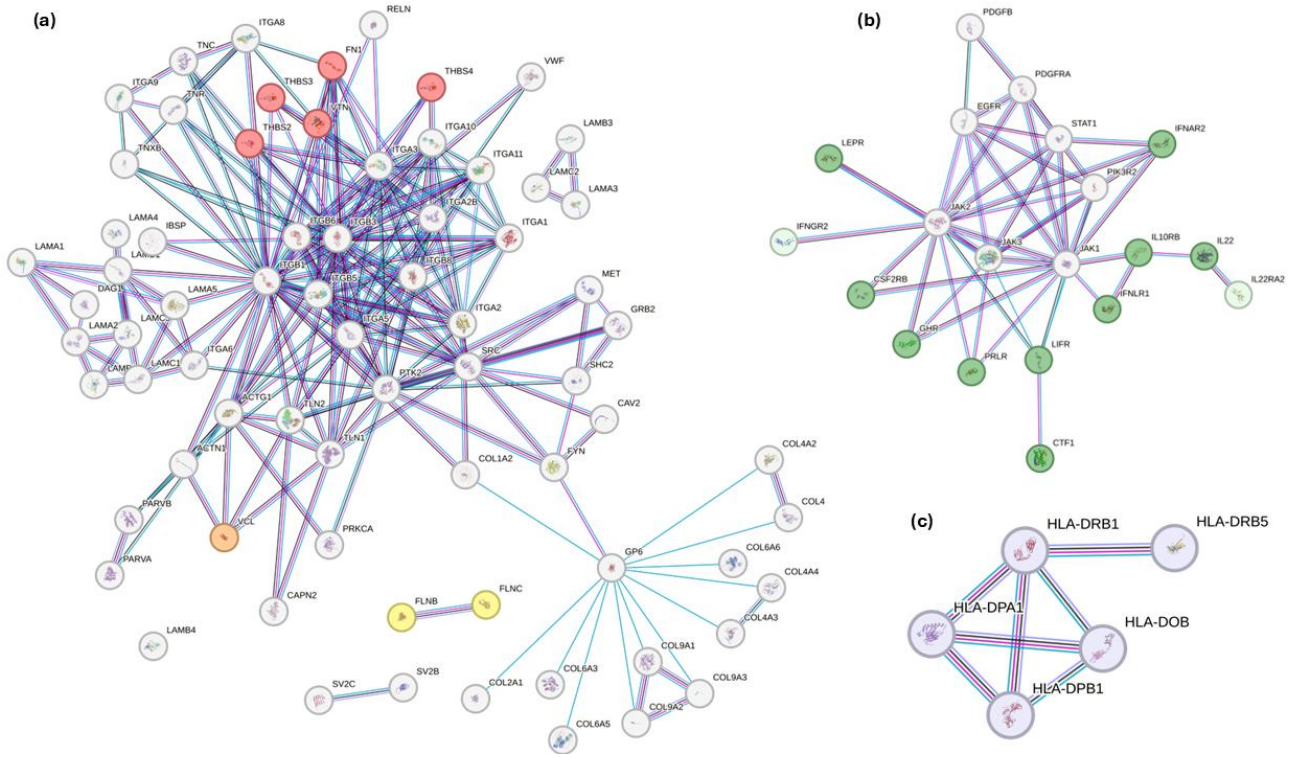


Figure 4.9. Gene-networks targeted by natural selection in West African populations and in the Bantu ethnic group. All the reported networks of genes were build using the STRING tool (available at https://string-db.org/) and by setting the confidence score to predict gene-gene associations (i.e. edges) at 0.9. Functional associations among genes were inferred by considering co-expression (i.e. black edges), experimental (i.e. pink edges) and database (i.e. light-blue edges) evidence. **(a)** Gene-networks belonging to the *ECM-receptor interaction*/*Focal adhesion* pathways observed in West African and Bantu populations. Genes that play a role in modulating diverse stages of Malaria infection are displayed in red, while *VCL* and *FLNB*/*C* genes, which have been annotated respectively in *Amoebiasis* and *Salmonella Infection* pathways are reported as orange and yellow circles. **(b)** Gene-network within the *JAK-STAT signalling* pathway obtained for Yoruba and Gambian populations. Genes enriched in the *Cytokine-cytokine receptor interaction* pathway (FDR = 6.23e-18) are displayed with dark green circles. Genes related with activation/modulation of immune responses according to literature evidence (Sayers et al. 2022) are reported in light green. **(c)** Significant network obtained for the Yoruba group encompassing *HLA* genes and comprised in the *Tuberculosis* pathway.

4.5.1.2 Selection signatures validated in the Yoruba population

As previously mentioned, the full selection pipeline including the *Trendsetter* validation step, has so far been applied only to some populations highlighted in bold in Table 1, due to the extended computational time required to generate population-specific genomic simulations. More specifically, these human groups were selected as examples of the overall human biodiversity and includes also two valuable case-studies to investigate human adaptations to extreme environments (i.e. Yakuts and high-altitude Tibetans). Concerning the African macro-area, the *Trendsetter* approach was completed for the Yoruba population and partially recapitulates the adaptive signals identified with the LASSI-*signet* methods. More specifically, several genomic windows located within or near the *VTN*, *HLA-DRB5*, *HLA-DRB1*, *HLA-DOB*, *HLA-DPA1*, *HLA-DPB1*, and *IL15* genes exhibited extremely low probabilities of being classified as either neutral genomic regions or as chromosomal segments having experienced hard selective sweeps, with the highest obtained

probabilities pointing to their belonging to the soft selective sweep class (Figure 4.10, Figure 4.11, Supplementary Figure S16 and Supplementary Table S4a). In detail, 72 out of 100 genomic windows comprised in the *VTN* gene and in its surrounding genomic regions were classified as affected by soft selective sweeps (Figure 4.10). Several of these windows presented output probabilities for neutral evolution of ~10%, compared to ~ 0.85% of being classified as soft selective sweeps and they localize particularly in both up and downstream genomic regions and within the *VTN* gene as well (Figure 4.10).

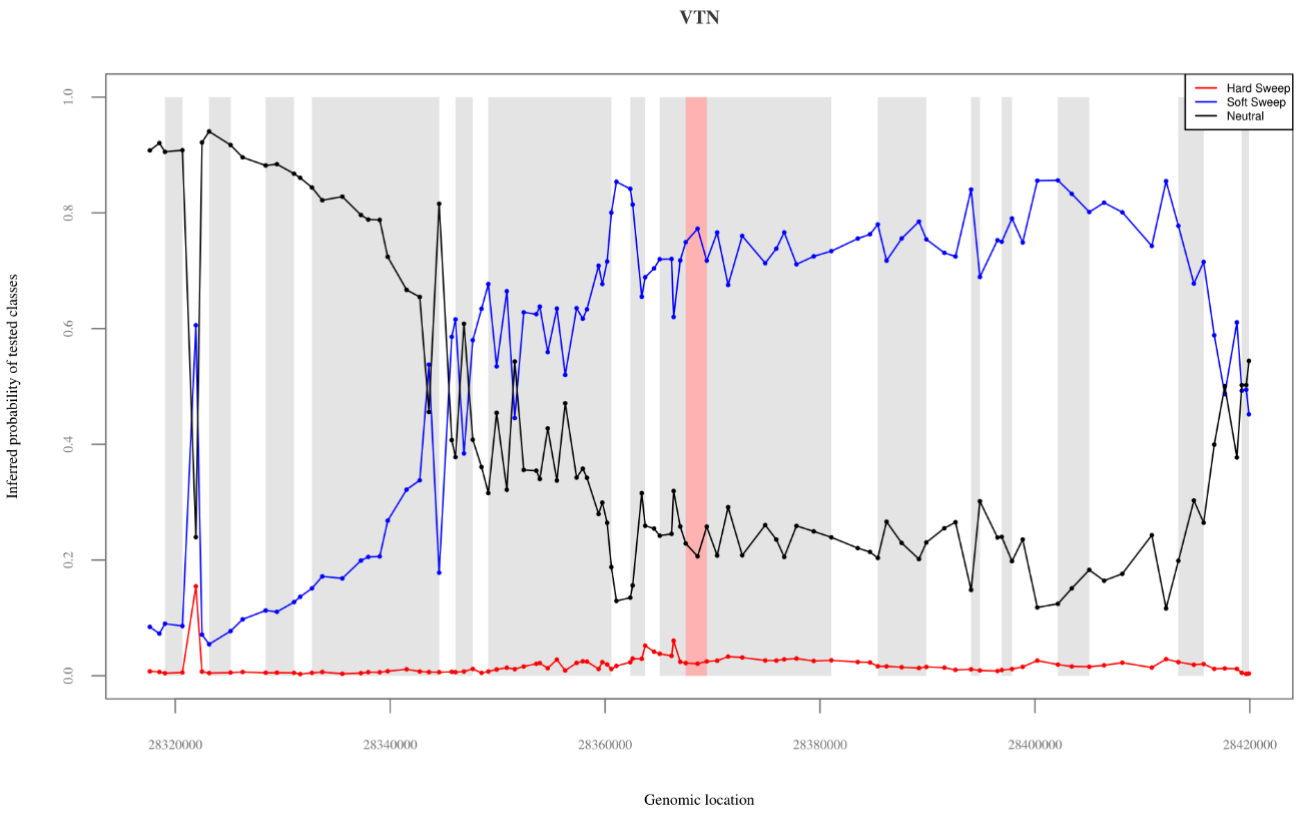


Figure 4.10. *Trendsetter* predicted probabilities at the *VTN* gene and its surrounding genomic regions for the Yoruba population. Distribution of *Trendsetter* output probabilities for the considered genomic windows of being classified as having experienced hard selective sweeps (i.e. red curve), soft selective sweeps (i.e. blue curve) and neutral evolution (i.e. black curve) across the putative adaptive *VTN* gene (i.e. red rectangle) and in flanking regions spanning 50k bases up- and down-stream the gene. The x axis reports the positions of the central SNVs in each genomic window identified according to the *Trendsetter* approach, while the y axis shows the values of *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent genomic regions in the considered genes that present values indicative of the action of natural selection according to the T *LASSI* statistic. Probabilities values associated to the soft selective sweeps class are consistently elevated in the final portion of the upstream regions and in almost the entire downstream regions of the *VTN* gene. Such a pattern is also observed at the genomic windows comprised within the *VTN* gene.

In line with these findings, ~68% (i.e. 2,036 out of 2,986) of the genomic widows included in the chromosomal intervals associated to *HLA* genes was assigned to the soft selective sweeps class and ~1% (i.e. 39 out of 2,986) to the hard selective sweeps class (Figure 4.11). Overall, in several of these windows the probability of being classified as soft selective sweeps reached values of ~99% respect to ~10^{-5}% and ~10^{-3}% of being assigned to the neutral and hard selective sweep classes, respectively (Figure 4.11). The genomic windows classified as hard selective sweeps, which were comprised between *HLA-DRB5* and *HLA-DRB1*, as well as in the *HLA-DPA1* and *HLA-DPB1* genes, showed associated probabilities peaking respectively at ~90% and ~76% (Figure 4.11). Remarkably, analogously to the *VTN* gene, no genomic windows assigned to the neutral class were identified for the *HLA* genes under investigation (Figure 4.11 and Supplementary Figure S16).

Finally, 46 out of 109 *IL15* genomic windows were classified as soft selective sweeps, with average values of *Trendsetter* output probability peaking at ~81%, compared to the average ~33% of being classified as neutral regions (Supplementary Figure S16).

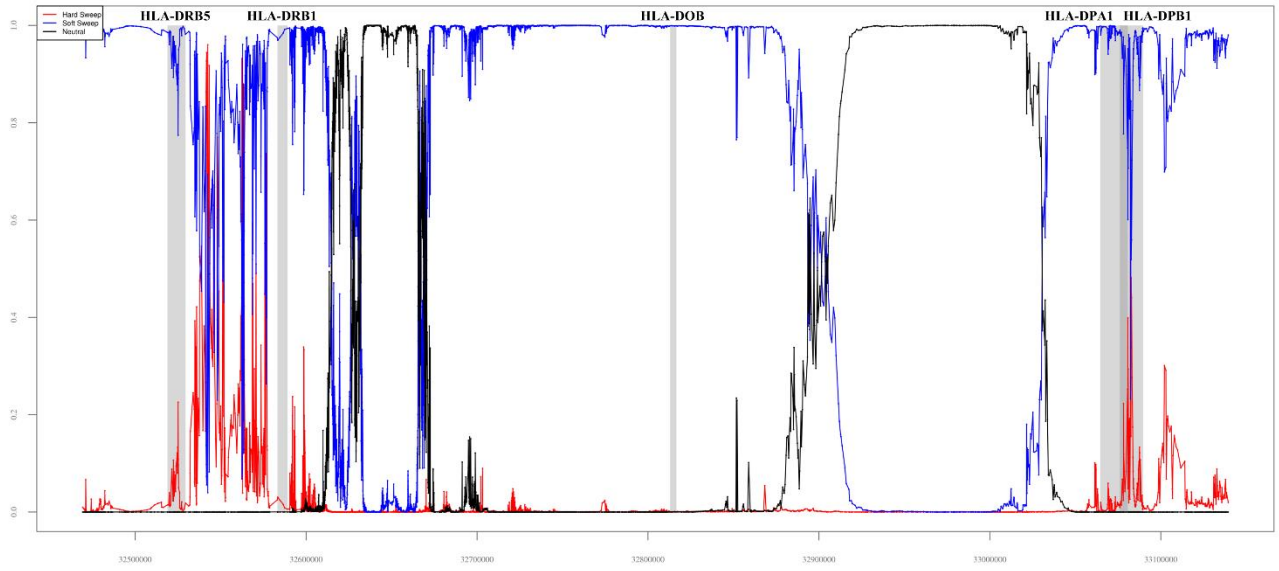


Figure 4.11. *Trendsetter* predicted probabilities at the *HLA* genes and their surrounding genomic regions for the Yoruba population. Distribution of *Trendsetter* output probabilities for the considered genomic windows of being classified as having experienced hard selective sweeps (i.e. red curve), soft selective sweeps (i.e. blue curve) and neutral evolution (i.e. black curve) across the putative adaptive *HLA* genes (i.e. grey rectangles) and in flanking regions. The x axis reports the positions of the central SNV in each genomic window identified according to *Trendsetter* approach, while on the y axis are shown the values of predicted probabilities associated to each of the three tested classes. Probabilities values assigned to the soft selective sweeps class are remarkably elevated for all genes, reaching values almost equal to one at these loci. Such trends are often maintained in both up and down stream genomic regions of *HLA* genes. Furthermore, some of the windows falling in the genomic portions between *HLA-DRB5* and *HLA-DRB1* genes and within *HLA-DPB1* are classified as having experienced hard selective sweeps, presenting probability values associated to this class reaching ~90% and ~76%, respectively.

4.5.2 Adaptive evolution of populations from the European macro-area

4.5.2.1 Results from LASSI-signet analyses

Overall, the LASSI-*signet* analyses pointed out putative adaptive gene networks involved in lipid metabolism (i.e. *Glycerolipid metabolism*, *Glycerophospholipid metabolisms*, and *Sphingolipid metabolism* pathways) for multiple European populations (Supplementary Table S4b). Specifically, a gene network belonging to the *Glycerolipid metabolism* pathway was identified for Finnish, Iberian, Russian, and Tuscan populations (Figure 4.12a and Supplementary Table S4b). This network includes multiple genes involved in diverse stages of lipid metabolism, such as lipid digestion (e.g. *LIPC* and *PNLIP* hepatic and pancreatic lipases), lipid new synthesis (e.g. *PLPP3*) and uptake (e.g. *MOGAT2* and *PLPP1*) (Sayers et al. 2022).

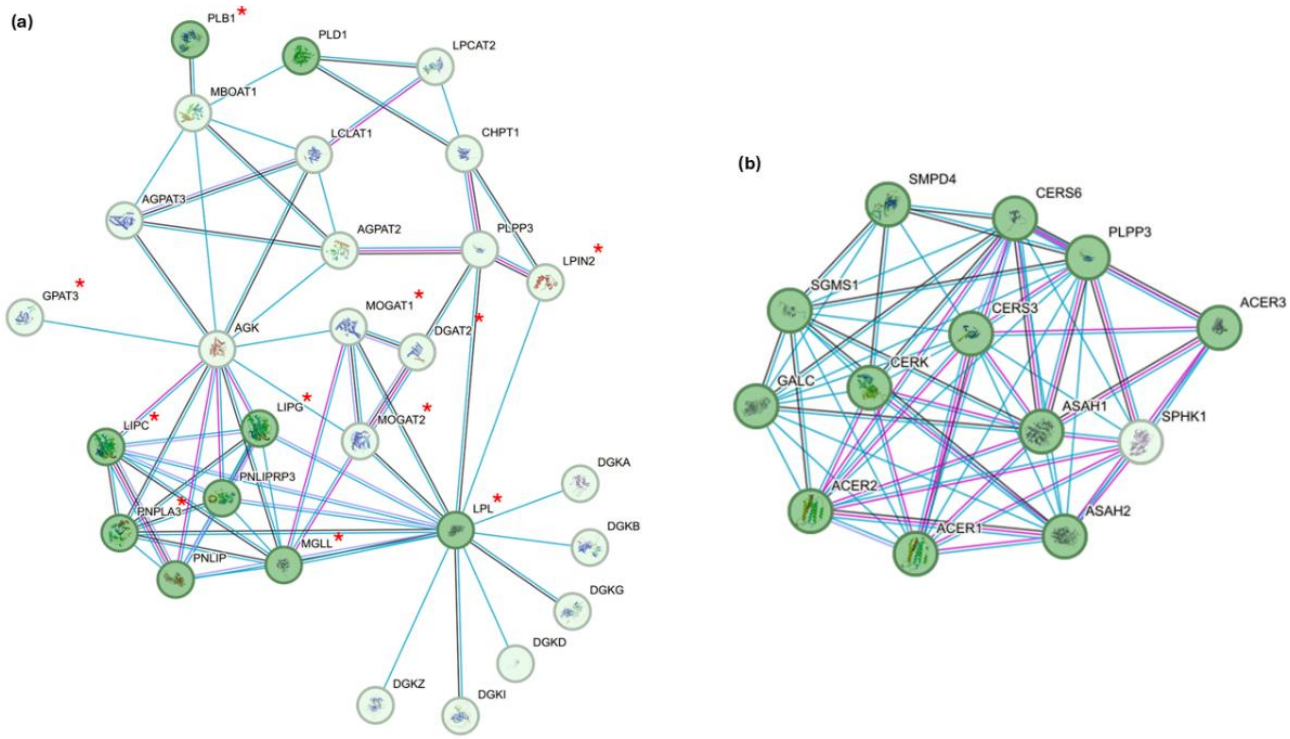


Figure 4.12. Gene-networks targeted by natural selection in European populations. Both networks were build using the STRING tool (available at https://string-db.org/) and by setting the confidence score to predict gene-gene associations (i.e. edges) at 0.9. Functional associations among genes were inferred by considering co-expression (i.e. black edges), experimental (i.e. pink edges) and database (i.e. light-blue edges) evidence. **(a)** Significant networks belonging to the *Glycerolipid metabolism* and/or *Glycerophospholipid metabolism* pathways were observed for Finnish, Iberian, Russian, Tuscan, and Sardinian populations. Genes reported in dark green are enriched in the *Lipase activity* process annotated in Gene Ontology database (FDR = 5.50e-12). Genes marked with the red asterisks are instead enriched in the Gene Ontology *Triglyceride metabolic process* (FDR = 8.71e-17). **(b)** Gene-network belonging to the *Sphingolipid metabolism* pathway observed for the British population. Genes enriched in *Ceramide metabolic process* of Gene Ontology (FDR = 4.98e-23) are reported as dark green circles. All the genes in the network were enriched in the *Sphingolipid metabolic process* annotated in the Gene Ontology database (i.e. FDR = 1.55e-23).

Moreover, in the Sardinian population two gene networks apparently related to such functions were identified. The first one belonged to the *Glycerophospholipid metabolism* pathway and the second one to the *Glycerolipid metabolism* pathway, despite the latter was associated with a not significant P-value (0.057). Nevertheless, multiple genes included in such networks resulted secondarily related to the aforementioned processes, being proved to play pleiotropic functions such as platelet activating factor biosynthetic processes (i.e. *CHPT1* gene), membrane phospholipid remodelling (i.e. *PLA2G6*) and regulation of the intracellular diacylglycerol (DAG) concentration (i.e. *DGKI*, *DGKB* and *DGKG*) (Sayers et al. 2022).

A significant network belonging to the *Sphingolipid metabolism* pathway was instead pointed out for the British population (Figure 4.12b and Supplementary Table S4b), including genes involved in metabolic processes of Ceramides and Sphingolipids, which regulate the composition of the epidermic stratum corneum (Denda et al. 1993; Choi and Maibach 2005) (Figure 4.12b and Supplementary Table S4b).

Furthermore, *ADCY3* and *KITLG* genes, which have been previously proposed as targets of natural selection in populations from Northern Italy (Sazzini et al. 2020) and in several Eurasian human groups (Yang et al. 2018) were respectively included in significant gene-networks belonging to the *Purine metabolism* pathway (as detected for Russian, Tuscan, Sardinian and Finnish populations) and to the *MAPK/Ras signalling* pathway (as obtained for Tuscan, Sardinian, Basque, Finnish and CEPH individuals) (Supplementary Table S4b). The haplotype structure of the *ADCY3* gene, as well as its comparison with haplotypes observed in archaic human species will be further discussed in the present thesis (see Figure 4.26 and paragraph 4.5.3.3.2).

Finally, in some European human groups, the LASSI-*signet* methodologies recognized multiple significant gene-networks belonging to immune-related pathways (i.e. *Chemokine signalling* and *Pathogenic Escherichia coli infection* pathways) (Supplementary Table S4b).

4.5.2.2 Selection signatures validated in the CEPH and Russian populations

The application of trained *Trendsetter* classifiers on CEPH and Russian datasets allowed us to validate some of the selection signatures described in the previous paragraph.

Although no LASSI-*signet* networks related to the *Glycerolipid metabolism* pathway were observed for CEPH, part of the genes related to such metabolic processes (i.e. *MOGAT2*, *PNLIP* and *PLPP1*) were pointed by *Trendsetter* as showing variation patterns indicative of the action of natural selection in this group (Figure 4.13).

More in detail, in the *MOGAT2* gene seven out of 13 genomic windows classified as having experienced soft selective sweeps with a predicted probability peaking at ~88% compared to ~9% and ~3% of the hard sweep and neutral classes, respectively, were identified (Figure 4.13a). Furthermore, all the genomic windows within *PNLIP* and *PLPP1* genes were assigned to the soft select sweeps class, with average values of the output probability equal to ~86% and ~93% (Figure 4.13b-c).

In line with this picture, selection signatures observed for the *MOGAT2*, *PNLIP* and *PLPP1* genes were further validated also in the Russian population (Figure 4.14a-b and Supplementary Table S4b). Furthermore, the

Trendsetter approach allowed us to corroborate signatures of natural selection described for the *ADCY3* gene in this human group (Figure 4.14c and Supplementary Table S4b).

Finally, the trained *Trendsetter* classifier recognized adaptive genomic regions within the *KITLG* gene in the CEPH population (Supplementary Figure S17).

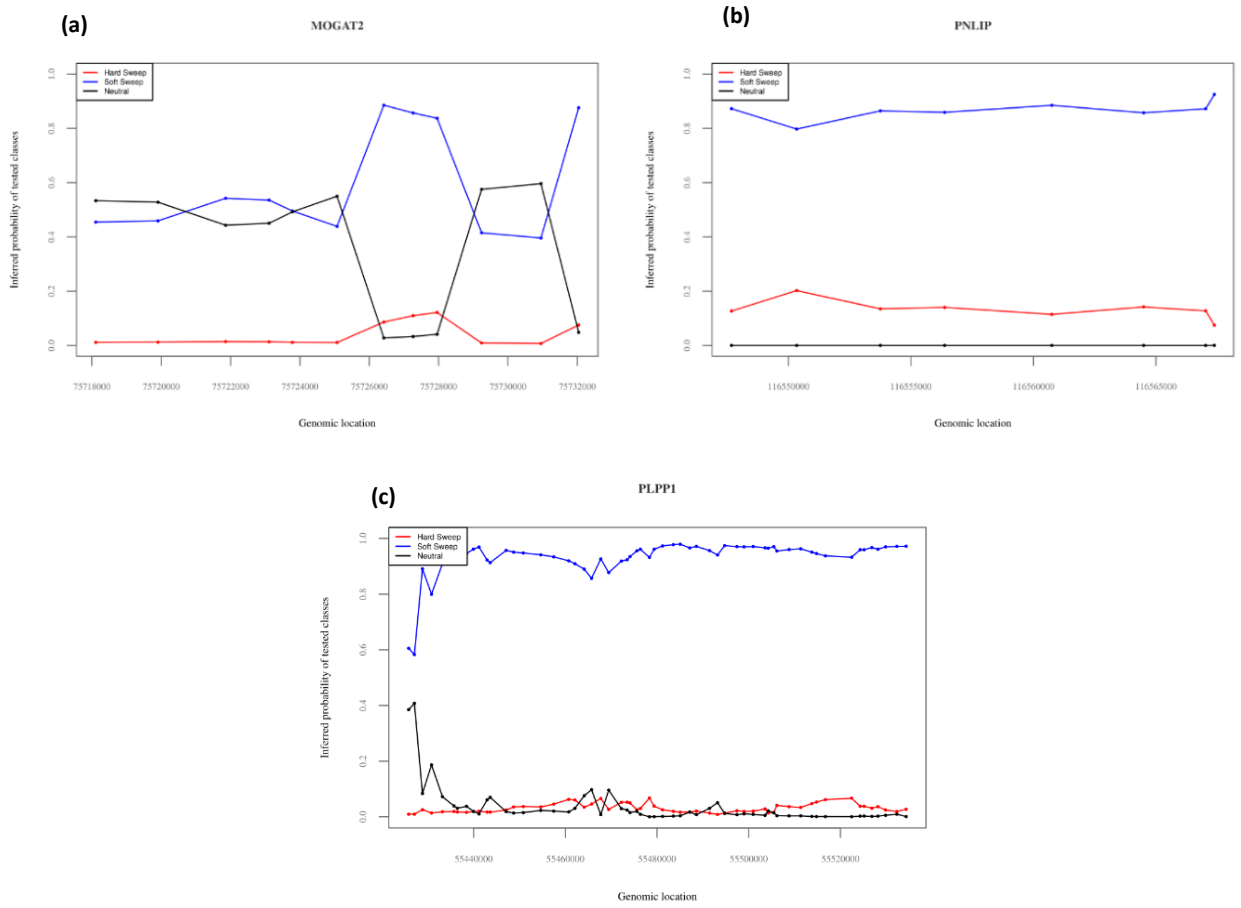


Figure 4.13. *Trendsetter* predicted probabilities across candidate adaptive genes modulating lipid metabolism in the CEPH population. Distribution of *Trendsetter* output probabilities for the considered genomic windows of being classified as having experienced hard selective sweeps (i.e. red curve), soft selective sweeps (i.e. blue curve) and neutral evolution (i.e. black curve) at **(a)** *MOGAT2*, **(b)** *PNLIP* and **(c)** *PLPP1* genes in the CEPH population. For all the plots, the x axis reports the positions of the central SNVs in each genomic window identified according to the *Trendsetter* approach, while the y axis shows the values of *Trendsetter* predicted probabilities associated to each of the three tested classes. **(a)** Predicted probabilities associated to the neutral class in the first two windows identified for the *MOGAT2* gene are slightly more elevated with respect to those assigned to the soft selective sweep class. However, such a trend rapidly inverts in the third and in the fourth window of this gene. Greater values associated to the soft sweep class are then obtained for the genomic windows located at the middle/end portions of the *MOGAT2* gene. **(b-c)** Output probabilities associated to the soft selective sweep class inferred for both *PNLIP* and *PLPP1* genes are outstanding with respect to those associated to the neutral and hard selective sweep classes across the entire loci. In particular, the greatest soft sweep-associated probability values were equal to ~92% and ~98% at *PNLIP* and *PLPP1*, respectively.

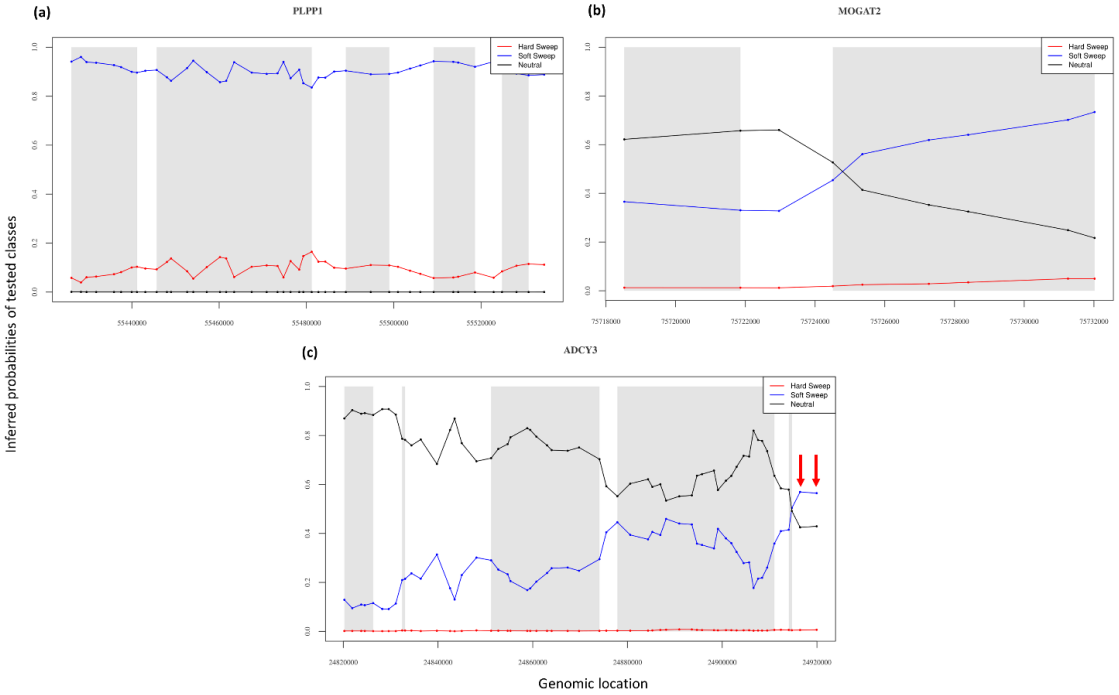


Figure 4.14. *Trendsetter* predicted probabilities at genes involved in *Glycerolipid metabolism* and *Purine metabolism* pathways in the Russian population. Distribution of *Trendsetter* predicted probabilities for the considered genomic windows of being classified as having experienced hard selective sweeps (i.e. red curve), soft selective sweeps (i.e. blue curve) and neutral evolution (i.e. black curve) across candidate adaptive genes included in LASSI-*signet* significant networks belonging to *Glycerolipid metabolism* **(a-b)** and *Purine metabolism* **(c)** pathways observed for the Russian population. For all the plots, the x axis reports the positions of the central SNVs in each genomic window identified according to the *Trendsetter* approach, while the y axis shows the values of *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent genomic regions in the considered genes that present values indicative of the action of natural selection according to the T *LASSI* statistic. **(a-b)** Output probabilities associated to the soft selective sweeps class inferred for the *PLPP1* gene are found to be outstanding with respect to those associated to the neutral one across the entire locus. A similar pattern is recognizable also in the middle/final portion of the *MOGAT2* gene, reinforcing signatures of natural selection identified for these two loci by adopting different statistical methods. **(c)** Output probabilities associated to the soft selective sweep class reach moderate values across the entire *ADCY3* gene. However, they overcome probabilities associated to the neutral class only in the final portion of the gene (i.e. red arrows).

4.5.2.3 Selection signatures at malaria-associated loci in the Sardinian population

The application of both LASSI and *signet* methodologies to the Sardinian dataset allowed us to identify a significant network belonging to the *Oxytocin signalling* pathway (Supplementary Table S4b).

Specifically, such network includes genes regulating pleiotropic functions (i.e. *EGFR*, *PRKCA* and *PRKCB*), as well as the *CD38* gene that has been previously proposed to modulate immune response in Malaria infections (Asito et al. 2008; Schwenk et al. 2013; Burel et al. 2016) (Supplementary Table S4b).

In line with this finding, also *CD36* and *CD81* genes, which have been previously proved to play cardinal roles in diverse stages of Malaria infection (Omi et al. 2003; Silvie et al. 2003; Silvie et al. 2006; Sinha et al. 2008; Cabrera et al. 2014), showed signatures ascribable to adaptive evolution in Sardinian population according to the LASSI analysis. In detail, these genes presented multiple genomic widows characterized by values of the

LASSI statistic included in the top 1% of the genomic distribution obtained (Figure 4.15). Particularly, the *CD36* gene presented a total of 15 top 1% LASSI genomic windows, all classified as regions having experienced hard selective sweeps (i.e. $m = 1$) (Figure 4.15). Furthermore, within the CD81 gene a total of 22 genomic windows showing T statistic falling in the top 1% of the obtained distribution were observed, being classified again as regions having experienced hard selective sweeps (Figure 4.16).

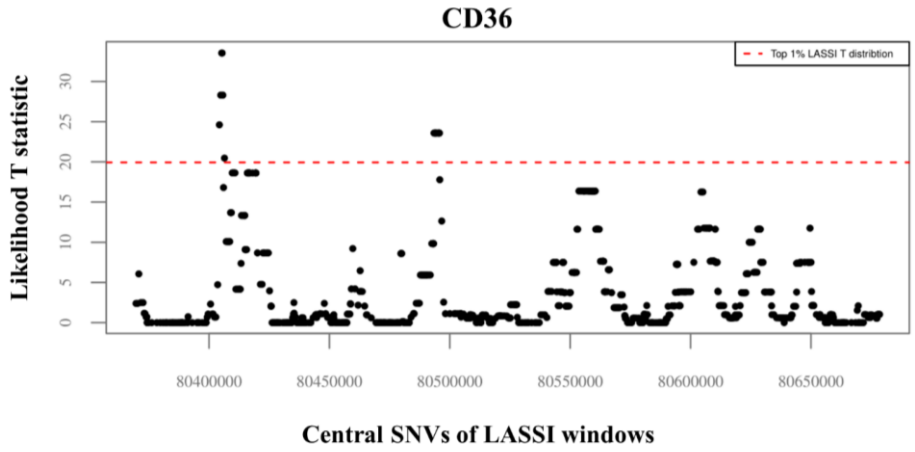


Figure 4.15. Values of the LASSI T statistic across *CD36* genomic windows. The x axis displays the genomic position of the SNVs located at the centre of the genomic windows considered in the LASSI analysis. The y axis reports the value of the likelihood *T* statistic calculated by the LASSI approach. The red dashed line indicates the top 1% threshold of the distribution of the *T* statistic. A total of 15 genomic windows (i.e. black dots above the threshold) is associated to top 1% *T* statistic values, with the greatest likelihood value obtained equal to ~33.53.

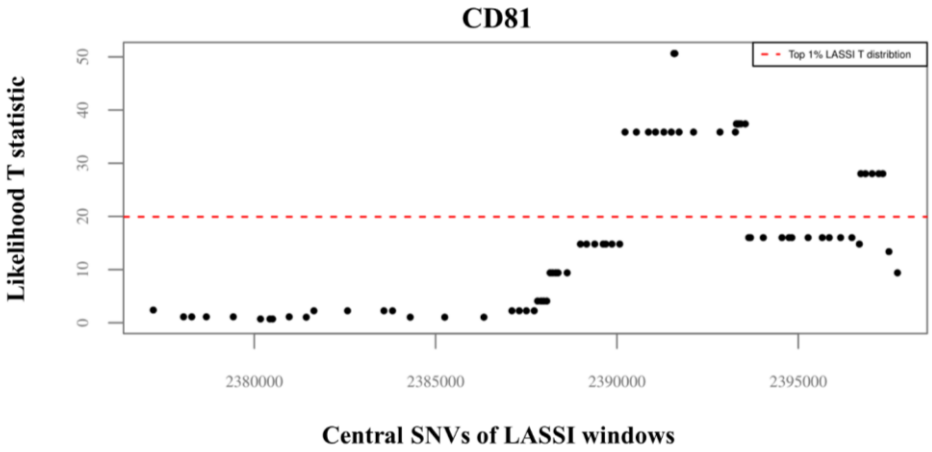


Figure 4.16. Values of the LASSI T statistic across *CD81* genomic windows. The x axis displays the genomic position of the SNVs located at the centre of the genomic windows considered in the LASSI analysis. The y axis reports the value of the likelihood *T* statistic calculated by the LASSI approach. The red dashed line indicates the top 1% threshold of the distribution of the *T* statistic. A total of 22 genomic windows showing values of the T statistic falling within the top 1% of the obtained distribution were observed for the *CD81* gene, presenting likelihood values peaking at 50.64 in the middle/end portions of the gene.

4.5.3 Adaptive evolution of populations from the North/East Asian macro-area

4.5.3.1 Signatures of polygenic adaptations in Han and Dai Chinese populations

The first two steps of the adopted pipeline of analyses allowed us to identify a putative adaptive gene-network belonging to the *JAK-STAT signalling* pathway for the Han Chinese population (Figure 4.17a and Supplementary Table S4c). Similarly to what observed for other populations (e.g. West African, Bantu, CEPH and Finnish populations; see Supplementary Table S4 a-b), such network includes immune-related genes that codify for cytokines (i.e. *IFNL3* and *IL7*), cytokine receptors (i.e. *IFNAR2* and *IL23R*) and interferon gamma receptor (i.e. *IFNGR2*) (Sayers et al. 2022) (Figure 4.17a). Moreover, such a network encompasses the *LEPR* gene, a locus that has been previously proposed to be involved in the increased heat dissipation by brown adipose tissue in modern East Asian populations (Sazzini et al. 2014) (Figure 4.17a).

In addition, according to our analyses, a significant gene-network belonging to the *Pentose and glucuronate interconversions* pathway was supposed to have been targeted by natural selection specifically in the Dai Chinese population (Figure 4.17b and Supplementary Table S4c). Particularly, this network includes several genes belonging to the UDP glucuronosyltransferase family 1 (e.g. *UGT1A7* and *UGT1A8*), which comprises enzymes involved in the glucuronidation pathway, a process responsible for the transformation of small lipophilic molecules, such as steroids, bilirubin, hormones, and drugs, into water-soluble, excretable metabolites (Sayers et al. 2022).

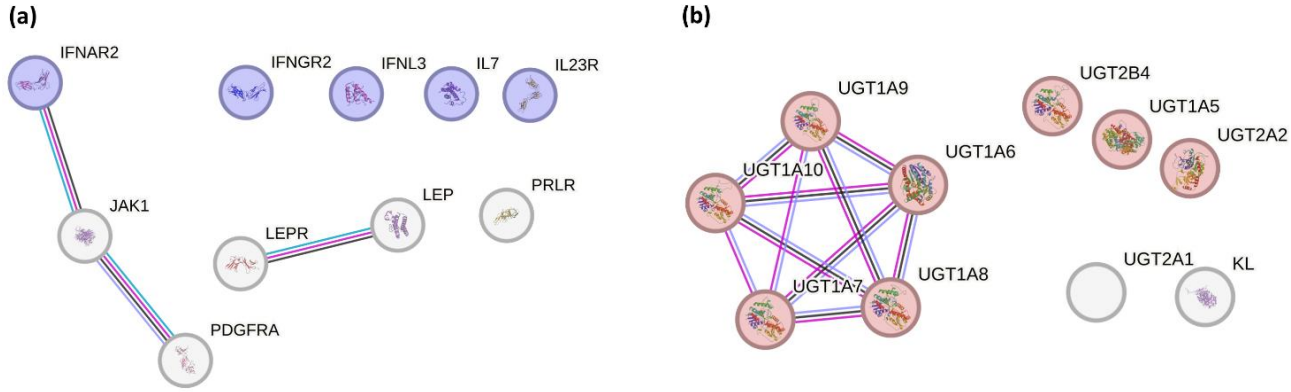


Figure 4.17. Gene-networks targeted by natural selection in the (a) Han and (b) Dai Chinese populations. Gene-networks were build using the STRING tool (available at https://string-db.org/) and by setting the confidence score to predict gene-gene associations (i.e. edges) at 0.9. Functional associations among genes were inferred by considering co-expression (i.e. black edges), experimental (i.e. pink edges) and database (i.e. light-blue edges) evidence. **(a)** A significant network belonging to the *JAK-STAT signalling* pathway was observed for the Han Chinese population. *IFNL3*, *IFNGR2*, *IFNAR2*, *IL23R*, and *IL7* immune-related genes are reported as blue circles in the network. Although these genes belong to the same biological processes (i.e. they are all annotated in the *Cytokine-mediated signalling* pathway according to the Gene Ontology database and in the *JAK-STAT signalling* and *Cytokine-cytokine receptor interaction* pathways according to the KEGG database), most of them do not establish direct interactions within the reconstructed network. **(b)** Gene-network belonging to the *Pentose and glucuronate interconversions* pathway was observed for the Dai Chinese population. Genes enriched in the *Cellular glucuronidation* process (i.e. FDR = 2.77e-18) and annotated in the Gene Ontology database are reported as red circles.

In line with what observed for European populations, also the *KITLG* gene was included in the putative adaptive networks identified by the LASSI-*signet* analyses for both Dai and Han Chinese populations, although the latter one presented a barely significant P-value of 0.052 (Supplementary Table S4c). In detail, these networks were encompassed in the *Rap1* and *PI3K-Akt signalling* pathways for the Dai and Han Chinese populations, respectively (Supplementary Table S4c).

Furthermore, implementation of the *Trendsetter* method allowed us to validate selection signals at *IL23*, *IL7* and *IFNL3* immune-related genes in the Han Chinese population. Again, several genomic windows within these loci were classified as having experienced soft selective sweeps (Figure 4.18 and 4.19). In particular, 35 out of 45 genomic windows at the *IL23* gene were classified as belonging to the soft selective sweep class, with an average predicted probability of ~66% in contrast to ~32% and ~1% of being assigned to the neutral and hard selective sweep classes, respectively (Figure 4.18a). Accordingly, probabilities values associated to the soft selective sweep class ranged from ~71% to ~51% for *IL7* genomic windows (Figure 4.18b).

In agreement with these findings, the *IFNL3* gene, as well as all the 40 windows encompassing the genomic regions surrounding it, were classified as having experienced soft selective sweeps, with predicted probability values peaking at ~89% (Figure 4.19). Finally, *Trendsetter* results obtained for the *LEPR* and *KITLG* genes further support signatures of natural selection previously identified for East Asian and Eurasian populations, respectively (Supplementary Figure S18 and S19) (Sazzini et al. 2014; Yang et al. 2018).

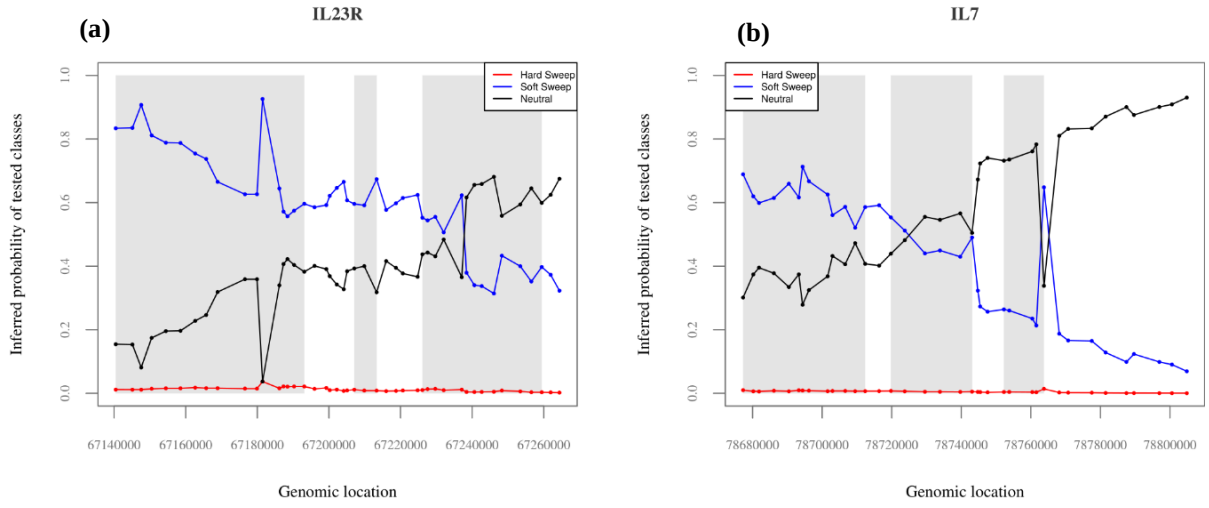


Figure 4.18. *Trendsetter* predicted probabilities at *IL23R* (a) and *IL7* (b) genes for the Han Chinese population. Distribution of *Trendsetter* predicted probabilities for the considered genomic windows of being classified as having experienced hard selective sweeps (i.e. red curve), soft selective sweeps (i.e. blue curve) and neutral evolution (i.e. black curve) across candidate adaptive genes related to immune responses in the Han Chinese population. For both the plots, the x axis reports the positions of the central SNVs of each of the genomic windows identified for *IL23R* and *IL7* genes. The y axis shows the values of the *Trendsetter* predicted probability associated to each of the three tested classes. The grey rectangles in the background represent genomic regions in the considered genes that present values indicative of the action of natural selection according to the LASSI statistic. **(a)** Output probabilities associated to the soft selective sweeps class inferred for the *IL23R* gene are found to be much more elevated with respect to those associated to the neutral one in the first/middle portions of the locus and present a peak reaching ~93%. **(b)** A similar pattern is recognizable also in the initial portion of the *IL7* gene, even though the predicted probability values obtained for such a locus are slightly less elevated with respect to those computed for *IL23R*.

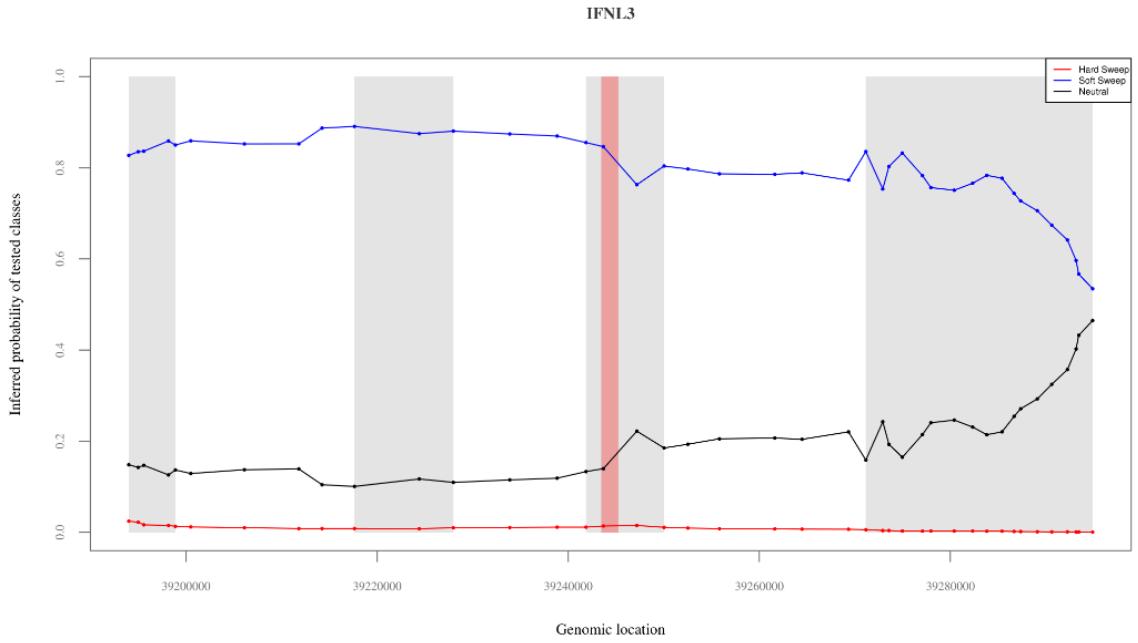


Figure 4.19. *Trendsetter* predicted probabilities at *IFNL3* gene and its surrounding genomic regions for the Han Chinese population. Distribution of *Trendsetter* predicted probabilities for the considered genomic windows of being classified as having experienced hard selective sweeps (i.e. red curve), soft selective sweeps (i.e. blue curve) and neutral evolution (i.e. black curve) across the putative adaptive gene *IFNL3* (i.e. red rectangle) and in its flanking genomic regions spanning 50k bases up- and down-stream the gene. The x axis reports the positions of the central SNVs in each genomic window identified according to the *Trendsetter* approach, while on the y axis are shown the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. Probabilities values associated to the soft selective sweeps class are found to be consistently elevated in the entire genomic region analysed. Specifically, the greatest values are detected for both the gene upstream genomic portions and the *IFNL3* locus.

4.5.3.2 Polygenic bases of high-altitude adaptation in populations of Tibetan ancestry

First and foremost, application of the LASSI analysis to the Tibetan dataset recapitulated the evidence previously collected for the *EPAS1* and *EGLN1* genes, which have been long proposed as key regulators of high-altitude adaptation in Himalayan populations by several studies (Beall et al. 2010; Simonson et al. 2010; Yi et al. 2010; Horscroft et al. 2017; Zhang et al. 2021). In fact, multiple chromosomal intervals within these loci presented values of T statistic that fall within top 5% of the related distribution (Ferraretti et al. 2024) (Supplementary Figure S20). Furthermore, some of the other genes ranking in the top 5% of such a distribution were found to be related to *Pathological hypertrophy signalling* pathways (Nakamura and Sadoshima 2018) (Figure 4.20), which were suggested as complex biological processes triggered by hypoxia thought the activation of Transient receptor potential canonical (TRPC) channels signalling mediated by HIF-1α up-regulation (Chu et al. 2012). In detail, these genes were: *AGT* (codifying for angiotensinogen, precursor of angiotensin II enzyme), *AGTR1* (also known as *ATR1*, encoding angiotensin II receptor type 1), *ADRA1B* (adrenoceptor alpha 1B), *TRPC6*, *RAPGEF4*, and *CAMK2B*/*D* (Figure 4.20).

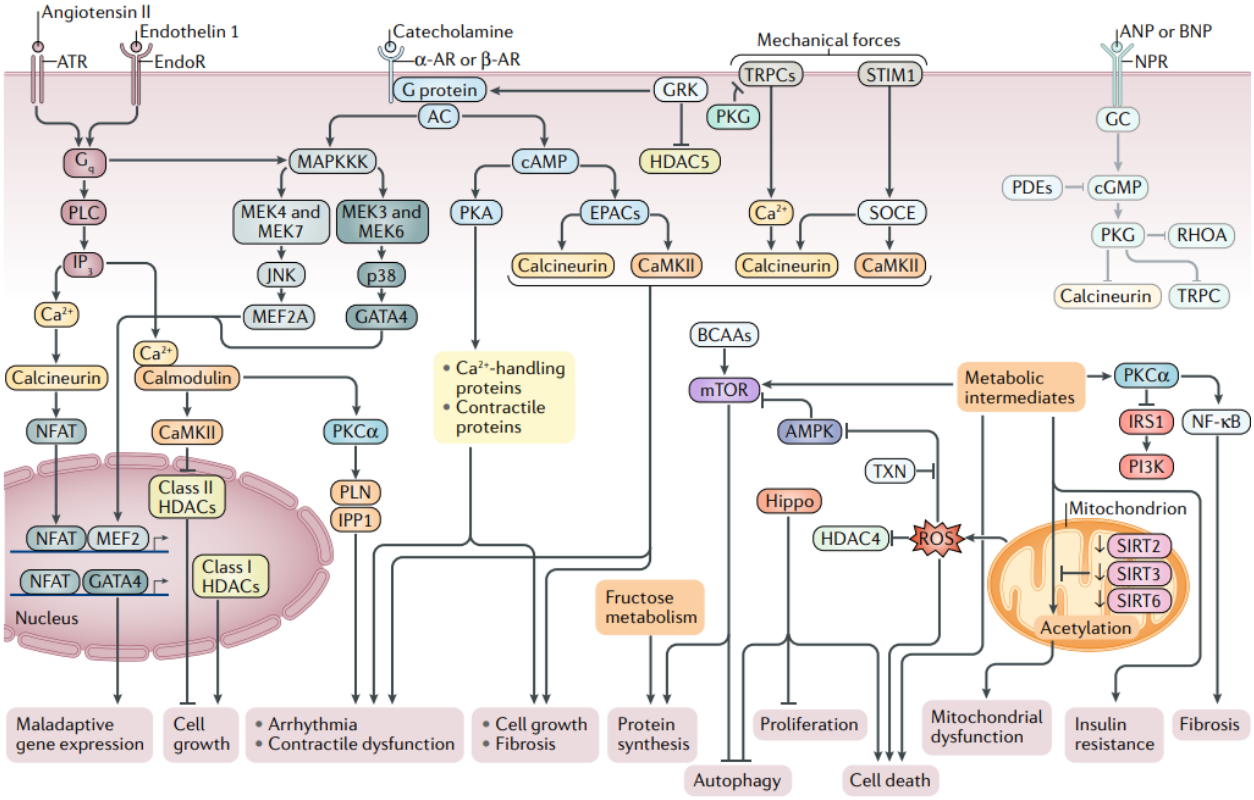


Figure 4.20. Schematic representation of the *Pathological hypertrophy signalling* pathways. Pathological hypertrophy (i.e. enlargement) of the heart is triggered by neuroendocrine hormones (e.g. angiotensin II, endothelin 1, catecholamines) and by mechanical forces, which trigger directly or indirectly increased reactive oxygen species (ROS) production and myocardial accumulation of metabolic intermediates, inducing cell death, fibrosis, and mitochondrial dysfunction. Natriuretic peptide-mediated increase in cyclic GMP (cGMP) levels activates protein kinase G (PKG) to inhibit cell growth, although the natriuretic peptide receptor (NPR) is desensitized in hypertrophy and heart failure. G protein-coupled receptor kinases (GRKs) are associated with β-adrenergic receptor (β-AR) desensitization, hypertrophy, insulin resistance, cell death, and mitochondrial dysfunction. Sustained activation of mechanistic target of rapamycin (mTOR) is detrimental because of autophagy suppression and deterioration of protein quality control mechanisms. Labels reported in the figure that indicate genes ranking in top 5% of the T statistic distribution: AGT (Angiotensin II); AGTR1 (ATR); ADRA1B (α-AR); CAMK2B/D (CAMKII); RAPGEF4 (EPACs); TRPC6 (TRPCs). The figure has been adapted from Nakamura & Sadoshima (2018).

As concerns results from the combined LASSI-*signet* analyses, among the significant gene-networks obtained we identified one belonging to the *Adrenergic signalling in cardiomyocytes* pathway, which resulted functionally related to the aforementioned functions since its chronic stimulation can induce pathological cardiac hypertrophy (Kanehisa and Goto 2000) (Figure 4.21 and Supplementary Table S4c). In particular, such a network comprised loci involved in the regulation of cardiac frequency, such as *CAMK2D* and *RYR2* (Sayers et al. 2022), as well as the *PPP2R5C* gene, a locus that has been previously proved to be indispensable for correct hearth development (Dyson et al. 2022) and has been suggested to regulate heart rate and conduction (Little et al. 2015; Jiao et al. 2025). Moreover, the network associated with the highest P-value obtained for the studied Tibetan population was found to belong to the *Phosphatidylinositol signalling system* pathway (Figure 4.21 and Supplementary Table S4c) and includes genes such as *PIKFYVE* and *PIP4K2A*, which are respectively linked to myocardial hypertrophy (Tronchere et al. 2017) and calcific aortic valve disease (Fang

et al. 2023). A further significant network belonging to the *ECM-receptor interaction* pathway was characterized by genes that have been proposed to regulate new vessel formation and angiogenesis (e.g. *LAMA1*, *LAMA2* and *DAG1*) (Hosokawa et al. 2002; Simon-Assmann et al. 2011; Biswas et al. 2022) (Figure 4.21 and Supplementary Table S2c), thus resembling findings previously advanced for Tibetan/Sherpa high-altitude populations (Gnecchi-Ruscone et al. 2018).

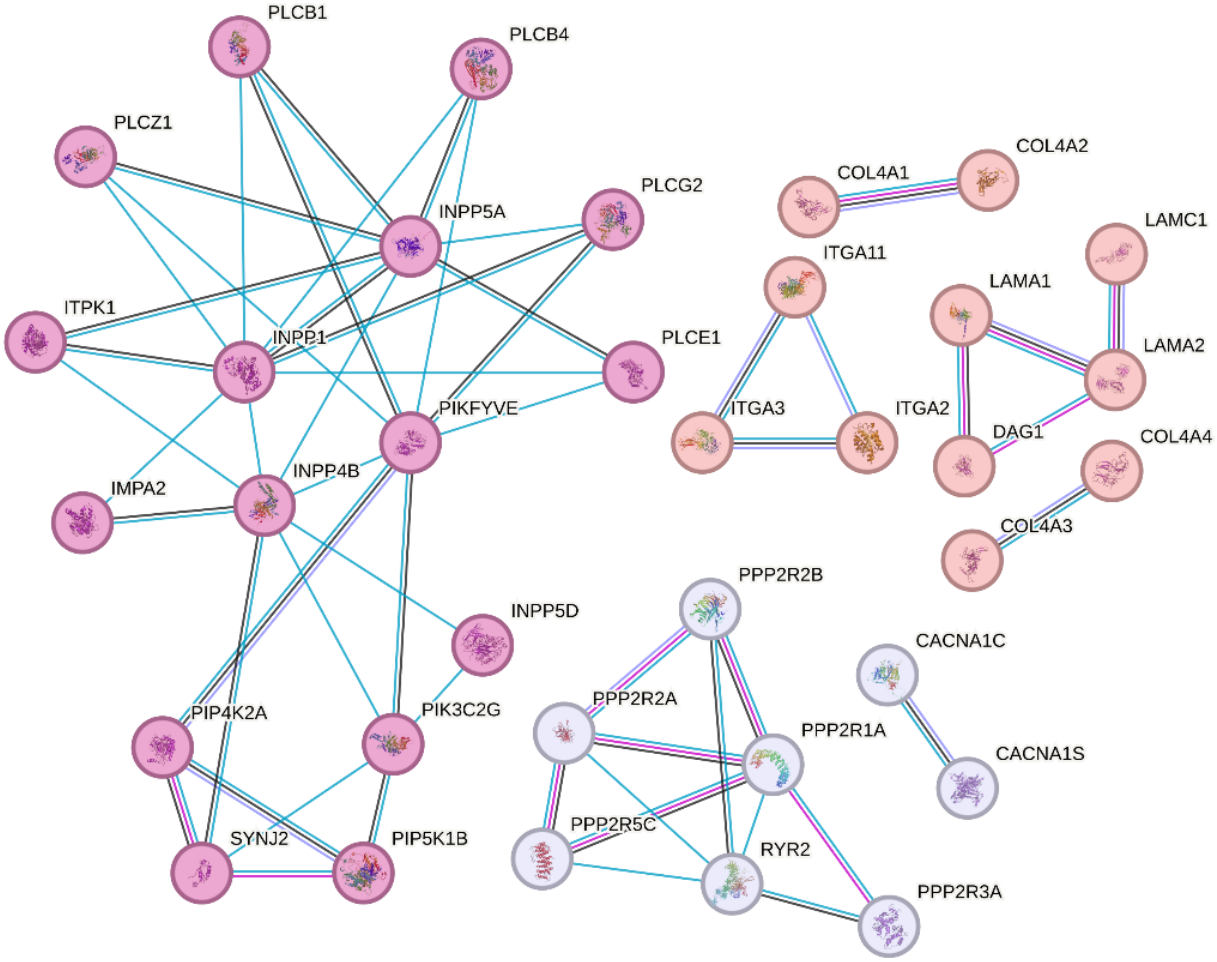


Figure 4.21. Gene-networks targeted by natural selection in the considered Tibetan population. All the reported gene-networks were build using the STRING tool (available at https://string-db.org/) and by setting the confidence score to predict gene-gene associations (i.e. edges) at 0.9. Functional associations among genes were inferred by considering co-expression (i.e. black edges), experimental (i.e. pink edges) and database (i.e. light-blue edges) evidence. Significant networks belonging to the *Phosphatidylinositol signalling system* and *Adrenergic signalling in cardiomyocytes* pathways are indicated as fuchsia and lilac circles, respectively. Genes included in the *ECM-receptor interaction* pathway are instead coloured in red.

Overall, the application of the *Trendsetter* method partially validated these results. In fact, all the genomic windows covering both the *PPP2R5C* and *PIKFYVE* genes presented variation patterns indicative of adaptive evolution according to such an approach (Figure 4.22). In particular, 41 out of 61 windows falling in the *PPP2R5C* gene were assigned to the soft selective sweep class, with the highest associated probability value

equal to ~0.88% (Figure 4.22a). The remining windows included in such a gene were instead classified as having experienced hard selective sweeps, with probability values reaching ~67% in the final portion of that locus (Figure 4.22a). In line with this picture, probability values for the soft selective sweep class inferred across the *PIKFYVE* gene were found to be consistently more elevated with respect to those associated to both the hard sweep and neutral classes (Figure 4.22b).

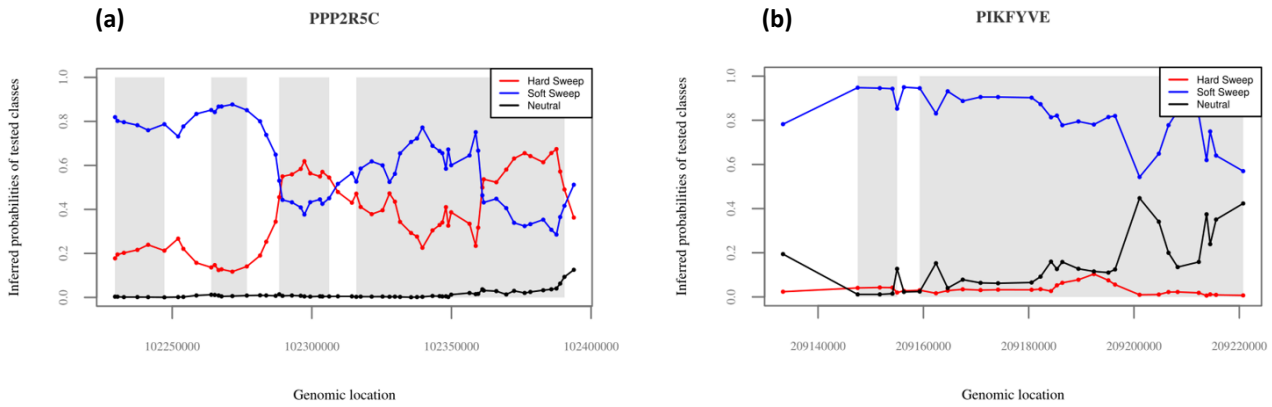


Figure 4.22. *Trendsetter* predicted probabilities at (a) *PPP2R5C* and (b) *PIKFYVE* genes in high-altitude Tibetans. Distribution of *Trendsetter* predicted probabilities for the considered genomic windows of being classified as having experienced hard sweeps (i.e. red curve), soft sweeps (i.e. blue curve) and neutral evolution (i.e. black curve) across the putative adaptive *PPP2R5C* and *PIKFYVE* genes. The x axis reports the positions of the central SNVs in each genomic window identified according to the *Trendsetter* approach, while on the y axis are shown the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent the genomic regions in the considered genes that present values indicative of the action of natural selection according to the T *LASSI* statistic. **(a)** Probabilities values associated to the soft selective sweep class are remarkably elevated across the entire gene (average of 60%). However, especially in the final portion of the gene the predicted probabilities associated to the hard sweep class overcome those of both the neutral and soft selective sweep classes. **(b)** *Trendsetter* probabilities indicative of adaptive evolution mediated by soft selective sweeps are constantly much more elevated across the entire gene with respect to those associated to the hard selective sweep and neutral classes.

Selection signatures pinpointed for the *DAG1* and *PIP4K2* genes were also validated by the *Trendsetter* approach (Supplementary Figure S21). In fact, although multiple windows falling in these genes were classified as genomic regions that evolved neutrally, in the initial portions of both loci we observed some signals supporting adaptive evolution according to the hard selective sweep model (Supplementary Figure S21). Finally, similarly to what observed for findings from the LASSI analysis, *Trendsetter* results were found to align with literature information supporting the adaptive role of both *EGLN1* and *VEGFC* genes in modulating high-altitude adaptations evolved by human populations from the Himalayas (Bigham et al. 2010; S. et al. 2010; Xu et al. 2011; Hu et al. 2017; Yang et al. 2017; Deng et al. 2019; Zheng et al. 2023) (Supplementary Figure S22).

4.5.3.2.1 Signatures of adaptive introgression from Denisovans

The Tibetan dataset analysed in the present thesis have been also used to investigate the contribution of archaic adaptive introgression events in the evolution of high-altitude adaptations of Himalayan populations, resulting in a recent publication (Ferraretti et al. 2024).

In particular, the rationale that guided this study was offered by previous evidence concerning the *EPAS1* gene, whose most frequent haplotype in Himalayan highlanders was proved to reduce their susceptibility to chronic mountain sickness and to have been introduced in the gene pool of their ancestors by admixture with Denisovans (Huerta-Sánchez et al. 2014; Zhang et al. 2021).

Based on this evidence, we aimed at further expanding the investigation of the impact of archaic introgression on more complex adaptive responses to hypobaric hypoxia evolved by populations of Tibetan/Sherpa ancestry. For this purpose, we combined results from a composite-likelihood method (i.e. VolcanoFinder) (Setter et al. 2020) with those of the gene network-based method (i.e. *signet*) implemented in the pipeline of analyses set up in the present thesis (Gouy and Excoffier 2020) to detect adaptive loci in introgressed chromosomal segments of Tibetan genomes (see Ferraretti et al. 2024 for methodological details).

According to such an approach, we identified multiple genes putatively involved in archaic introgression events. Specifically, the most validated signatures ascribable to adaptive introgression were detected for the *TBC1D1*, *RASGRF2*, *PRKAG2*, and *KRAS* genes. We supposed that these loci have plausibly contributed to the adaptive modulation of cellular hypoxic stress, angiogenesis and of certain cardiovascular traits (Kanehisa and Goto 2000; Kranenburg et al. 2004; Vichaiwong et al. 2010; Zhang et al. 2013) that may be protective in high-altitude Himalayan peoples.

The results obtained for these genes are in part schematized in Figure 4.23 and Supplementary Figures S23. For further information concerning these results, as well as the detailed methodological steps that have been carried out to pinpoint and validate them, see Ferraretti et al. (2024).

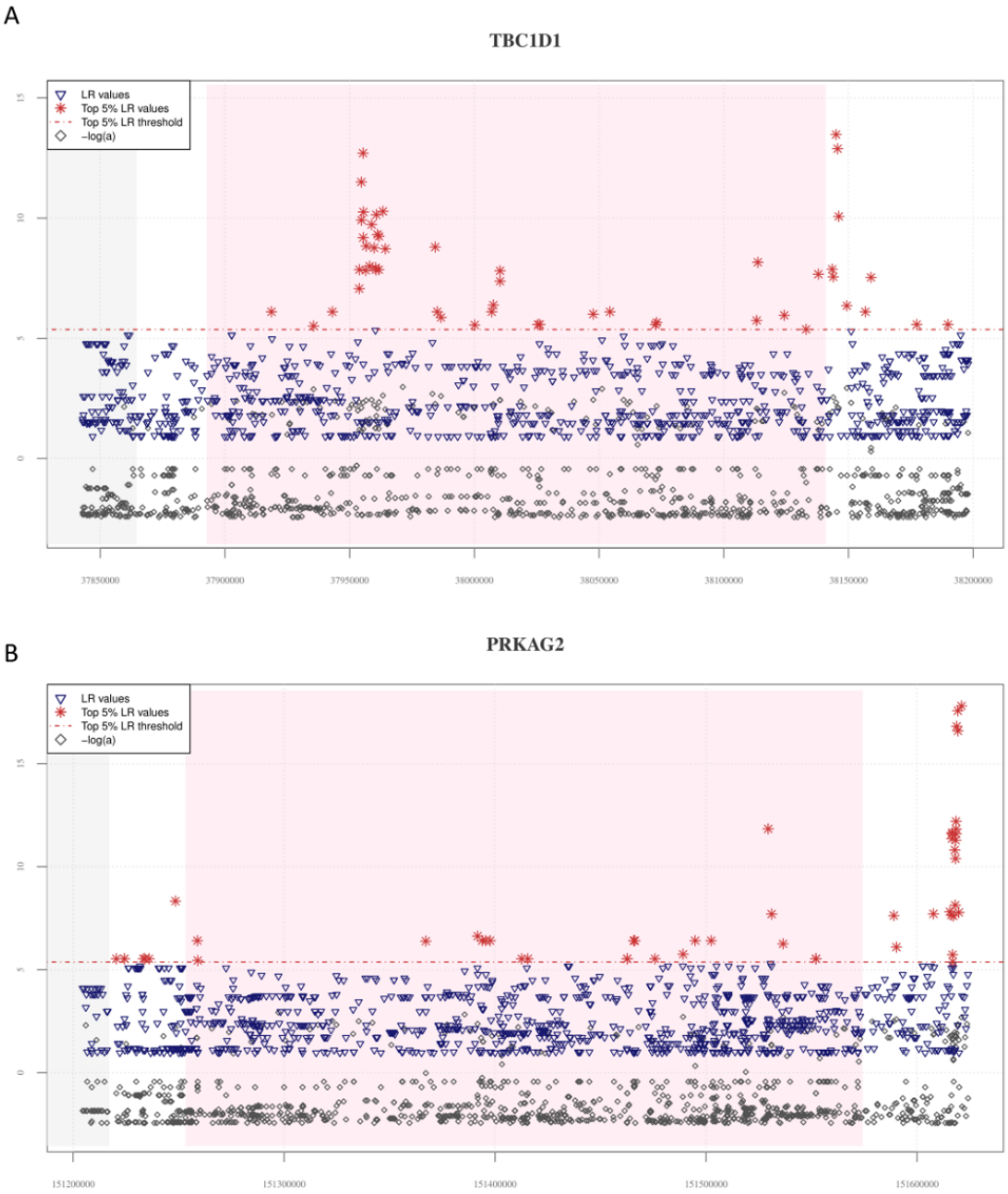


Figure 4.23. Distribution of *VolcanoFinder* statistics suggestive of adaptive introgressed loci across the *TBC1D1* and *PRKAG2* genes. On the x-axis are reported the genomic positions of each SNV, while on the y-axis are displayed the related statistics obtained. The pink background indicates the chromosomal interval occupied by the considered genes, while the grey background identifies those genes (i.e. *PGM2* in the *TBC1D1* downstream genomic region and the *RHEB* gene in the upstream *PRKAG2* region) possibly involved in regulatory transcription mechanisms. The dashed red line identifies the threshold set to filter for significant likelihood ratio (LR) values (i.e. top 5%) obtained according to the VolcanoFinder approach. **(A)** A total of 50 significant LR values (red stars) and collectively elevated −logα values (grey diamonds) were obtained for both *TBC1D1* and its downstream regions, suggesting widespread signatures of adaptive introgression. A remarkable concentration of significant LR values (i.e. 19 SNVs) was especially observed in the first portion of the gene. (B) The entire *PRKAG2* region comprise 46 SNVs showing significant LR values, with the greatest peaks being located in the downstream region associated to such gene. Peaks detected for the LR statistic are accompanied by peaks of −logα values. The figure has been taken from Ferraretti et al. (2024).

4.5.3.3 Signatures of polygenic adaptations in the Yakut population

As briefly mentioned in section 4.2, the developed pipeline of analyses was used also to investigate the genetic bases of complex adaptations evolved by high-latitude Eurasian populations to cope with selective pressures associated to sub-arctic climates (Ferraretti et al., under review). In detail, the described statistical approaches were firstly applied on whole genome sequence data available for the Yakut North East Asian population (Bergström et al. 2020) and, secondly, the obtained results were compared with those achieved by analysing a Russian "control" group, which was chosen due to its relatively shared ancestry with the Yakuts according to the performed population structure analyses (Figure 4.2, Supplementary Figures S1a and S9). We followed such a rationale with the aim of distinguishing between Yakut-specific adaptive events and adaptations plausibly shared among other high-latitude Eurasian populations, which have been long subjected to similar cold-induced selective pressures that however are more blunted with respect to those characterizing the Yakutia region. Furthermore, Sprime (Browning et al. 2018) and Haplostrips (Marnetto and Huerta-Sánchez 2017) analyses, two independent approaches able to identify putative chromosomal segments impacted by allele introgression from archaic species, were used to explore Neanderthal and Denisovan introgression in the Yakut genomes, then crosschecking the obtained results with those from the selection tests in order to investigate potential patterns of adaptive introgression.

4.5.3.3.1 Yakut-specific adaptations and selective events shared among high-latitude Eurasian populations

Among the biological functions/genes pointed out by the adopted approaches and likely related to adaptive events evolved to cope with extreme cold environments, a significant gene network including the *RCAN2*, *HIF1A*, *ATP1B2*, *PIK3R2*, *RXRA*, *SRC*, and *THRB* loci and belonging to the *Thyroid hormone signalling* pathway was observed exclusively in the Yakut population (Figure 4.24, Supplementary Table S4c). These results may indicate adaptations evolved specifically in such human group and modulating thyroid functioning, in accordance with previous evidence suggesting the crucial role of thyroid hormones in regulating physiological responses to cold stress (Tsibulnikov et al. 2000).

Similarly, we identified a significant Yakut-specific gene network belonging to the *AGE-RAGE signalling pathway in diabetic complications*, which included genes (e.g. *NFKB1* and *PRKCD*) whose functions were previously found to be related to the regulation of glucose homeostasis and the development of insulin resistance (Bezy et al. 2011; Chen et al. 2013; Guo et al. 2021) (Figure 4.24, Supplementary Table S4c). In line with this picture, also the *HTR1F* gene, which has been already proposed to be related to similar functions (Almaça et al. 2016), was found to belong to the *Serotonergic synapse* putative adaptive pathway identified for the Yakuts (Supplementary Table S4c).

Moreover, a gene network observed in both the Yakut and control groups and belonging to the *PI3K-Akt signaling* pathway included some loci (e.g. *INSR* and *EFNA5*) involved in the modulation of insulin secretion, binding, and activation of insulin downstream responses (Konstantinova et al. 2007; Yang Chen et al. 2019) (Figure 4.24, Supplementary Tables S4c).

Two additional significant gene networks belonging to the *Glycerolipid metabolism* and *cGMP-PKG signalling* pathways were also detected in both the examined populations, suggesting that natural selection anciently targeted these complex biological traits, as well as the genes involved in their modulation, long before the divergence of these human groups from a common ancestral gene pool. Finally, after extensive literature review, we assessed that some genes (e.g. *NTRK1, GPAT3, CAMK2D,* and *CLDN10*) identified within multiple significant Yakut-specific networks were previously proved to exhibit altered expression in mice Brown Adipose Tissue (BAT), white adipose tissue, and liver after cold exposure (Shore et al. 2013; Wang and Wahl 2014; Labbé et al. 2015; Ghosh et al. 2021) (Supplementary Table S5).

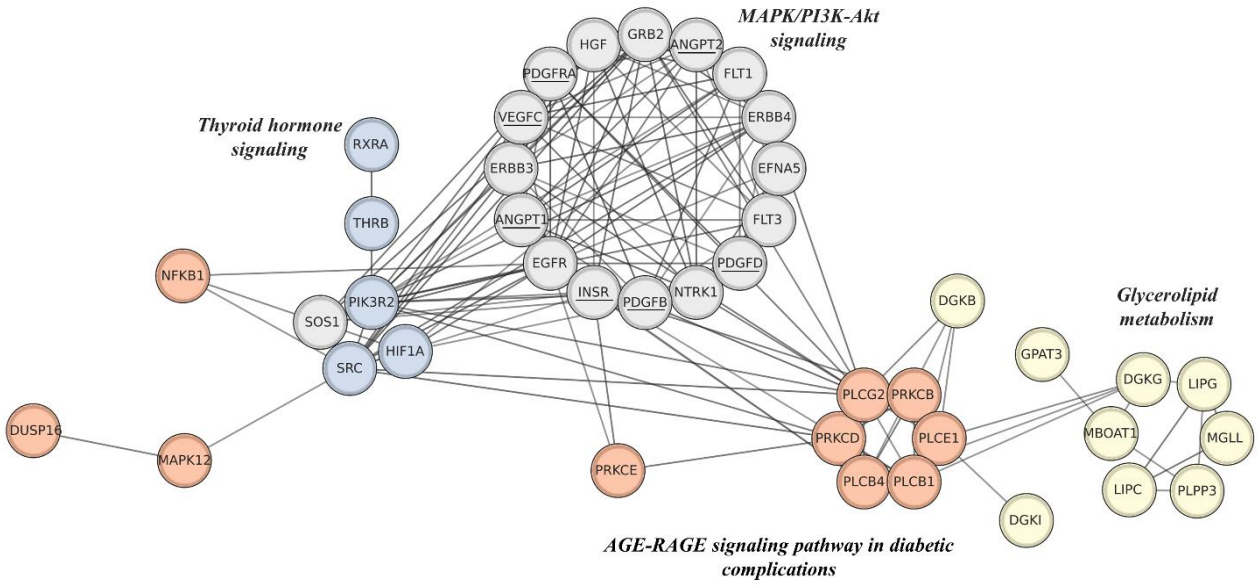


Figure 4.24. Gene-networks targeted by natural selection in the Yakut population. Networks of genes belonging to the *Thyroid hormone signalling* pathway and to the *AGE-RAGE signalling pathway in diabetic complications*, whose variation patterns resulted significantly impacted by the action of natural selection exclusively in the Yakut population, are displayed as light blue and orange circles, respectively. Significant gene networks within the *MAPK/PI3K-Akt signalling* and *Glygerolipid metabolism* pathways detected in both Yakut and control human groups are instead shown with grey and yellow circles, respectively. All the genes reported in the figure and included in the *Glycerolipid metabolism* pathway showed adaptive evolution in both Yakuts and control groups, while only the underlined genes in the *MAPK/PI3K-Akt signalling* pathway presented the same shared pattern. Gene networks were built using the STRING tool (available at https://string-db.org/) by setting the confidence score to 0.7. Genes belonging to the same pathway were displayed as circular networks (i.e. adopting the circular layout in Cytoscape version 3.10.3) in order to highlight the functional connections inferred among different significant networks/pathways.

Among them, six genes (e.g. *ADCY3, GPAT3, INSR, PDE4D, PDGFRA,* and *PLCB1*) were found to be included also in significant networks observed for the Russian control population (Supplementary Table S5). As concerns instead the results validated by the *Trendsetter* approach, *SRC* and *RCAN2* genes were found to be supported as putative adaptive loci in the Yakut population (Supplementary Figure S24). More in detail, extremely low probabilities of being classified as neutral regions when compared with those relative to the hard and/or soft selective sweep classes were obtained for several genomic windows included in these genes.

For instance, multiple *SRC* chromosomal intervals presented neutral output probabilities of ~ 0.01%, compared to ~ 0.84% and to ~ 0.13% of being classified as having experienced soft and hard selective sweeps, respectively (Supplementary Figure S24). Comparable results were also obtained for multiple genes belonging to the *AGE-RAGE signalling pathway in diabetic complications* and to the *Serotonergic synapse* pathway (i.e. *HTRF1*, *NFKB1*, *PRKCB*, *PRKCD*) (Figure 4.25, Supplementary Figures S24 and S25).

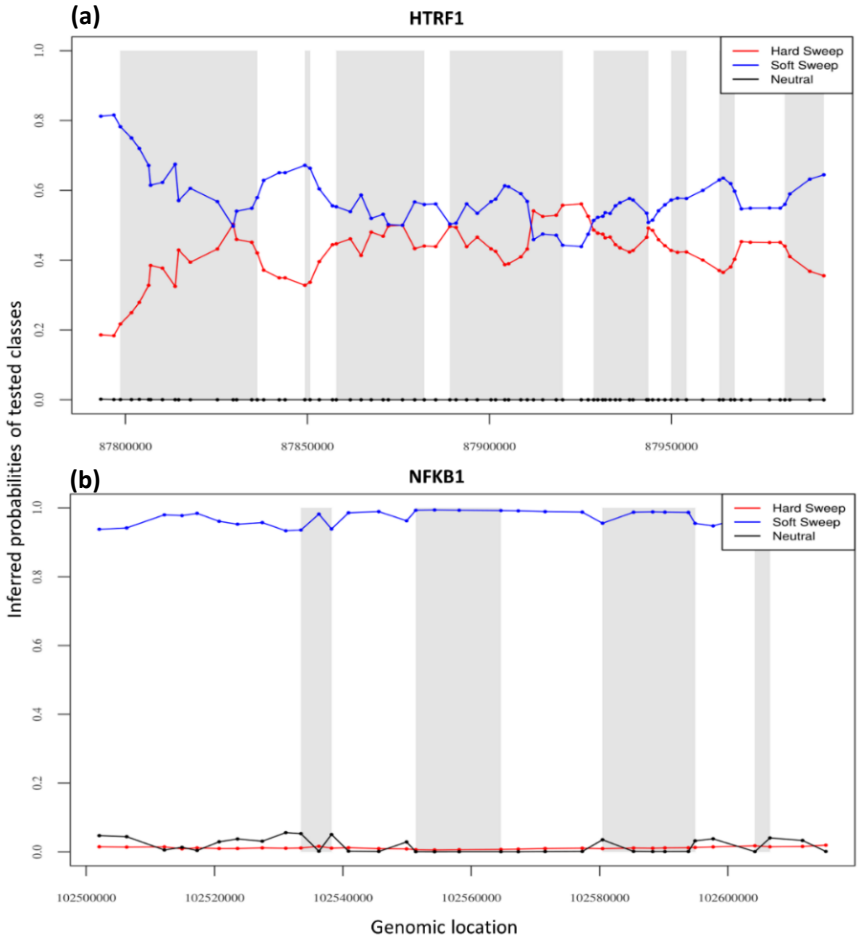


Figure 4.25. *Trendsetter* predicted probabilities at (a) *HTRF1* and (b) *NHKB1* genes in the Yakut population. Distribution of *Trendsetter* predicted probabilities for the considered genomic windows of being classified as having experienced hard sweeps (i.e. red curve), soft sweeps (i.e. blue curve) and neutral evolution (i.e. black curve) across the putative adaptive *HTRF1* and *NFKB1* genes, respectively. In both the plots, the x axis reports the positions of the central SNVs in each genomic window identified according to the *Trendsetter* approach, while the y axis shows the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent genomic regions in the considered genes that present values indicative of the action of natural selection according to the T *LASSI* statistic. **(a)** The predicted probabilities inferred for all the genomic windows falling within the *HTRF1* gene and associated to the soft selective sweep class are substantially more elevated with respect to those computed for the neutral class, suggesting that variation patterns at this locus have been intensively shaped by selective events. More in detail, predicted probabilities associated to the soft sweep class are predominant across the entire gene (i.e. ranging from 0.8 to 0.5), except for a region in the middle of it, in which predicted probabilities associated to the hard sweep class overcome them. **(b)** Predicted probabilities associated to the soft sweep class were outstanding across the entire *NFKB1* gene with respect to those computed for both the hard sweep and neutral scenarios, reaching values constantly higher than 0.90, and thus supporting the occurrence of selective events at this gene conformed with the soft selective sweep model.

As regards putative adaptive genes belonging to the *Glycerolipid metabolism*, *PI3K-Akt signalling*, and *cGMP-PKG signalling* pathways, the *GPAT3*, *ERBB4*, *PRKCB* and *PRKG1* genes were found to be supported by the *Trendsetter* approach as having evolved adaptively in the Yakut population (Supplementary Figures S24 and S25).

4.5.3.3.2 Signatures of adaptive introgression from archaic human species

The crosscheck between results obtained with the *Sprime* and LASSI-*signet* approaches (for methodological details, see the attached manuscript by Ferraretti et al., under review) allowed us to shortlist several genes participating to putative adaptive networks, such as those belonging to the *Glycerolipid metabolism* and *Thyroid hormone signalling* pathways, and whose variation patterns have been shaped by archaic introgression.

Interestingly, the expression of some of these genes (i.e. *ADCY3*, *CAMK2D*, *PDE4D*, *KCNN2*, *EGFR*, and *IMPA2*) was previously proposed to be altered in mouse BAT after cold exposure (Shore et al. 2013) (Supplementary Table S5).

To further validate these results and to explicitly test for the archaic source responsible for the observed introgression signatures, we run the Haplostrips algorithm to compare putative adaptive introgressed haplotypes at Yakut and control groups with those of three different archaic genomes (i.e. Neanderthal Altai, Neanderthal Vindija, and Altai Denisovan), as well as with those observed at the Yoruba outgroup population. A population of African ancestry was indeed chosen as the outgroup by assuming that it has experienced no/negligible Neanderthal and/or Denisovan admixture.

Overall, such an approach corroborated the occurrence of adaptive introgression events at the *ADCY3*, *PLPP3*, *ATP1B2* and *CD247* loci (Figure 4.26, Supplementary Figures S26, S27 and S28). The most compelling results were obtained for the *ADCY3* gene, which turned out to be extremely conserved among Yakut and control populations and was previously proposed to be involved in the modulation of polygenic adaptive events able to improve thermogenesis (Sazzini et al. 2020). In detail, most of the Yakut and control *ADCY3* haplotypes were found to present the lowest number of pair-wise differences with respect to the Denisovan genome when compared with the outgroup population of African ancestry (Figure 4.26a).

An even more clear picture of tight genetic affinity between Yakut/control and Denisovan haplotypes was depicted by limiting such a comparison to the *ADCY3* chromosomal intervals pointed out by the *LASSI* analysis as targeted by positive selection (Figure 4.26b and c).

Similar patterns were observed also for the *PLPP3*, *ATP1B2* and *CD247* genes, but involving alleles of putative Neandertal origin (Supplementary Figures S26, S27 and S28).

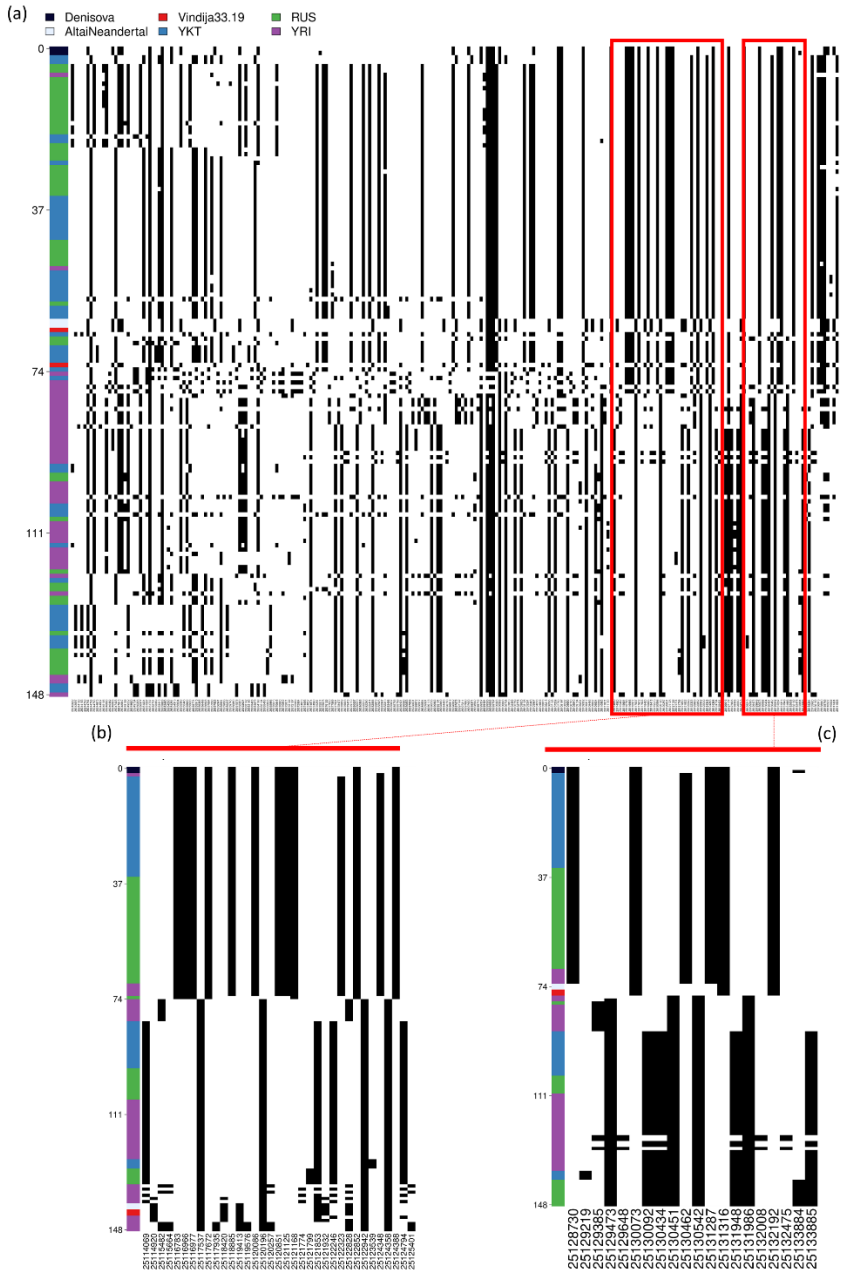


Figure 4.26. Structure of the putative adaptive introgressed *ADCY3* haplotypes. **(a)** Haplostrips plot showing patterns of *ADCY3* haplotype diversity among modern human populations (i.e. Yakut, Russian, Yoruba) and archaic samples (i.e. Denisovan Altai, Neanderthal Altai and Neanderthal Vindija). Haplotypes (in rows) are ranked according to the number of differences with respect to the Denisovan sequence. SNVs are displayed in columns with the ancestral/derived states reported in white/black colours. Most haplotypes belonging to both Yakut (i.e. blue) and control (i.e. green) populations present the smallest amount of pairwise differences with the Denisovan ones (i.e. black), even with respect to Vindija (i.e. red) and Altai (i.e. light blue) Neanderthals. Haplotypes of the Yoruba outgroup population (i.e. purple) cluster mainly at the bottom of the plot, attesting a remarkable diversity when compared to the Denisovan sequence and showing opposite patterns with respect to non-African populations. Red squares localize the positions of overlapping putative adaptive *ADCY3* chromosomal segments identified by *LASSI* analysis. **(b-c)** Haplostrips plots build for overlapping genomic windows presenting patterns indicative of Yakut adaptive evolution according to the T *LASSI* statistic and falling within *ADCY3*. In both genomic regions, similar patterns of variation are detectable with respect to those observed across the entire *ADCY3*. In detail, most Yakut and control haplotypes are extremely conserved, presenting the smallest amount of pairwise differences with respect to the Denisovan one, thus suggesting the occurrence of archaic adaptive introgression at the *ADCY3* in both groups.

4.5.3.3.3 Haplotype and genotype frequencies at Yakut candidate adaptive loci

By focusing on selective events specifically occurred in the Yakut population, we explored allele composition and frequencies of putative adaptive haplotypes at genes included in significant networks belonging to the *Thyroid hormone signalling* and *AGE-RAGE signalling pathway in diabetic complications*, as well as at adaptive loci belonging to different pathways but known to play a role in these biological functions.

For the *THRB*, *SRC*, *RCAN2*, *INSR*, *NFKB1*, *PRKCB*, and *HTR1F* genes, we observed significant differences in the frequency of multiple SNVs (i.e. adjusted FDR < 0.05) between the Yakut and control populations. Interestingly, some of these variants have been previously reported in the GTEx database as expression Quantitative Trait Loci (eQTLs) able to regulate the expression of the identified candidate adaptive genes in diverse tissues (e.g. white adipose tissue, skeletal muscle, thyroid and blood) according to studies aimed at examining the physiological responses activated by cold exposure (Ribeiro et al. 2001; Blondin et al. 2015; Xu et al. 2019; Zekri et al. 2021) (Figure 4.27a).

Overall, the frequencies of putative adaptive haplotypes containing these eQTLs were found to be remarkably higher in the Yakut population (i.e. ranging from 0.79 to 0.92) with respect to those observed in the control group (Supplementary Table S6). In particular, such a pattern was observed for all the tested genes except for *THRB*.

However, when considering genotype instead of haplotype frequencies, homozygotes for all these candidate adaptive eQTLs were consistently more represented in the Yakuts with respect to the control population (Figure 4.27b). In detail, candidate adaptive eQTLs at *INSR* and *THRB* have been previously associated, in the homozygous state, to increased expression of these genes in white adipose tissue, while those at *HTR1F* and *SRC* were found to increase expression in white adipose tissues/thyroid and liver, respectively, reducing instead the expression of *SRC* in the pancreas (Supplementary Figure S29) (info available at https://gtexportal.org/home/). Finally, homozygotes for candidate adaptive eQTLs at *RCAN2* have been demonstrated to decrease expression of such a gene in the skeletal muscle, as well as of *PLA2G7* in the blood (Supplementary Figure S29) (info available at https://gtexportal.org/home/).

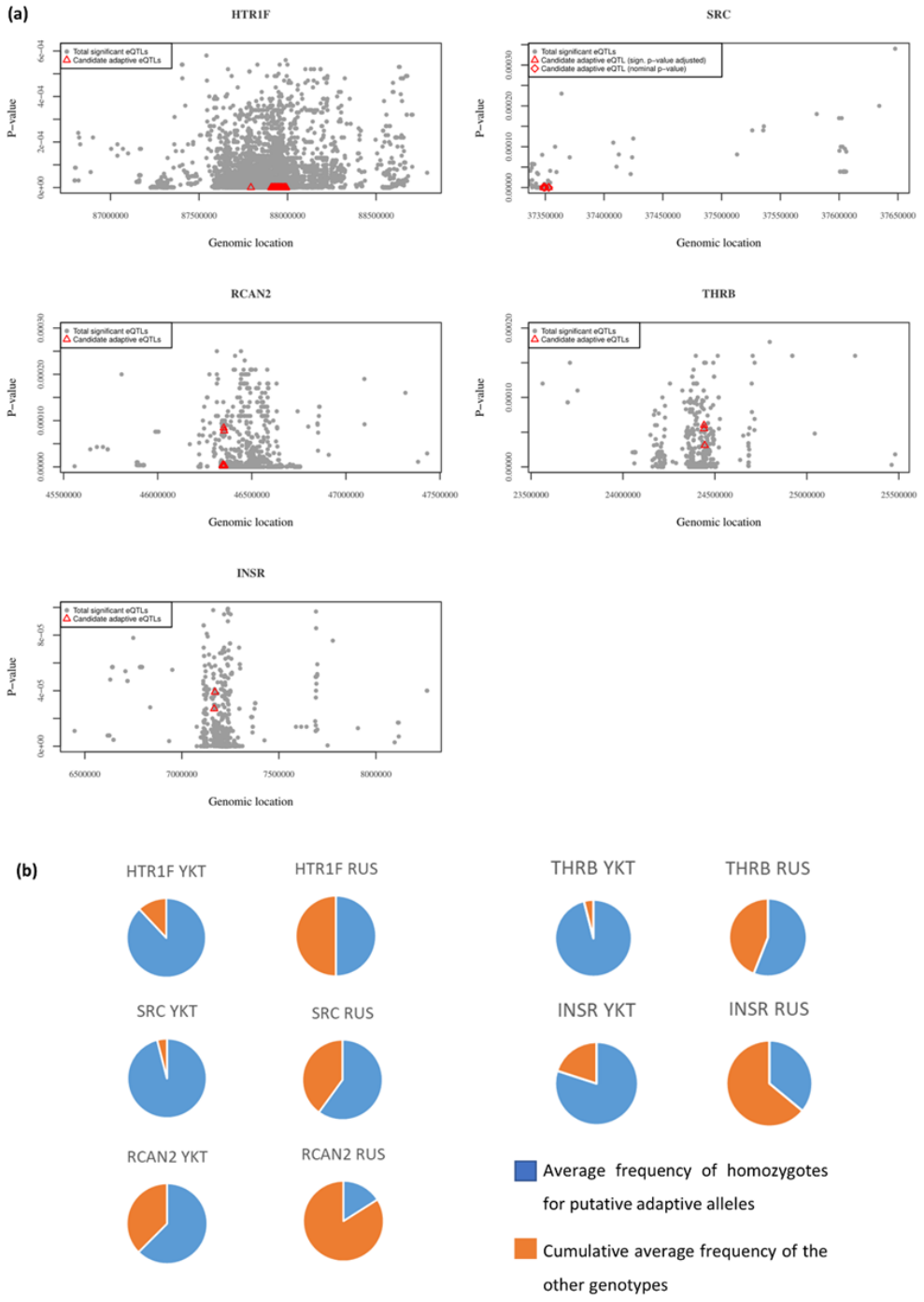


Figure 4.27. Candidate adaptive eQTLs on the *HTR1F*, *SRC*, *RCAN2*, *THRB*, and *INSR* genes. **(a)** The x axis reports the genomic location of eQTLs at the considered putative adaptive genes, while in the y axis are displayed the relative P-values for their association with altered gene expression levels. eQTLs influencing the expression of the *HTR1F*, *SRC*, *RCAN2*, *THRB*, and *INSR* genes in different district of the body are reported as grey circles, while those affecting the expression of the same genes in target tissues potentially involved in the modulation of physiological responses to cold exposure (i.e. cold target tissues: adipose tissue, thyroid, liver and pancreas, skeletal muscle) and that were characterized by significant shifts in frequency between Yakuts and the control population are displayed as red triangles and rhomboids. **(b)** Pie chart displaying the average frequency of homozygotes for the putative adaptive eQTL alleles (i.e. blue) and of the cumulative average frequency of the remaining genotypes in Yakut (YKT) and control (RUS) populations (i.e. orange). Homozygote genotypes associated to the augmented expression of the *HTR1F*, *SRC*, *INSR*, *THRB* genes, as well as to the decreased the expression of the *SRC* and *RCAN2* genes in cold target tissues are much more frequent in the Yakuts with respect to the controls.

4.5.4 Adaptive evolution of populations from the American macro-area

4.5.4.1 Signatures of polygenic adaptations in indigenous Mexican populations

When subjected to the LASSI-*signet* combined analysis, the Maya indigenous population from Mexico was found to be characterized by a significant network belonging to the *Insulin Resistance* pathway containing genes such as *IRS1*, *PTPRF* and *PRKCZ* that are known to regulate insulin signalling processes (Sayers et al. 2022) and whose enhanced/reduced expression was associated with the development of insulin resistance/Type 2 diabetes (De Lorenzo et al. 2013; Zou et al. 2013) (Figure 4.28a and Supplementary Table S4d).

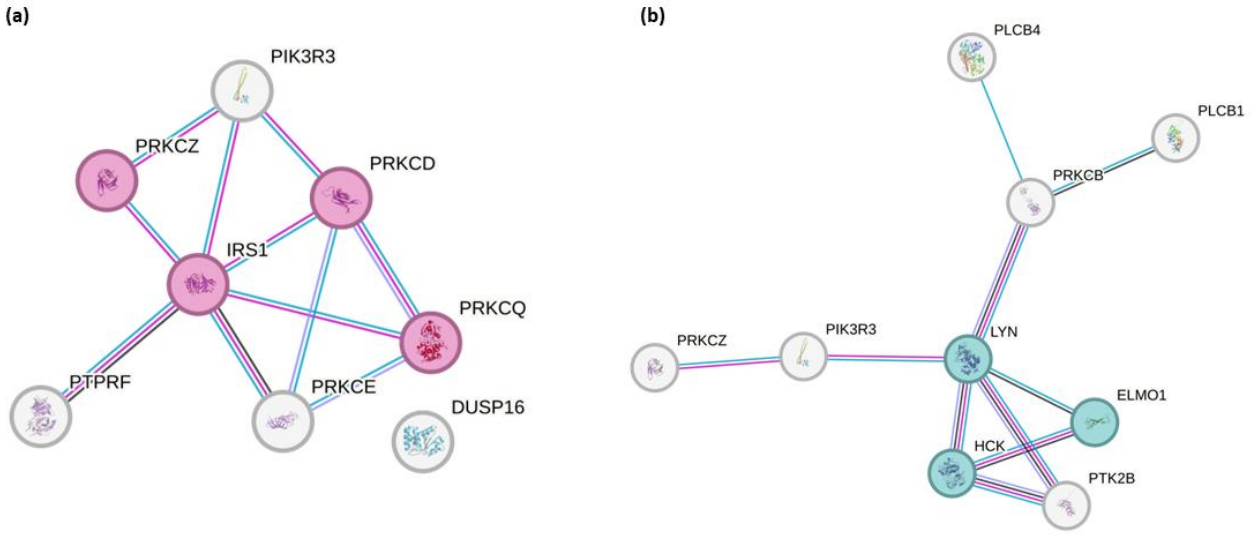


Figure 4.28. Gene-networks targeted by natural selection in indigenous populations from Mexico. The reported networks were build using the STRING tool (available at https://string-db.org/) and by setting the confidence score to predict gene-gene associations (i.e. edges) at 0.9. Functional associations were inferred by considering co-expression (i.e. black edges), experimental (i.e. pink edges) and database (i.e. light-blue edges) evidence. **(a)** Significant network belonging to the *Insulin Resistance* pathway was observed for the Maya population. Genes reported in pink are enriched in the *Negative regulation of insulin receptor signalling* pathway annotated in the Gene Ontology database (i.e. the most significant enriched term characterized by FDR = 1.15e-05). **(b)** Putative adaptive genes belonging to the *Chemokine signalling* pathway were observed for the Nahua and Pima populations. Loci for which the involvement in immune related functions is well established by previous evidence (Smolinska et al. 2011; Li et al. 2014; Sarkar et al. 2017) are reported as sea green circles.

Moreover, the *FLT3* gene, included in the *Acute myeloid leukemia* pathway, was proposed to modulate similar pathological phenotypes in mice (Chilton et al. 1995) (Supplementary Table S4d). However, such a gene was found to be included in other putative adaptive gene-networks obtained for Eurasian populations and has been confirmed by the *Trendsetter* approach to have adaptively evolved in the CEPH and Yakut groups as well (Supplementary Table S4b-c). Always in Maya population, we detected a significant network belonging to the *Axon guidance* pathway and including the *PPP3CA* gene, a locus that has previously proposed to modulate adaptive responses during *T. Cruzi* infection in Native populations from Brazil and to show genetic signatures

ascribable to the action of natural selection in such human group, as well as in Native populations form South Mexico (Ojeda-Granados et al. 2022; Couto-Silva et al. 2023). Accordingly, the *THBS1* gene, a locus that has been proposed to play a role in *T. cruzi* early infection (Nde et al. 2012) and to include risk variants for severe forms of Malaria (Kanchan et al. 2015), presented multiple windows associated with top1% T LASSI values (Supplementary Figure S30). Specifically, the investigation of *PPP3CA* selection signatures has been addressed in a study aimed at investigating the genetic bases of local adaptations evolved by the Wichí population from South America (Ferraretti et al., in preparation).

Similarly to what observed for the Mayas, we identified multiple putative adaptive networks belonging to immune-related pathways also for the Pima and Nahua populations from Northern and Central Mexico. In detail, these human groups presented pervasive selective signatures at the *Chemokine signalling* pathway, which included genes such as *LYN*, *ELMO1* and *HCK* that were proved to activate and/or regulate immune responses against different pathogens (Nelson et al. 2009; Li et al. 2014; Sarkar et al. 2017) (Figure 4.28b and Supplementary Table S4d).

In line with this picture, *VAV2* and *VAV3* genes, known to modulate immune-system responses (Roth et al. 2016) were found included in a significant LASSI-*signet* network identified for the Nahuas (Supplementary Table S4d). We also detected a gene-network belonging to the *Fc gamma R-mediated phagocytosis* pathway for both Pima and Nahua populations, although in the latter group it reached only a barely significant P-value (0.051) (Supplementary Table S4d).

Finally, the Nahua population presented some specific signatures involving loci within the *Morphine Addiction* pathway (Supplementary Table S4d) which were previously suggested to enhance biological tolerance towards addictive compounds and their detrimental effects in indigenous populations from Central Mexico (Ojeda-Granados et al. 2022).

Part of the described selection signals were further validated by the *Trendsetter* approach. In particular, for the Maya population, we observed some genomic windows classified as having experienced hard/soft selective sweeps and falling within the *PTPRF, PRKCZ, FLT3, PPP3CA* and *THBS1* genes (Figure 4.29, Supplementary Figure S31) as well as for other genes (e.g. *RHOA*, *NFATC3*) included in significant networks (Supplementary Table S4d).

Accordingly, several genes participating to the biological functions proposed as significantly impacted by the action of natural selection in both Nahua (e.g. *KCNJ6*, *GNAS*, *VAV2 and VAV3*) and Pima (e.g. *HCK*) populations were further validated (Supplementary Table S4d).

Finally, signatures of adaptive evolution for the *HCK* gene were supported by the *Trendsetter* approach also in the Maya, even though the gene-network within the *Chemokine signalling* pathway, to which this gene belongs, was associated to a non-significant P-value in this population (i.e. 0.058) (Supplementary Table S4d).

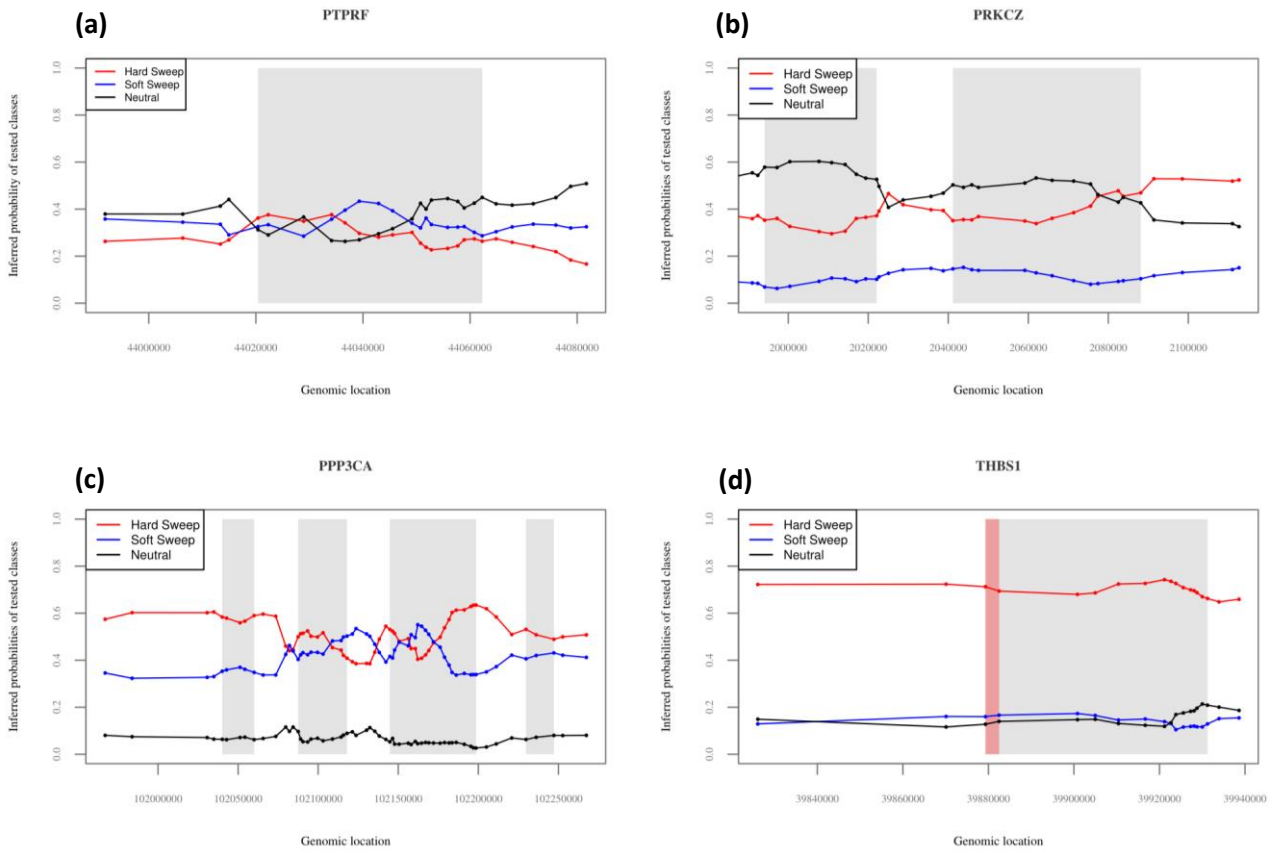


Figure 4.29. *Trendsetter* predicted probabilities at genes related to insulin resistance (a-b) and parasitic infections (c-d) in the Maya population. Distribution of *Trendsetter* predicted probabilities for the considered genomic windows of being classified as having experienced hard sweeps (i.e. red curve), soft sweeps (i.e. blue curve) and neutral evolution (i.e. black curve) across the putative adaptive *PTPRF* **(a)**, *PRKCZ* **(b)**, *PPP3CA* **(c)** and *THRS1* **(d)** genes. For all the plots, the x axis reports the positions of the central SNVs in each genomic window identified according to the *Trendsetter* approach, while the y axis shows the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent genomic regions in the considered genes that present values indicative of the action of natural selection according to the T *LASSI* statistic. **(a)** Some of the genomic windows falling in the middle part of the *PTPRF* gene are assigned to the soft selective sweep class, suggesting that variation patterns at this locus have been shaped by selection events. **(b)** Probabilities associated to the hard sweep class reach moderate values across the gene and then overcome those related to the neutral class in the final portion of the *PRKCZ* locus. **(c**) Probability trends associated to both hard and soft sweep classes are much more elevated with respect to those computed for the neutral class across the entire *PPP3CA* gene. **(d)** Predicted probabilities associated to the hard selective sweep class were outstanding with respect to those associated to both soft sweep and neutral scenarios, reaching values with an average of 0.70, thus supporting strong signatures ascribable to the action of natural selection on the *THRS1* locus (red rectangle), as well as on its surrounding genomic regions.

4.5.4.2 Polygenic bases of high-altitude adaptations in Andean Aymara

As a further population representative of genomic variation observable in the American macro-area, a cohort of 24 individuals belonging to the high-altitude Bolivian Aymara ethnic group (Figure 4.4) was extensively analysed as reported in the study by Ferraretti et al. (2025), which was aimed at uncovering the genetic bases of convergent complex adaptations evolved by high-altitude populations from the Andes and the Himalayas. Specifically, in such a study, which has been already launched before the implementation of the present PhD project, multiple selection scans were performed (i.e. H12, nSL) and combined with the *signet* gene-network analysis, with the LASSI method implemented in the pipeline of analyses described so far being used to further validate the obtained results (Figure 4.30).

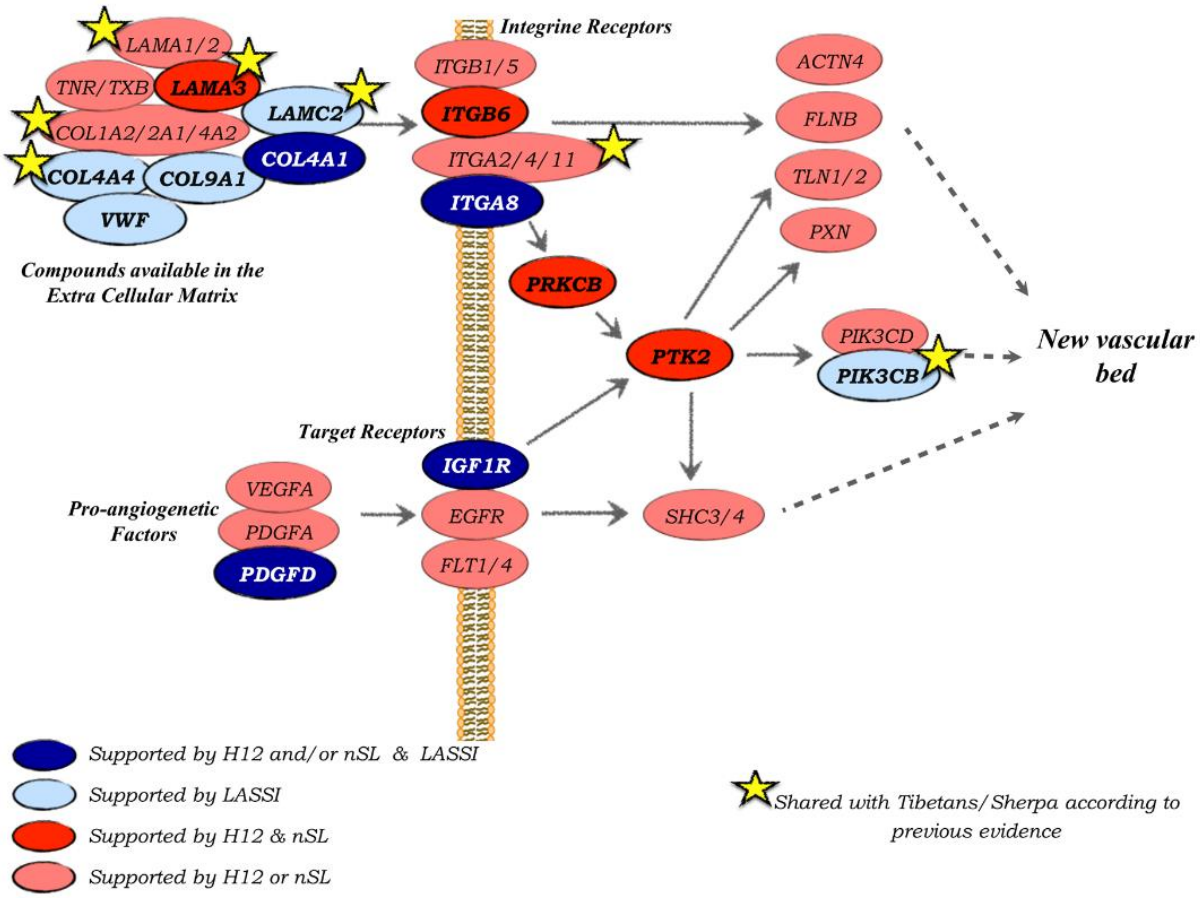


Figure 4.30. Scheme displaying functional interactions between adaptive genes belonging to the *PI3K-Akt signalling* and *Focal adhesion* pathways contributing to improved angiogenesis in Andean populations. During the early phases of angiogenesis, molecular compounds made available in the extracellular matrix, such as those building the basement vascular membrane (e.g. *COL4A1*) and angiogenetic factors (e.g. *VEGFA* and *PDGFD*), interact with integrin and target receptors (e.g. *IGF1R*) stimulating protein kinases (e.g. *PRKCB* and *PTK2*), which subsequentially activate those signalling cascades essential for the migration of endothelial cells and for the formation of new vascular structures. The yellow stars mark those genes previously identified as loci putatively mediating biological adaptation to hypobaric hypoxia in Tibetan/Sherpa populations (Arciero et al. 2018; Gnecchi-Ruscone et al. 2018; Deng et al. 2019), thus providing evidence for partial genetic convergence between Andean and Himalayan adaptive traits in addition to the remarkable convergence observed at the biological function/pathway level. The figure has been adapted from Ferraretti et al. (2025).

Such methodologies were applied on WGS data for 24 high-altitude Aymara individuals (Lindo et al. 2018) and on an imputed validation genotype dataset (i.e. 8,024,216 SNVs) comprising 130 individuals belonging to additional Bolivian Aymara groups, as well as from Aymara, Quechua, and Uros high-altitude Peruvian populations (Reich et al. 2009; Gnecchi-Ruscone et al. 2019). Furthermore, a control dataset including 24 individuals from three different Amazonian low-altitude groups (i.e. Ashaninka, Cashibo, and Shipibo) (Gnecchi-Ruscone et al. 2019) was assembled and subjected to the same analyses in order to retain only high-altitude specific selection signatures. For further details concerning the definition of the low-land control group, as well as the imputation of the validation dataset, see Ferraretti et al. (2025).

By following such a workflow, we demonstrated that high-altitude Andean human groups experienced pervasive selective events at angiogenic pathways, which resemble those previously attested for Himalayan populations despite partial convergence at the single-gene level was observed (Ferraretti et al. 2025). Such results are in part schematized in Figure 4.30 and in Supplementary Figure S32. Moreover, the outputs of the LASSI-*signet* combined analyses obtained for high-altitude Aymara are reported in Supplementary Table S4d.

5. DISCUSSION

In the present study, a pipeline of analyses capable to identify genomic signatures attributable to selective events involved in the evolution of complex adaptive traits was developed (Figure 3.2) and used to analyse a large panel of whole genome sequence data representative of several world-wide human populations (Figure 4.4).

At first, in order to identify groups presenting no internal genetic substructures and thus useful to be considered as reliable proxies of the ancestral gene pools that have long experienced specific selective pressures, genotype- (i.e. PCA and ADMIXTURE) and haplotype-based (i.e. ChromoPainter and fineSTRUCTURE) population structures analyses were performed (Figures 4.1, 4.2 and 4.3, Supplementary Figures S1-S4 and S6-S10). Genetically homogeneous populations identified by means of these approaches have been then subjected to the above-mentioned pipeline of analyses to investigate the contribution of polygenic adaptations to their adaptive history. In detail, the core of such an integrated approach is constituted by the combination of the likelihood-based LASSI (Harris and DeGiorgio 2020) and the network-based *signet* (Gouy et al. 2017) methods (Figure 3.2), which collectively allowed the identification of sets of genes that, as a whole, have been significantly impacted by pervasive natural selection and play a role in the modulation of complex biological traits/functions. This framework was especially useful to focus also on relatively weak selective signatures, which are those more likely contributing to the evolution of polygenic adaptations (Figures 4.9, 4.12, 4.17, 4.21, 4.24, 4.28, 4.30, Supplementary Figure S15 and Supplementary Table S4).

Results obtained with the LASSI-*signet* methods were further validated by means of the *Trendsetter* supervised machine-learning approach (Mughal and DeGiorgio 2019), which allowed us to corroborate signatures of adaptive evolution for a relevant fraction of the putative adaptive genes/biological functions pinpointed by the first two analyses (Figures 4.10, 4.11, 4.13, 4.14, 4.18, 4.19, 4.22, 4.25, 4.29, Supplementary Figures S16-S19, S21, S22, S24, S25 and Supplementary Table S4). To ensure the achievement of robust validations, we first trained the *Trendsetter* models and we evaluated their reliability (Figure 4.8, Supplementary Figures S11-S14) by generating thousands of genomic simulations with the software *discoal* (Kern and Schrider 2016), which incorporated inferences drawn by means of the SMC++ algorithm (Terhorst et al. 2017) and concerning population-specific Ne changes through time (Figure 4.6). Notably, Ne estimates obtained via the SMC++ analysis aligned with those reported by previous studies (Terhorst et al. 2017; Bergström et al. 2020), even though different individuals and different numbers of them per population were considered. Moreover, the trained *Trendsetter* algorithms demonstrated high accuracy and precision in identifying selective events, particularly in distinguishing them from a scenario of neutral evolution (Figure 4.8, Supplementary Figure S14, Supplementary Tables S2 and S3). Importantly, these results support the validity of this approach to confirm the selection signatures identified with the combined LASSI-*signet* methods (Supplementary Table S4), also according to the classification probabilities obtained, which were well-calibrated and further

corroborated the reliability of the implemented approach (Figure 4.7, Supplementary Figures S11, S12, and S13).

In addition, the contribution of archaic adaptive introgression events to the evolution of complex adaptive traits was investigated in some human groups that represent well-characterized case studies of human adaptation to extreme environmental settings, such as high-altitude populations of Tibetan ancestry and high-latitude Eurasian groups from different Russian and Northern Siberian regions (Figures 4.22, 4.25, Supplementary Figures S23, S26, S27 and S28). Results obtained for these case studies corroborated the hypothesis that inbreeding events having involved archaic hominins have played a non-negligible role in shaping the complex architecture of adaptive variability of the *H. sapiens* species.

The following paragraphs provide an in-depth examination of the biological relevance of the genes/functional pathways pointed out by the analyses mentioned above as putatively implicated in the modulation of the polygenic adaptations evolved by the considered populations. Such pieces of evidence are presented based on the type of the selective pressures that have been hypothesized as the driving forces responsible for the evolution of such adaptive responses (i.e. pathogen-, climate- and hypoxia-induced selective pressures) and by placing particular emphasis on the putative adaptive loci validated also by the *Trendsetter* approach.

5.1 Pervasive pathogen-driven selection signatures in world-wide human populations

Multiple studies have so far suggested that pathogen-driven selective events may have represented some of the predominant evolutionary processes that contributed to shape the biological adaptations evolved by modern human populations globally (Novembre and Han 2012; Cagliani and Sironi 2013; Couto-Silva et al. 2023). Overall, findings provided by the present study deeply agree with such evidence, as attested by the several immune-related gene-networks/loci showing selection signatures conform to adaptive evolution under a model of polygenic adaptation (Figures 4.9, 4.15, 4.16, 4.17, 4.28b-c, 4.29c-d and Supplementary Table S7).

For instance, when examining West African populations, we identified several genes (e.g. *HLA* loci, *VTN*, *THBS2*, *THBS3*, *THBS4*, and *FN1* genes) presenting selection signatures that could be related to the improvement of host immune responses to the highly endemic tuberculosis and malaria infections (Molina-Moya et al. 2018; Weiss et al. 2025) (Figure 4.9, Supplementary Table S4a). Among them, *VTN*, *IL15* and *HLA* genes were validated as putative adaptive loci in the Yoruba population also by the *Trendsetter* algorithm (Figures 4.10, 4.11 and Supplementary Figure S16).

In detail, the *VTN* gene codifies for the vitronectin, an adhesive glycoprotein that possesses anti-microbial properties thanks to its heparin-binding domain (Sayers et al. 2022). Remarkably, vitronectin is a serum protein captured by the malaria parasite in the stages of merozoites, trophozoite and schizont, according to a strategy proposed to enhance pathogenicity by avoiding the recognition of non-self-components by the host immune system (Figure 5.1) (Tougan et al. 2018; Tougan et al. 2020). Specifically, parasite-*VTN* contact is established at the *P. falciparum* antigen SERA5 (serine repeat antigen 5), in its P47 N-terminal domain, which directly

bind the vitronectin, as attested by their co-localization in the schizont and merozoite (Figure 5.1) (Tougan et al. 2018).

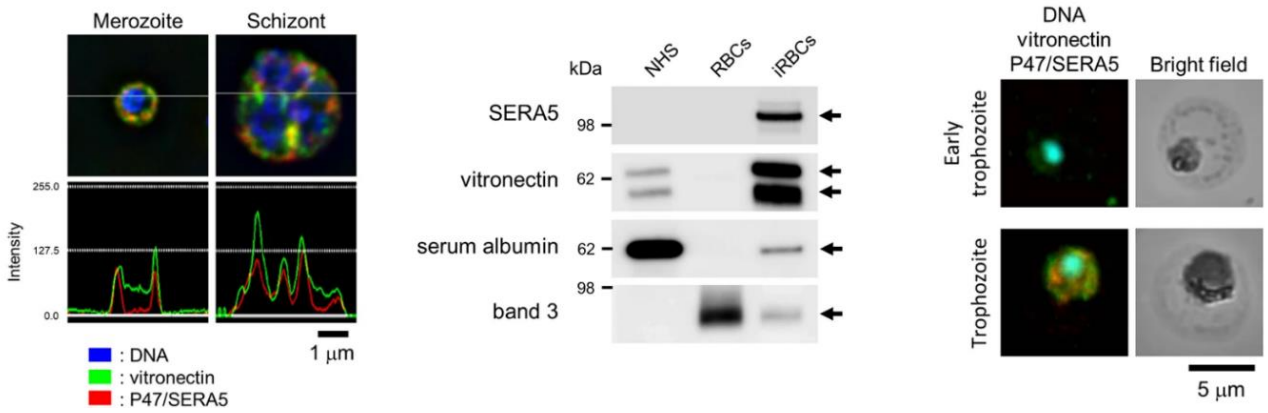


Figure 5.1. Co-localization of vitronectin and P47 domain of SERA5. From left to right: 1) representative immunofluorescence assay (IFA) images of merozoite and schizont stages under deconvolution microscopy. The upper panels show the localization of P47 domain/SERA5 (red colour) during merozoite and schizont stages, while the lower panels display the intensity of green (*VTN*) and red signals. 2) Western blotting of naïve human serum (NHS), red blood cells (RBC) lysate, and infected red blood cells (iRBC) lysate. Band 3 is RBC-specific protein and was used as loading control. 3) Representative IFA images of early trophozoite and trophozoite. The figure has been adapted from Tougan et al. (2018).

As concerns instead the *HLA* genes, the class II of these loci is known to play crucial roles in host defence from different pathogens (Sayers et al. 2022). In fact, these genes (especially HLA-*DRB1*) have been previously associated with protection and/or susceptibility to *Mycobacterium tuberculosis* in both European and East Asian populations (Tong et al. 2015; Oliveira-Cortez et al. 2016). Such results are partially validated for the West African Mali population, in which susceptibility to *Mycobacterium africanum* infections seems to be regulated by a specific polymorphism at HLA-*DRB1* (Kone et al. 2019). Moreover, recent evidence supports the association between the HLA-*DRB5* and HLA-*DPB1* genes with susceptibility to *M. tuberculosis* in diverse African human groups (Odera et al. 2022; Croock et al. 2025). Furthermore, a specific HLA-*DRB1* haplotype was proved to be significantly associated with protection from severe malarial anaemia in Gambian children (Hill et al. 1991), thus linking the regulation of immune system mediated by *HLA* genes also to malaria infections. Also, a previous study identified SNVs in the vicinity of multiple *HLA* loci (i.e. *DRB5*, *DRB9*, *DRA* and *DPA2*) showing significant values of the Fst statistic in West African populations, thus suggesting the occurrence of multiple independent selection events involving HLA genomic region in these human groups (Bhatia et al. 2011).

Finally, *IL15* encodes a cytokine that regulates T and natural killer cells activation and proliferation (Sayers et al. 2022). Remarkably, the combined IL15/IL15Rα (i.e. the IL15 receptor) treatment in mice was proved to modulate the dysregulated inflammation during Plasmodium infection by promoting T follicular helper cells differentiation and antibody generation, a process that correlates with improved survival from reinfection (Bravo et al. 2024).

Overall, given the biological functions modulated by these genes during malaria and tuberculosis infections, we propose that the putative adaptive haplotypes identified at *VTN*, *HLA* and *IL15* loci may play adaptive/protective roles in West African human populations by reducing the probability of contracting these infections and/or of developing severe forms of the related diseases. In accordance with this hypothesis, signatures of natural selection, although being not validated by *Trendsetter* in the Yoruba population, were found at several other immune-related loci (e.g. *IL19*, *IL22RA2* and *IFNA6*), as well as at *THBS2*, *THBS3*, *THBS4*, and *FN1* malaria-associated genes (Roberts et al. 1985; Eda and Sherman 2004) (Figure 4.9).

Endemic malaria constitutes a serious health burden for several non-African populations as well. A cardinal example is represented by groups inhabiting the Sardinia Island in the Western Mediterranean, which is notable for a long history of endemic malaria that plagued the population until the near eradication of mosquitoes shortly after the second World War (Kosoy et al. 2011). In the present study, when analysing Sardinian WGS data we identified both strong and weak selection signatures respectively at *CD36*/*81* and *CD38* genes, which are potentially related to malaria infection (Omi et al. 2003; Silvie et al. 2003; Silvie et al. 2006; Asito et al. 2008; Sinha et al. 2008; Schwenk et al. 2013; Cabrera et al. 2014) (Figures 4.15, 4.16 and Supplementary Table S4b). In fact, the expression and/or the genetic variability of such genes have been previously implicated in the development of malaria in diverse populations, but not in the Sardinian one.

In particular, the *CD36* receptor was proved to contribute to malaria pathology by directly interacting with *Plasmodium falciparum*-infected erythrocytes, and by carrying genetic variants associated with protection from cerebral malaria in Thai (Omi et al. 2003) and with severe malaria in both African and Indian populations (Omi et al. 2003; Sinha et al. 2008). The receptor protein encoded by *CD81* is instead required on both human and rodent hepatocytes to make possible *Plasmodium falciparum* and *Plasmodium yoelii* sporozoite infections (Silvie et al. 2003; Silvie et al. 2006). Finally, the proliferation of CD4+ T cells expressing the CD38 ectoenzyme was significant associated with a reduced blood-stage *P. falciparum* burden, providing results consistent with an important functional role played by these cells in protective immunity against malaria in humans (Burel et al. 2016). Given these lines of evidence, our results suggest that the selective events at *CD36*/*38*/*81* loci inferred for the Sardinian population may have been involved in the modulation of adaptive responses that determine blunted forms of malaria infection in such human group.

As regards instead the European populations more in general, we identified multiple gene networks significantly impacted by the action of natural selection, which belonged to immune-related functional pathways such as *Chemokine signalling*, in both CEPH and Sardinian groups, and *Pathogenic Escherichia coli infection* in the Finnish population (Supplementary Table S4b). This further supports the cardinal role of pathogen-induced selective pressures in having shaped the adaptive variability of such human groups.

In parallel, the Han Chinese population analysed was instead found to be characterized by selection signatures at loci previously related to the development of both hepatitis B (HBV) and C (HCV) infections and/or associated to related co-morbidities in such human groups (Figures 4.17a, 4.18 and 4.19, Supplementary Table S4c) (Peng et al. 2013; Wang et al. 2013; Xu et al. 2013; Lee et al. 2015; Hou et al. 2018). In detail, the most relevant putative adaptive genes were *IL23R*, *IL7* and *IFNL3*, which codify for specific cytokines and/or

cytokine receptor (Sayers et al. 2022). Interestingly, *IL23R* carries genetic polymorphisms that have been associated to the increased risk of developing HBV-related hepatocellular carcinoma (Peng et al. 2013; Xu et al. 2013), and its elevated expression was proved to be indispensable for the differentiation of CD4+ T cells during Hepatitis B infection (Wang et al. 2013). The serum levels of interleukin 7 were instead significantly reduced in chronic HCV infected patients, while *IL7* stimulation promotes viral clearance during chronic hepatitis C virus infection (Hou et al. 2018). Finally, *IFNL3* carries genetic variants significantly associated to the increased HCV spontaneous clearance, but also with the increased risk of developing hepatocellular carcinoma (Lee et al. 2015).

HBV and HCV infections have long been a major health problem in China (Liang et al. 2015; Jing et al. 2020), determining an elevated prevalence of hepatitis-related liver cancer and/or cirrhosis with respect to other countries. For instance, hepatocellular carcinoma (HCC) presents a particularly high prevalence in East and Southeast Asia, while Chinese people alone account for more than 50% of HCC cases worldwide. However, only a minority of lifelong chronic carriers of HBV eventually develops HCC, indicating the importance of host genetic risk/protective factors influencing the development of HBV-related comorbidities (Liang et al. 2009; Peng et al. 2013; Xu et al. 2013). This finding, along with the selection signatures identified at *IL23R*, *IL7* and *IFNL3* hepatitis-related genes, suggest that the variability of such loci may constitute part of the genetic factors conferring protection to Han Chinese population during HBV/HCV infections and/or that reduce its susceptibility to hepatitis-related comorbidities.

When considering populations of Native American ancestry, also the Mayas showed compelling evidence of adaptive evolution likely driven by endemic pathogens, such as the *Trypanosoma cruzi* parasite, which causes Chagas disease (Medina-Torres et al. 2010; Carabarin-Lima et al. 2013). Particularly, such evidence concerns the *PPP3CA* and *THBS1* genes, which are loci known to be linked to *T. cruzi* infection (Figure 4.29c-d, Supplementary Figure S30 and Supplementary Table S4d) (Simmons et al. 2006; Nde et al. 2012; Couto-Silva et al. 2023). Interestingly, *PPP3CA* has been proposed to carry signatures attributable to positive selection in other indigenous human populations, particularly those from Brazil (Couto-Silva et al. 2023) and southern Mexico (Ojeda-Granados et al. 2022). Remarkably, these selective events have been functionally linked to a lower percentage of cardiomyocytes infected by *T. cruzi*, potentially due to the diminished *FLJ20021* expression, which has been associated to a milder form of the Chagas disease (Nunes et al. 2018; Couto-Silva et al. 2023). The expression of *THBS1* was instead shown to be upregulated during *T. cruzi* early infection in human coronary artery smooth muscle cells, and accordingly, its knockdown dramatically reduces the parasite activity in the host (Simmons et al. 2006; Nde et al. 2012). We therefore argue that the adaptive haplotypes identified at both *PPP3CA* and *THBS1* loci may protectively modulate immune responses during *T. cruzi* infection in the Mayas, thus resulting in the blunted forms of Chagas disease observed in such population.

Furthermore, we identified broader pathogen-driven selection signatures in indigenous Mexican populations, particularly involving genes (e.g. *LYN*, *ELMO1*, *RHOA* and *HCK*) that play key roles in the regulation and activation of the immune system as a defence mechanism against diverse types of infections (Figure 4.28b, Supplementary Table S4d) (Smolinska et al. 2011; Li et al. 2014; Sarkar et al. 2017; Bros et al. 2019). Taken

as a whole, results obtained for indigenous Mexican populations proved once again the crucial role played by pathogen-related selective pressures in shaping the adaptive variability of modern human populations.
Finally, after the application of the *Trendsetter* method to several populations of different ancestry, we detected some positive selection signatures involving sepsis-related genes (Supplementary Table S7), although such results appeared to be more congruent with the occurrence of hard/soft selective sweeps rather than reflecting the evolution of polygenic adaptations. Particularly, selection signatures at the *CTCF* and *LRBA* genes were found to be shared among European, East Asian and Native American populations (Supplementary Table S7). In detail, the enduring increase in *CTCF* binding within the major histocompatibility complex (MHC)-class II has been associated with postoperative sepsis (Siegler et al. 2021), while a rare homozygous mutation at the *LRBA* gene was proved to cause primary immunodeficiency syndrome, a condition that often evolves in sepsis (Burns et al. 2012). Moreover, *RAG1*/*2* and *MAD1L1* genes, which showed adaptive evolution in both CEPH and Han Chinese populations, were respectively proved to carry mutations causative of Severe Combined Immunodeficiency (SCID), which are rare genetic disorders often leading to neonatal sepsis (De Villartay et al. 2005; Cossu 2010) and have been thus proposed as sepsis-diagnostic biomarkers (Chen et al. 2024). Furthermore, mutations at the *CD79A* and *DCLRE1C* genes, which showed selection signatures in both Yoruba and Maya populations, have been associated to SCID immunodeficiencies as well (Ghadimi et al. 2023; Yu et al. 2024). Finally, adaptive evolution at *CASP12* (Xue et al. 2006), a gene that has been previously proposed to confer resistance to peritonitis and septic shock (Saleh et al. 2006), were further validated by our analyses in the CEPH group (Supplementary Table S7). Overall, literature evidence, combined with our findings, suggest that these loci may play adaptive roles in modern human populations by providing protection against uncontrolled immune system activation/reaction and the subsequent development of sepsis, thus further supporting how pathogen-driven selective pressures represented also an indirect driving force of adaptive evolution in humans.

5.2. Polygenic adaptations and dis-adaptations triggered by climate-induced selective pressures

Climatic factors, such as temperature and humidity, play an important role in determining species distributions and have likely influenced also the phenotypic/genomic variation of modern human populations (Hancock et al. 2011). Part of the signatures of positive selection emerged in the present study may indeed involve the genetic bases of complex biological adaptations evolved by the ancestors of present-day Eurasian human groups (e.g. CEPH, Finnish, and Russians) to cope with cold continental climates and/or with sub-arctic climate (Yakuts). Specifically, these results concern genes such as *LIPC*, *LIPG*, *PLPP1*, *MOGAT2*, *PNLIP*, *INSR* and *NFKB1* (Figures 4.12a, 4.13, 4.14a-b, 4.24, 4.25b and Supplementary Table S4b-c), which have been proved to contribute to the regulation of lipid digestion, uptake, and new synthesis (Sayers et al. 2022), as well as in the modulation of insulin responses (Coto et al. 2018).
In detail, *LIPG*/*C* and *PNLIP* encode members of the lipase family that are implicated in the hydrolysis of triglycerides and are essential for the efficient digestion of dietary fats (Sayers et al. 2022). Mutations and/or

alterations of the expression of these genes have been associated to weight loss in obese individuals, congenital pancreatic lipase deficiency (e.g. those at *PNLIP*), as well as to familial combined hyperlipidaemia and type 2 diabetes (e.g. those at *LIPC*) (Sayers et al. 2022). Furthermore, *PLPP1* and *MOGAT2* genes codify for enzymes that catalyse diacylglycerol production, which in tourn i) can be used to produce triacylglycerol and glycolipids or ii) is hydrolysed to produce free fatty acids (Kanehisa and Goto 2000; Sayers et al. 2022).

The *INSR* gene encodes instead a receptor that, after binding insulin, activates downstream signalling, thus regulating glucose uptake, release, as well as the synthesis and storage of carbohydrates, lipids and proteins (Sayers et al. 2022). Notably, mutations at both *INSR* and *NFKB1* genes have been associated to increased risk of developing type 2 diabetes (Coto et al. 2018) and to inherited severe insulin resistance syndromes, such as type A insulin resistance, Donohue and Rabson-Mendenhall syndromes (Sayers et al. 2022). Furthermore, the *ADCY3* gene that showed selection signatures in diverse Eurasian human groups (Figure 4.14c and Supplementary Table S4b) also presented haplotype patterns conformed with a scenario of adaptive introgression events from Denisovans presumably occurred in the ancestors of both Russian and Yakut populations (Figure 4.26). Again, variation at this locus has been previously demonstrated to be associated with type 2 diabetes (Grarup et al. 2018), adiposity and insulin resistance (Keele et al. 2018; Tian et al. 2018). Remarkably, the remodelling of triacylglycerol and glycerophospholipids in Brown Adipose Tissue (BAT), (Marcher et al. 2015) as well as increased lipase activities with consequent enhanced free fatty acids (FFA) release (Calderon-Dominguez et al. 2016; Mallov), have been observed as a result of cold exposure in the mouse model. Interestingly, enhanced insulin-mediated glucose uptake in both white and brown adipose tissues was also detected in rats during cold exposure (Vallerand et al. 1987), a process that has been proposed to contribute to improve BAT metabolism and/or to sustain tissue energy demand during cold exposure (Hankir and Klingenspor 2018). Furthermore, a potential crosstalk between BAT and glucose regulatory pathways has been proposed also in humans (Lee et al. 2016) because of the correlation between a greater BAT activation with reduced glycemia and/or protection against diabetes (Schrauwen et al. 2015). That being so, considering the biological functions regulated by these genes, our results suggest that the selection signatures detected in continental Europeans, as well as in high-latitude Eurasian populations, may represent those metabolic adjustments that allowed these human groups to cope with cold continental/sub-arctic climates.

Nevertheless, partially overlapping results concerning genes belonging to the *Glycerolipid metabolism* pathway were detected also for populations currently living in Mediterranean climates (e.g. Iberian and Sardinian populations), even though these signatures are associated with non-significant or marginally significant p-values and/or involved loci not directly related to the aforementioned functions (Supplementary Table S4b). These findings may reflect the genetic bases of ancient adaptations that have been likely evolved in the early ancestors of contemporary Eurasian populations prior to their genetic divergence, and which remain barely detectable, although with reduced statistical robustness, even in populations no longer exposed to cold-induced selective pressures.

Strictly related to these topics, the Yakut population from Northern Siberia presented selection signatures at genes participating to the *Thyroid hormone signalling* pathway (Figure 4.24, Supplementary Table S4c) and/or

involved in the modulation of cold-activated biological functions. Such results were further validated by eQTL (at *THRB, SRC, RCAN2, INSR* and *HTR1F* genes; Figure 4.27 and Supplementary Figure S29) and *Trendsetter* (at *HTR1F*, *SRC* and *RCAN2* genes; Figure 4.1a, Supplementary Figure S24a-b, Supplementary Table S4c) analyses.

In detail, *THRB* encodes for a nuclear receptor for the triiodothyronine (T3) hormone (Sayers et al. 2022), which has been shown to directly regulate adaptive thermogenesis in BAT, as well as to induce browning process of white adipocytes by regulating the expression of the UCP1 uncoupling mitochondrial protein (Figure 5.2) (Yau and Yen 2020; Zekri et al. 2021; Machado et al. 2022).

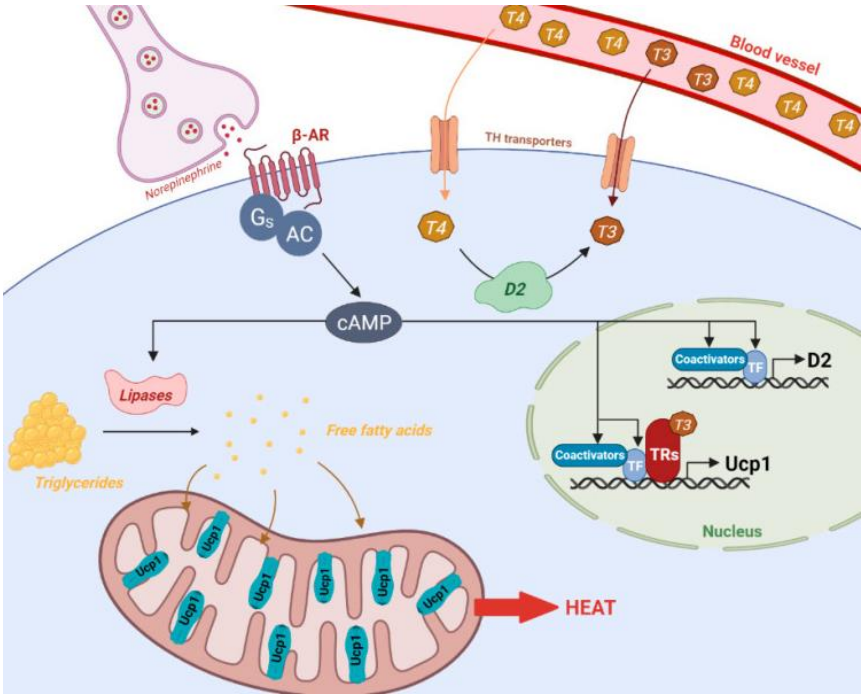


Figure 5.2. Regulation of adaptive thermogenesis in brown adipocyte by norepinephrine and thyroid hormone. Sympathetic neurons release norepinephrine that binds to β-adrenergic receptors (β-AR), which are coupled to stimulatory guanine nucleotide binding protein (Gs) that in tourn activates adenylate cyclase (AC) to produce cAMP. Such adrenergic signalling activates transcription factors (TF) and coactivators involved in the regulation of type 2 deiodinase (D2). Both adrenergic signalling and thyroid hormone receptors (TRs) regulate UCP1 expression. Triglycerides are hydrolysed into free fatty acids by lipases and transported to mitochondria to sustain β-oxidation and activate UCP1, which uncouples ATP production from respiration, comporting an increased mitochondrial activity and heat production. The figure has been adapted from Zekri et al. (2021).

Furthermore, *RCAN2* was identified as a T3-responsive gene (Miyazaki et al. 1996) and its knockout in mice was found to be related to a reduced body weight and white adipose mass compared to control subjects, a condition that is significantly enhanced in a regime of high-fat diet (Sun et al. 2011). The *SRC* proto-oncogene instead encodes for a tyrosine-protein kinase that i) enhances the expression of the cold-induced neuroprotective gene *RBM3* thus mediating neuroprotective effects of mild hypothermia (Yuan et al. 2021), ii)

regulates fatty acid synthesis (Zhao et al. 2024), iii) modulates cold-induced activation of *TRPM8*, a receptor acting upstream of cold-induced thermogenesis, response to cold and thermoception processes (Manolache et al. 2020; Sayers et al. 2022), and iv) causes impairment of glucose-induced insulin secretion in pancreatic β-cells when its expression and/or activity is dysregulated (Sato et al. 2016). Finally, the *HTR1F* receptor was shown to regulate the secretion of glucagon in pancreatic alpha cells thought the binding with serotonin (Almaça et al. 2016). In fact, serotonin is known to regulate energy metabolism in several others peripheral tissues, such as BAT and white adipose tissue, skeletal muscle and liver (Choi et al. 2020).

Interestingly, short-time sojourners in Antarctica have been reported to present increased production and uptake of thyroid hormones, a condition that results in reduced free T3/T4 circulating levels (Levy et al. 2013). This physiological adjustment, known as "polar T3 syndrome", has been observed also in the Yakut population when comparing their levels of circulating T3/T4 and thyroid-stimulating hormone (THS) between summer and winter seasons, suggesting enhanced capacity to increase metabolic heat production during seasonal severe cold (Levy et al. 2013). Moreover, Yakut young adults displayed improved BAT thermogenesis in response to mild cooling when compared with control groups (Levy et al. 2022).

Taken together, adaptive evolution at *Thyroid hormone signalling* genes and at loci contributing to lipid/glucose metabolisms may represent the determinants of Yakut enhanced thyroid hormone uptake (and the consequent augmented BAT activity), as well as of their fine-tuning of lipid/glucose metabolisms and insulin signalling in different districts of the body, which conjunctly mediate complex cold adaptations in such population.

Intriguingly, we identified selection signatures potentially related to insulin signalling also in the Maya population. Particularly, such results concern multiple genes involved in the *Insulin Resistance* pathway (e.g. *IRS1*, *PTPRF* and *PRKCZ*, Figure 4.28a, Figure 4.29a-b and Supplementary Table S4d), as well as the *FLT3* gene, which have been supported by the *Trendsetter* approach in the Mayas, but also in the CEPH and Yakut populations (Supplementary Table S4d).

Concerning the functional roles played by these genes, *IRS1* encodes the *insulin receptor substrate 1*, which is a protein phosphorylated by the insulin receptor tyrosine kinase and whose mutations are associated with type II diabetes and susceptibility to insulin resistance (Sayers et al. 2022). Moreover, *IRS1* together with *PRKCZ, PRKCD* and *PRKCQ* genes participates at the enriched Gene Ontology term of *Negative regulation of insulin receptor signalling pathway* (Figure 4.28a), suggesting a functional correlation of such genes that conjunctly can determine increased glycemia and tissue insulin resistance. In detail, the expression of *PRKCZ* was found to be significantly reduced in type 2 diabetes patients with respect to control groups as a result of the CpG sites hypermethylation at *PRKCZ* promoter, suggesting the involvement of such gene in diabetes mellitus pathogenicity (Zou et al. 2013). The overexpression of *PTPRF* tyrosine phosphatase has been instead detected in obese insulin-resistance tissues and has been shown to induce insulin resistance in transgenic mice by inhibiting the insulin receptor tyrosine kinase and phosphorylation activity (De Lorenzo et al. 2013; Sayers et al. 2022). Finally, treatment with Flt3-ligand significantly decreased the risk of progression to diabetes in

non-obese diabetic (NOD) mice (Chilton et al. 1995) and has also been shown to increase in the innate resistance against different pathogens (Vollstedt et al. 2003).

Overall, these results provide evidence that may support the evolution of dis-adaptations (i.e. evolutionary mismatch scenarios) involving insulin-related loci in the Maya population. Briefly, the evolutionary mismatch hypothesis stands that genetic variation that in ancestral environments might have been involved in the modulation of biological adaptations, in the present context, which is remarkably different as a result of recent modifications in the environmental/ecological conditions and/or population habits, turned out to increase the susceptibility of developing complex diseases (Li et al. 2017).

Notably, diverse indigenous populations from North, Central, and South America exhibit a significantly higher prevalence of type 2 diabetes compared to their non-indigenous (i.e. admixed) counterparts (Benyshek et al. 2001; Leung 2016). Particularly, a higher prevalence of diabetes has been reported in the southeast area of Mexico, a region occupied mostly by Maya indigenous groups (Lara-Riegos et al. 2015). Moreover, it has been suggested that genetic susceptibility of Mexican mestizos to diabetes, obesity and dyslipidaemia is related to their fraction of genetic background ascribable to the Native American ancestry (Lorenzo et al. 2001; Cossrow and Falkner 2004; Aguilar-Salinas et al. 2009).

The increased predisposition of Native American populations to type 2 diabetes has often been explained with the thrifty genotype hypothesis (Benyshek et al. 2001), a theory that has, however, faced substantial criticisms in more recent years. Such hypothesis proposes the adaptive evolution of a “thrifty” genotype/phenotype in response to periods of famine and food scarcity, which, however, in modern contexts leads to fat accumulation, especially under a diet rich in fats and processed food (Speakman 2008). Therefore, given the uncertainty of this previous evidence, and considering both the rapid spread of human populations in the American continent (Oppenheimer 2012; Willerslev and Meltzer 2021) and a quite long isolation of their ancestors in Beringia, lasted approximately 15,000 years (Tamm et al. 2007), here we propose that selection signatures related to insulin signalling identified for the Mayas may represent a reminiscence of cold-induced adaptations originally evolved by ancestral human populations that from Northern Siberia moved to Beringia. More specifically, we argue that indigenous populations from south Mexico may be predisposed to greater fat accumulation/type 2 diabetes due to the relatively rapid transition from severe cold to wet tropical climates, rather than to a higher frequency of thrifty genotypes in these human groups respect to others. Our hypothesis is based on the evidence that reduced exposure to seasonal cold may have a dual effect on energy expenditure, both minimizing the need for physiological thermogenesis and reducing thermogenic capacity (Fontana et al. 2024). These processes in hot/warm climates can lead to fat accumulation in the adipose tissue and thus to obesity (Koch et al. 2021; Huang and Hong 2024). Obesity, in turn, causes an increased circulation of free fatty acids, which leads to greater use of these compounds to fuel cellular processes instead of glucose, comporting increased glycaemia (Felber and Golay 2002). As a compensatory response, insulin secretion is increased. However, over time, this metabolic imbalance can lead to insulin resistance and, ultimately, to the development of type 2 diabetes (Felber and Golay 2002).

Adaptive evolution at the *FLT3* insulin-related gene in Maya, CEPH and Yakut populations seem to corroborate such hypothesis (Supplementary Figure S13 and Supplementary Table S4b-c-d). Furthermore, the less distinct *Trendsetter* probability trends observed for the insulin-related *PRKCZ* and *PTPRF* genes (Figure 4.29 a-b), when compared to the clearer patterns observed in immune genes such as *PPP3CA* and *THBS1* (Figure 4.29 c-d), may further support the occurrence of an evolutionary mismatch. Nevertheless, all putative adaptive genomic windows detected for the *FLT3* gene were classified as having experienced hard selective sweeps in the Maya ancestors (Supplementary Figure S13), differently from what observed for the CEPHs and Yakuts, for which the selection signatures at the same locus were indicative of soft selective sweeps. Such a pattern could be explained either by an increased effect of the genetic drift having affected the adaptive variability of the ancestral Beringian populations during and after their migration into the American continent, and/or by local adaptation evolved by the ancestors of Native American populations, possibly conferring increased immunity to viral infections (Vollstedt et al. 2003), which however may have also increased the risk of developing type 2 diabetes.

5.3. Polygenic adaptations evolved by high-altitude populations

Hypobaric hypoxia represents an intense selective pressure currently acting on diverse modern human populations living at high altitudes and was proved to have influenced both epigenetic (Childebayeva et al. 2021; Lin et al. 2023) and genomic variability of highlander ethnic groups (Beall et al. 2010; Yi et al. 2010; Horscroft et al. 2017; Gnecchi-Ruscone et al. 2018; Julian and Moore 2019; Zhang et al. 2021; Borda et al.). To further comprehend the complex architecture of biological adaptations evolved by human populations to cope with hypobaric hypoxia, we investigated i) the genetic bases of polygenic adaptations evolved by high-altitude groups of Tibetan ancestry (Figures 4.21, 4.22, Supplementary Figures S21 and S22 and Supplementary Table S4c), ii) the role played by Denisovan adaptive introgression in shaping the variability of such complex adaptive traits (Figure 4.23 and Supplementary Figure S23) (Ferraretti et al. 2024), and iii) the possible convergent polygenic adaptations shared by Andean and Himalayan high-altitude populations (Figure 4.30, Supplementary Figure S32 and Supplementary Table S4d) (Ferraretti et al. 2025). Overall, the analyses performed pointed out pervasive signatures ascribable to the action of natural selection at genes/pathways involved in the promotion of angiogenesis (e.g. *LAMA1*/*2/3*, *LAMC2*, *DAG1* and *COL4A1*) (Figure 4.30 and Supplementary Figure S23), as well as in the regulation of some cardiovascular traits (e.g. *IGF1R*, *PDGFD*, *PIKFYVE* and *PPP2R5C*) (Figure 4.20). Part of these signatures were found to be shared at the single gene-level by both Andean Aymara and Tibetan high-altitude populations (Ferraretti et al. 2025), also showing patterns of variation suggestive of Denisovan adaptive introgression events occurred specifically in the ancestors of Tibetans (Ferraretti et al. 2024).

More in detail, Tibetan highlanders were found to be characterized by adaptive evolution at loci involved in the *Adrenergic signalling in cardiomyocytes*, *Phosphatidylinositol signalling system* and *ECM-receptor interaction* pathways (Figure 4.21 and Supplementary Table S4c), with results concerning specifically the

PPP2R5C, *PIKFYVE*, *DAG1* and *PIP4K2* genes that have been corroborated also by the *Trendsetter* approach (Figure 4.22 and Supplementary Figure S21).

In particular, *PPP2R5C* codifies for the B56γ protein that is part of the protein phosphatase 2A (PP2A) regulatory subunit B family (Sayers et al. 2022). Remarkably, inactivation of both B56γ and B56δ proteins were proved to cause the arrest of fetal development and severe cardiac defects in the mouse model (Dyson et al. 2022). Moreover, *PPP2R5C* knockout in mice determines increased PP2A activity, leading to a decreased heart rate and to heart conduction defects (Little et al. 2015; Jiao et al. 2025). The inactivation of *PIKFYVE* was instead shown to suppress the excessive production of mitochondrial reactive oxygen species (ROS) and apoptosis in cardio-myoblasts, to reverse cardiac mitochondrial damage, to improve cardiometabolic profile, left ventricular function and to attenuate cardiac hypertrophy in obese mice (Tronchere et al. 2017). Furthermore, the interaction of *PIP4K2A* mRNA transcript with the Human antigen R (HuR) RNA-binding protein was proved to positively modulate *PIP4K2A* expression and consequently regulate autophagy and calcific aortic valve disease (CAVD) progression (Fang et al. 2023). Notably, CAVD is defined as a slowly progressive condition that ranges from mild valve aortic sclerosis to severe calcifying aortic valve stenosis, a pathological process that was proposed to be accelerated by impairments in the mitochondria functioning, oxidative stress and inflammation processes, as well as by the activation of the HIF1 hypoxia signalling pathway (Bouhamida et al. 2023). Finally, the dystroglycan 1 protein codified by the *DAG1* gene was proved to directly bind the Hippo pathway effector Yap to inhibit cardiomyocyte proliferation in mice (Morikawa et al. 2017). Nevertheless, dystroglycan may also modulate angiogenesis via interaction with the alpha chain of laminin-1 (Javerzat et al. 2009).

Specifically, cardiac hypertrophy is a physiological process that comports the increase in size of cardiomyocytes, and which initially develops as an adaptive response to multiple physiological and pathological stimuli (Nakamura and Sadoshima 2018). However, it can progress in pathological cardiac hypertrophy, which generally leads to heart failure (Nakamura and Sadoshima 2018). Notably, cardiac hypertrophy can be induced by prolonged hypoxia (Chu et al. 2012; Nakamura and Sadoshima 2018; Johnson et al. 2023) potentially comporting either detrimental or adaptive physiological effects (Cazorla et al. 2006; Nakamura and Sadoshima 2018; Chen et al. 2025). In fact, a recent study conducted on Han Chinese patients affected by hypertrophic cardiomyopathy (HCM) and living in low- and high-altitude regions, showed that, while high-altitude may promote myocardial fibrosis, it may also be linked to some adaptive benefits in left ventricular function in HCM patients (Chen et al. 2025). Furthermore, it has been shown that high-altitude Tibetan people exhibit reduced incidence of major adverse cardiovascular events with respect to low-altitude control groups, possibly indicating the involvement of protective cardiac mechanisms in the modulation of high-altitude adaptations, as previously proposed (Kolář and Ošťádal 2004; Mallet et al. 2018; Ferraretti et al. 2024; Lei et al. 2024).

Mitochondrial dysfunctions and increased production of ROS are physiological processes induced by hypoxia as well (Guzy et al. 2005; Huan et al. 2023). Remarkably, high-altitude native individuals exhibit lower mitochondrial density compared to lowlanders (Li et al. 2016; Murray and Horscroft 2016) a trait that has been

suggested to represent an adaptive mechanism to protect tissues/cells from ROS damage (Hoppeler et al. 2003; Heather et al. 2012).

Finally, enhanced angiogenesis is a complex process proposed to be triggered by hypoxia (Krock et al. 2011; Tsipis et al. 2014) and to play a key role in high-altitude adaptation by ensuring adequate tissue perfusion and oxygenation under hypoxic conditions (Beall 2007; Gnecchi-Ruscone et al. 2018).

Therefore, based on the evidence supporting the functional connection of the *PIKFYVE, PIP4K2A*, *PPP2R5C* and *DAG1* genes with the aforementioned physiological processes, we propose that the selection signatures detected at these loci may represent the genetic bases of polygenic adaptations that contribute to enhance angiogenesis, reduce susceptibility to hypoxia-induced fatal cardiovascular events, including those associated with pathological cardiac hypertrophy, and mitigate cellular oxidative stress in high-altitude populations of Tibetan ancestry.

In line with these findings, in the study by Ferraretti et al. (2024), we propose compelling evidence supporting adaptive introgression events having shaped variation patterns at genes related to cardiovascular functions/diseases (e.g. *PRKAG2*), promotion of angiogenesis (e.g. *KRAS* and *RASGRF2*) and regulation of responses to the hypoxic stress (e.g. *TBC1D1* and *PRKAG2*) (Figure 4.23, Figure 2-figure supplement 3 and 4 in Ferraretti et al. 2024). Overall, the observed haplotype structure at these loci may be attributable to Denisovan archaic introgression (Supplementary Figure S23, Figure 4 in Ferraretti et al. 2024), suggesting that interbreeding with archaic hominins has had a greater impact than previously thought in shaping the adaptive genomic variation of present-day Tibetans (Ferraretti et al. 2024).

As regards the functional roles exploited by these genes, *TBC1D1* encodes for a protein targeted by AMP-activated protein kinases (AMPK) in response to the increased cellular AMP/ATP ratio caused by stresses that induce a lower ATP production (e.g. deprivation of oxygen and/or glucose) or that boost ATP consumption (e.g. intense muscle contraction) (Kanehisa and Goto 2000; Vichaiwong et al. 2010). Mutations at *PRKAG2*, which is another gene contributing to the *AMPK signalling* pathway, were shown to cause PRKAG2 cardiac syndrome, an inherited disease characterized by ventricular pre-excitation, supraventricular arrhythmias, and cardiac hypertrophy (Zhang et al. 2013; Porto et al. 2016). Specifically, dysregulation of AMPK activity mediated by reduced expression of *PRKAG2* has been proposed as a possible cause for the development of this disease (Zhang et al. 2013). Conversely, enhanced activation of the *AMPK signalling* pathway during pregnancy, along with overexpression of *PRKAG2*, has been observed in the placentas of women living at high altitudes when compared to those of low-landers (Lorca et al. 2021) suggesting a role of this gene also in the establishment of a successful pregnancy in hypobaric hypoxia condition.

The proteins encoded by *RASGRF2* and *KRAS* genes are instead strictly linked from a functional perspective and exert their functions through the *Ras signalling* pathway (Kanehisa and Goto 2000). In detail, *RASGRF2* codifies for a calcium-regulated nucleotide exchange factor that activates the RAS protein codified by the proto-oncogene *KRAS* (Kanehisa and Goto 2000; Sayers et al. 2022). Notably, the *Ras* and *MAPK/ERK signalling* pathways have been proposed to improve angiogenesis by promoting *VEGF* expression (Supplementary Figure S23) (Kanehisa and Goto 2000; Kranenburg et al. 2004).

Therefore, in line with previous observations, we proposed that adaptive evolution at *TBC1D1* and *PRKAG2* genomic regions in high-altitude Tibetan groups might have contributed to the development of protective mechanisms that reduce cardiovascular risk associated to the hypoxic stress, while those observed at *KRAS* and *RASGRF2* genes seem to be implicated in the modulation of improved angiogenetic responses to hypoxia (Ferraretti et al. 2024).

In accordance with this view, results described in the study by Ferraretti et al. (2025) further support the adaptive role played by pro-angiogenetic pathways and specific cardiovascular traits in the modulation of polygenic adaptations evolved by different high-altitude populations. More specifically, we detected selection signatures shared by both Andean and Himalayan populations (Hu et al. 2017; Arciero et al. 2018; Gnecchi-Ruscone et al. 2018; Deng et al. 2019) at genes (i.e. *LAMC2*, *COL4A4*, *PLCE1*, *PIK3CB* and *PRKCE*) and/or pathways (i.e. *Focal adhesion* signalling cascades) known to promote angiogenesis and regulate cardiovascular functions (Figure 4.30, Supplementary Figure S32 and Supplementary Table S4d) (Kanehisa and Goto 2000; Ferraretti et al. 2025).

Particularly, silencing *LAMC2* expression was linked to cell cycle arrest and to the significant suppression of tumour cells migration and invasion, as well as to *in vivo* angiogenesis of the malignant tissue (Garg et al. 2014; Pei et al. 2019). Moreover, *LAMC2* expression was found to be substantially upregulated in Placental Accreta Spectrum tissues, stimulating trophoblast over-invasion (Wang et al. 2023). In particular, the placental accreta spectrum is a serious complication of pregnancy caused by the pathologic adherence of the placenta due to abnormal trophoblast neovascularization and invasion into the uterine wall (Bartels et al. 2018; Cahill et al. 2018). This invasion has been related to improved stimulation of pro-angiogenetic factors in both the trophoblast embryotic tissue and the maternal placental basal plate (Goh et al. 2013; Bartels et al. 2018).

Moreover, *COL4A4* and *PIK3CB* genes seem to be involved in the modulation of physiological processes established in preparation to pregnancy (Oefner et al. 2015; Tong et al. 2020). In particular, *COL4A4* expression increases during decidualization of the human endometrium (Oefner et al. 2015) potentially contributing to the regulation of the morphological and functional changes (e.g. increased vascular permeability and vascular remodelling and angiogenesis) required for the establishment of a successful pregnancy (Plaisier 2011; Okada et al. 2018). In accordance, *PIK3CB* turned out to be up-regulated in the decidua of woman developing early-onset pre-eclampsia, thereby plausibly influencing abnormal placental development (Tong et al. 2020). As concerns the remaining set of genes, *PLCE1* was associated to the inhibition of apoptosis and angiogenesis in tumours (Yunzhao Chen et al. 2019), while *PRKCE* has been demonstrated to exert a cardio-protective role against ischaemic injury (Scruggs et al. 2016).

Furthermore, the genes pinpointed by the LASSI-*signet* approach as putative adaptive loci specifically in Aymara high-altitude population from the Andes resulted to be linked to similar biological functions (Kanehisa and Goto 2000; Ferraretti et al. 2025) (Supplementary Table S4d). In detail, *COL4A1* codifies for a type IV collagen alpha protein that represents a main constituent of the vascular basement membrane (Kalluri 2003; Sayers et al. 2022). Both the degradation and remodelling of such membrane are the very first steps of angiogenesis induced by multiple growth factors (Kalluri 2003). In addition, mutations at *COL4A* genes have

been associated with cerebrovascular disease (Sayers et al. 2022) and intracerebral haemorrhages due to systemic small-vessel disease (Kuo et al. 2012), thus suggesting their crucial role in modulating the formation of new vascular structures. Accordingly, the *PDGFD* gene codifies for an angiogenetic factor whose activity promotes tumour angiogenesis in mice due to enhanced VEGF expression (Li et al. 2003). *PDGFD* down-regulation was also observed in the placenta of women with preeclampsia, a condition characterised by placental hypoperfusion and elevated expression of antiangiogenetic factors (Sitras et al. 2009). Finally, the *IGF1R* receptor is known to bind the IGF1 angiogenetic factor and to be overexpressed in different tumours where it acts as an anti-apoptotic agent by enhancing cell survival (Sayers et al. 2022; Xu et al. 2022). Moreover, a homozygous chromosomal mutation at *IGF1* causes intrauterine growth retardation, which is accompanied by a low average placental weight (Woods et al. 2025).

Overall, considering the consistent amount of evidence provided by previous studies, we hypothesised that adaptive evolution at these loci might result in the promotion of angiogenesis in diverse district of the body, including placenta, and/or in the regulation of protective cardiovascular functions, thus contributing to modulate complex adaptive traits that are observable in both Andean and Himalayan high-altitude populations (Ferraretti et al. 2025).

6. CONCLUSION

Overall, the present dissertation proposes an innovative combination of multiple statistical approaches (Gouy et al. 2017; Mughal and DeGiorgio 2019; Harris and DeGiorgio 2020) that in an integrated manner can detect a range of selective events that have plausibly contributed to the evolution of complex (i.e. polygenic) biological adaptations by modern human populations. Remarkably, such a pipeline of analyses offers an alternative conceptual and methodological framework for investigating the genetic bases of biological polygenic adaptations, which is profoundly distinct from the approaches conventionally adopted by both GWAS and studies focused on the computation of polygenic risk scores. Its flexibility makes it particularly well-suited for application in targeted case-studies and to population cohorts not necessarily characterized by large sample sizes. Particularly, by applying such a pipeline of analyses on WGS data, we were able to provide new insights concerning the genetic bases underpinning complex biological adaptations evolved by a panel of populations that are well representative of a remarkable fraction of the human biodiversity (Figure 5.3).

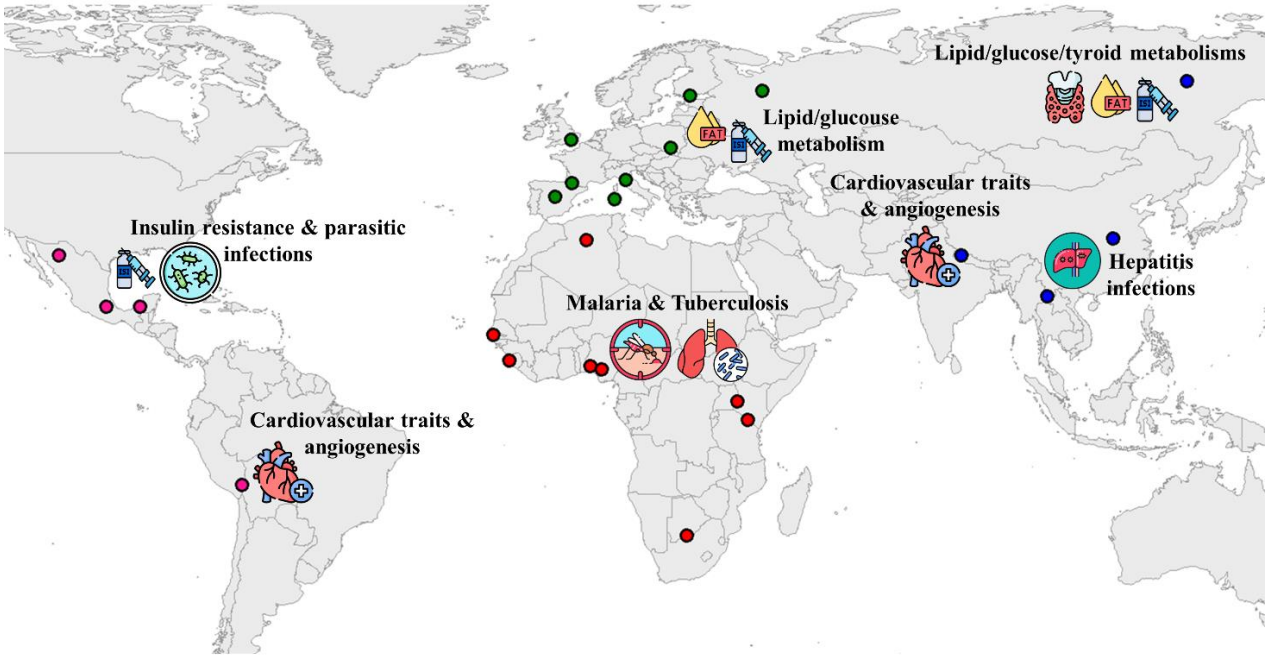


Figure 5.3. Key selection signatures ascribable to the evolution of polygenic adaptations. The reported map mirrors the geographic location of the human populations reported in Figure 4.4 and displays the main findings described in the present thesis as concerns polygenic local adaptations inferred for human groups from each of the considered geographical macro-areas.

Taken together, the obtained results encompassed soft selective sweeps-like patterns, as well as more subtle genomic signatures ascribable to the action of natural selection, both of which seem to have contributed to the evolution of complex polygenic adaptive traits in the considered populations. Remarkably, these findings are

consistent with the recently introduced concept of *loci heterogeneity*, which is supposed to underly the genetic architecture of polygenic adaptations (Barghi et al. 2020).

Notably, among the most pervasive adaptive signatures identified, we found those related to pathogen-driven selection, which have been likely mediated either by endemic infectious diseases (such as those described for the Yoruba, Han Chinese and Maya populations) and/or are imputable more generally to different types of pathogen insults. For instance, genes related to immune system over-activation and sepsis showed patterns conformed with adaptive evolution across several human populations from different geographical macro-areas. Moreover, complex metabolic adaptations concerning lipid digestion and glucose uptake probably evolved in response to cold-induced selective pressures have significantly shaped the genomic variability of different populations of Central European ancestry, as well as of some high-latitude Eurasian human groups. Interestingly, we showed that part of such adaptive traits may have evolved as results of adaptive introgression processes, which involved archaic hominin species and the ancestors of both Yakut and Russian populations. In addition, we identified Yakut-specific selection signatures at thyroid hormone-related genes and we propose them as key regulators of complex adaptations evolved by such population to cope with a sub-arctic climate.

Unprecedented results supporting dis-adaptations of the Maya population from Southern Mexico and concerning specifically loci involved in the development of insulin resistance and/or type 2 diabetes were also obtained. Particularly, we hypothesize that they might be informative of an evolutionary mismatch attributable to complex adaptations evolved by the ancestors of Native American populations during their migratory route across Beringia that turned out to be detrimental after the rapid occupation of the southern regions of the American continent, where completely different climatic conditions with respect to those of northern Siberia are experienced.

Finally, the present project expanded our knowledge also regarding the genomic architecture of high-altitude polygenic adaptations. In fact, in the studies by Ferraretti et al. (2024) and Ferraretti et al. (2025) we described, respectively, new insights concerning the archaic adaptive introgression events experienced by the ancestors of high-altitude Tibetan populations and the convergent adaptations evolved by Andean and Himalayan high-altitude human groups. Overall, according to our analyses, high-altitude populations showed extensive selection signatures at genes/pathways promoting angiogenesis and regulating cardio-vascular functions, thus potentially constituting the genetic bases of complex biological traits that exert an adaptive role at high-altitudes.

Strikingly, selection signatures associated with the evolution of complex cardiovascular and angiogenesis-related traits, as well as adaptations mediated by adjustments at glycerolipid and insulin signalling pathways, which have been identified across populations whose ancestors have long experienced similar selective pressures, but that involved different genes in different human groups, illustrate paradigmatic examples of *non-parallelism*, which is proposed to be another hallmark of polygenic adaptations (Barghi et al. 2020).

In conclusion, our analytical pipeline revealed a mosaic of selective signals that is consistent with a scenario of polygenic adaptive evolution, thereby shedding light on several complex adaptations evolved by our species to cope with a diversified range of selective pressures.

7. REFERENCES

A WK, Cecilia C-H, O SM, L CAJ. 2025. Intrauterine Growth Retardation and Postnatal Growth Failure Associated with Deletion of the Insulin-Like Growth Factor I Gene. New England Journal of Medicine 335:1363–1367. Available from: https://doi.org/10.1056/NEJM199610313351805.

Adhikari K, Mendoza-Revilla J, Sohail A, Fuentes-Guajardo M, Lampert J, Chacón-Duque JC, Hurtado M, Villegas V, Granja V, Acuña-Alonzo V, et al. 2019. A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. Nat Commun 10.

Aguilar-Salinas CA, Canizales-Quinteros S, Rojas-Martfnez R, Mehta R, Ma TVM, Arellano-Campos O, Riba L, Gómez-Pérez FJ, Tusié-Luna MT. 2009. Hypoalphalipoproteinemia in populations of Native American ancestry: An opportunity to assess the interaction of genes and the environment. Curr Opin Lipidol 20:92–97.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664.

Almaça J, Molina J, Menegaz D, Pronin AN, Tamayo A, Slepak V, Berggren PO, Caicedo A. 2016. Human Beta Cells Produce and Release Serotonin to Inhibit Glucagon Secretion from Alpha Cells. Cell Rep 17:3281–3291.

Arciero E, Kraaijenbrink T, Asan, Haber M, Mezzavilla M, Ayub Q, Wang W, Pingcuo Z, Yang H, Wang J, et al. 2018. Demographic history and genetic adaptation in the himalayan region inferred from genome-wide SNP genotypes of 49 populations. Mol Biol Evol 35:1916–1933.

Asito AS, Moormann AM, Kiprotich C, Ng'ang'A ZW, Ploutz-Snyder R, Rochford R. 2008. Alterations on peripheral B cell subsets following an acute uncomplicated clinical malaria infection in children. Malar J 7.

Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, et al. 2015. A global reference for human genetic variation. Nature 526:68–74.

Balentine CM, Bolnick DA. 2022. Parallel evolution in human populations: A biocultural perspective. Evolutionary Anthropology: Issues, News, and Reviews 31:302–316. Available from: https://doi.org/10.1002/evan.21956.

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

Barghi N, Tobler R, Nolte V, Jakšić AM, Mallard F, Otte KA, Dolezal M, Taus T, Kofler R, Schlötterer C. 2019. Genetic redundancy fuels polygenic adaptation in Drosophila. PLoS Biol 17.

Bartels HC, Postle JD, Downey P, Brennan DJ. 2018. Placenta accreta spectrum: A review of pathology, molecular biology, and biomarkers. Dis Markers 2018.

Beall CM. 2007. Two routes to functional adaptation: Tibetan and andean high-altitude natives. In the Light of Evolution 1:239–255.

Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, Li C, Li JC, Liang Y, McCormack M, et al. 2010. Natural selection on EPAS1 (HIF2α) associated with low hemoglobin concentration in Tibetan highlanders. Proc Natl Acad Sci U S A 107:11459–11464.

Becerra-Valdivia L, Higham T. 2020. The timing and effect of the earliest human arrivals in North America. Nature 584:93–97.

Benjamin V, M AJ. 2014. Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. Science (1979) 343:1017–1021. Available from: https://doi.org/10.1126/science.1245938.

Benyshek DC, Martin JF, Johnston CS. 2001. A reconsideration of the origins of the type 2 diabetes epidemic among native Americans and the implications for intervention policy. Med Anthropol 20:25–64. Available from: https://doi.org/10.1080/01459740.2001.9966186.

Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, Chen Y, Felkel S, Hallast P, Kamm J, et al. 2020. Insights into human genetic variation and population history from 929 diverse genomes. Science (1979) 367.

Bergström A, Stringer C, Hajdinjak M, Scerri EML, Skoglund P. 2021. Origins of modern human ancestry. Nature 590:229–237. Available from: https://doi.org/10.1038/s41586-021-03244-5.

Bezy O, Tran TT, Pihlajamäki J, Suzuki R, Emanuelli B, Winnay J, Mori MA, Haas J, Biddinger SB, Leitges M, et al. 2011. PKCδ regulates hepatic insulin sensitivity and hepatosteatosis in mice and humans. Journal of Clinical Investigation 121:2504–2517.

Bhatia G, Patterson N, Pasaniuc B, Zaitlen N, Genovese G, Pollack S, Mallick S, Myers S, Tandon A, Spencer C, et al. 2011. Genome-wide comparison of African-ancestry populations from CARe and other cohorts reveals signals of natural selection. Am J Hum Genet 89:368–381.

Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, Mei R, Scherer SW, Julian CG, Wilson MJ, López Herráez D, et al. 2010. Identifying Signatures of Natural Selection in Tibetan and Andean Populations Using Dense Genome Scan Data. PLoS Genet 6:e1001116. Available from: https://doi.org/10.1371/journal.pgen.1001116.

Biswas S, Shahriar S, Giangreco NP, Arvanitis P, Winkler M, Tatonetti NP, Brunken WJ, Cutforth T, Agalliu D. 2022. Mural Wnt/β-catenin signaling regulates Lama2 expression to promote neurovascular unit maturation. Development 149:dev200610. Available from: https://doi.org/10.1242/dev.200610.

Blondin DP, Labbé SM, Phoenix S, Guérin B, Turcotte ÉE, Richard D, Carpentier AC, Haman F. 2015. Contributions of white and brown adipose tissues and skeletal muscles to acute cold-induced metabolic responses in healthy men. Journal of Physiology 593:701–714.

Borda V, Alvim I, Mendes M, Silva-Carvalho C, Soares-Souza GB, Leal TP, Furlan V, Scliar MO, Zamudio R, Zolini C, et al. The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. Available from: www.pnas.org/cgi/doi/10.1073/pnas.2013773117.

Bouhamida E, Morciano G, Pedriali G, Ramaccini D, Tremoli E, Giorgi C, Pinton P, Patergnani S. 2023. The Complex Relationship between Hypoxia Signaling, Mitochondrial Dysfunction and Inflammation in Calcific Aortic Valve Disease: Insights from the Molecular Mechanisms to Therapeutic Approaches. Int J Mol Sci 24.

Bravo M, Dileepan T, Dolan M, Hildebrand J, Wolford J, Hanson ID, Hamilton SE, Frosch AE, Burrack KS. 2024. IL-15 Complex–Induced IL-10 Enhances Plasmodium -specific CD4+ T Follicular Helper Differentiation and Antibody Production. The Journal of Immunology 212:992–1001.

Bros M, Haas K, Moll L, Grabbe S. 2019. Rhoa as a key regulator of innate and adaptive immunity. Cells 8.

Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM. 2018. Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. Cell 173:53-61.e9.

Burel JG, Apte SH, Groves PL, Klein K, McCarthy JS, Doolan DL. 2016. Reduced Plasmodium Parasite Burden Associates with CD38+ CD4+ T Cells Displaying Cytolytic Potential and Impaired IFN-γ Production. PLoS Pathog 12.

Burns SO, Zenner HL, Plagnol V, Curtis J, Mok K, Eisenhut M, Kumararatne D, Doffinger R, Thrasher AJ, Nejentsev S. 2012. LRBA gene deletion in a patient presenting with autoimmunity without hypogammaglobulinemia. Journal of Allergy and Clinical Immunology 130:1428–1432.

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature 562:203–209.

Cabrera A, Neculai D, Kain KC. 2014. CD36 and malaria: friends or foes? A decade of data provides some answers. Trends Parasitol 30:436–444. Available from: https://www.sciencedirect.com/science/article/pii/S1471492214001275.

Cagliani R, Sironi M. 2013. Pathogen-Driven Selection in the Human Genome. Int J Evol Biol 2013:1–6.

Cahill AG, Beigi R, Heine RP, Silver RM, Wax JR. 2018. Placenta Accreta Spectrum. Am J Obstet Gynecol 219:B2–B16.

Calderon-Dominguez M, Mir JF, Fucho R, Weber M, Serra D, Herrero L. 2016. Fatty acid metabolism and the basis of brown adipose tissue function. Adipocyte 5:98–118.

Carabarin-Lima A, González-Vázquez MC, Rodríguez-Morales O, Baylón-Pacheco L, Rosales-Encina JL, Reyes-López PA, Arce-Fonseca M. 2013. Chagas disease (American trypanosomiasis) in Mexico: An update. Acta Trop 127:126–135. Available from: https://www.sciencedirect.com/science/article/pii/S0001706X13001058.

Cazorla O, Aït Mou Y, Goret L, Vassort G, Dauzat M, Lacampagne A, Tanguy S, Obert P. 2006. Effects of high-altitude exercise training on contractile function of rat skinned cardiomyocyte. Cardiovasc Res 71:652–660.

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 4.

Chen J, Meng Y, Zhou J, Zhuo M, Ling F, Zhang Y, Du H, Wang X. 2013. Identifying candidate genes for type 2 diabetes mellitus and obesity through gene expression profiling in multiple tissues or cells. J Diabetes Res 2013.

Chen J, Si J, Li Q, Zhang W, He J. 2024. Unlocking the potential of senescence-related gene signature as a diagnostic and prognostic biomarker in sepsis: insights from meta-analyses, single-cell RNA sequencing, and in vitro experiments. Aging Vol. 16.

Chen Yang, Huang L, Qi X, Chen C. 2019. Insulin receptor trafficking: Consequences for insulin sensitivity and diabetes. Int J Mol Sci 20.

Chen Yunzhao, Wang D, Peng H, Chen X, Han X, Yu J, Wang W, Liang L, Liu Z, Zheng Y, et al. 2019. Epigenetically upregulated oncoprotein PLCE1 drives esophageal carcinoma angiogenesis and proliferation via activating the PI-PLCϵ-NF-κB signaling pathway and VEGF-C/ Bcl-2 expression. Mol Cancer 18.

Chen Zixian, Sun Y, Yang N, Nan J, Cao Likun, Zhao L, Liu S, Xu J, Li Yuxi, He X, et al. 2025. High altitudes, deeper insights: multicenter cardiovascular magnetic resonance study on hypertrophic cardiomyopathy. Eur Radiol 35:3883–3894.

Childebayeva A, Goodrich JM, Leon-Velarde F, Rivera-Chira M, Kiyamu M, Brutsaert TD, Dolinoy DC, Bigham AW. 2021. Genome-Wide Epigenetic Signatures of Adaptive Developmental Plasticity in the Andes. Genome Biol Evol 13.

Chilton PM, Rezzoug F, Fugier-Vivier I, Weeter LA, Xu H, Huang Y, Ray MB, Ildstad ST. 1995. Flt3-Ligand Treatment Prevents Diabetes in NOD Mice. Available from: http://diabetesjournals.org/diabetes/article-pdf/53/8/1995/653898/zdb00804001995.pdf.

Choi MJ, Maibach HI. 2005. Role of Ceramides in Barrier Function of Healthy and Diseased Skin.

Choi W, Moon JH, Kim H. 2020. Serotonergic regulation of energy metabolism in peripheral tissues. Journal of Endocrinology 245:R1–R10.

Chu W, Wan L, Zhao D, Qu X, Cai F, Huo R, Wang N, Zhu J, Zhang C, Zheng F, et al. 2012. Mild hypoxia-induced cardiomyocyte hypertrophy via up-regulation of HIF-1α-mediated TRPC signalling. J Cell Mol Med 16:2022–2034.

Cossrow N, Falkner B. 2004. Race/ethnic issues in obesity and obesity-related comorbidities. In: Journal of Clinical Endocrinology and Metabolism. Vol. 89. p. 2590–2594.

Cossu F. 2010. Genetics of SCID. Ital J Pediatr 36.

Coto E, Díaz-Corte C, Tranche S, Gómez J, Alonso B, Iglesias S, Reguero JR, López-Larrea C, Coto-Segura P. 2018. Gene variants in the NF-KB pathway (NFKB1, NFKBIA, NFKBIZ) and their association with type 2 diabetes and impaired renal function. Hum Immunol 79:494–498. Available from: https://www.sciencedirect.com/science/article/pii/S0198885918300818.

Couto-Silva CM, Nunes K, Venturini G, Castro E Silva MA, Pereira L V, Comas D, Pereira A, Hünemeier T. 2023. Indigenous people from Amazon show genetic signatures of pathogen-driven selection. Available from: http://ensembl.org/.

Croock D, Swart Y, Schurz H, Petersen DC, Möller M, Uren C. 2025. Confirmation of HLA-II associations with TB susceptibility in admixed African samples. Available from: https://elifesciences.org/reviewed-preprints/99200v3.

Dannemann M, Racimo F. 2018. Something old, something borrowed: admixture and adaptation in human evolution. Curr Opin Genet Dev 53:1–8.

Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi M, Excoffier L. 2013. Evidence for polygenic adaptation to pathogens in the human genome. Mol Biol Evol 30:1544–1558.

Delaneau O, Zagury JF, Marchini J. 2013. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods 10:5–6.

Denda M, Horii I, Takahashi M, Hara M, Tagami H. 1993. Age-and sex-dependent change in stratum corneum sphingolipids. Springer-Verlag.

Deng L, Zhang C, Yuan K, Gao Y, Pan Y, Ge X, He Y, Yuan Y, Lu Y, Zhang X, et al. 2019. Prioritizing natural-selection signals from the deep-sequencing genomic data suggests multi-variant adaptation in Tibetan highlanders. Natl Sci Rev 6:1201–1222.

Deschamps M, Laval G, Fagny M, Itan Y, Abel L, Casanova JL, Patin E, Quintana-Murci L. 2016. Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. Am J Hum Genet 98:5–21.

Dyson JJ, Abbasi F, Varadkar P, McCright B. 2022. Growth arrest of PPP2R5C and PPP2R5D double knockout mice indicates a genetic interaction and conserved function for these PP2A B subunits. FASEB Bioadv 4:273–282.

Eda S, Sherman IW. 2004. Plasmodium falciparum-infected erythrocytes bind to the RGD motif of fibronectin via the band 3-related adhesin. Exp Parasitol 107:157–162. Available from: https://www.sciencedirect.com/science/article/pii/S0014489404001092.

Edwards M, Bigham A, Tan J, Li S, Gozdzik A, Ross K, Jin L, Parra EJ. 2010. Association of the OCA2 polymorphism His615Arg with melanin content in East Asian populations: Further evidence of convergent evolution of skin pigmentation. PLoS Genet 6.

Enard D, Petrov DA. 2018. Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. Cell 175:360-371.e13. Available from: https://doi.org/10.1016/j.cell.2018.08.034.

Espinosa-Cantellano M, Marti´nez A, Marti´nez-Palomo M. 2000. Pathogenesis of Intestinal Amebiasis: From Molecules to Disease.

Fan S, Hansen MEB, Lo Y, Tishkoff SA. 2016. Going global by adapting local: A review of recent human adaptation. Available from: https://www.science.org.

Fang J, Qian Y, Chen J, Xu D, Cao N, Zhu G, Hu W, Hu H, Qian N, Yang S, et al. 2023. Human antigen R regulates autophagic flux by stabilizing autophagy-associated mRNA in calcific aortic valve disease. Cardiovasc Res 119:2117–2129. Available from: https://doi.org/10.1093/cvr/cvad077.

Felber J-P, Golay A. 2002. PAPER Pathways from obesity to diabetes. Int J Obes 26:39–45. Available from: www.nature.com/ijo.

Ferraretti G, Abondio P, Alberti M, Dezi A, Sherpa PT, Cocco P, Tiriticco M, Di Marcello M, Gnecchi-Ruscone GA, Natali L, et al. 2024. Archaic introgression contributed to shape the adaptive modulation of

angiogenesis and cardiovascular traits in human high-altitude populations from the Himalayas. Elife 12. Available from: https://elifesciences.org/articles/89815.

Ferraretti G, Rill A, Abondio P, Smith K, Ojeda-Granados C, De Fanti S, Alberti M, Izzi M, Sherpa PT, Cocco P, et al. 2025. Convergent evolution of complex adaptive traits modulates angiogenesis in high-altitude Andean and Himalayan human populations. Commun Biol 8:377. Available from: https://doi.org/10.1038/s42003-025-07813-6.

Ferraretti G, Alberti M, Cicolini R, Pognant Viù S, Sarno S, Sazzini M, under review. Complex adaptive evolution at thyroid hormone, insulin and glycerolipid pathways improved energy metabolism in Eurasian populations from high-latitude cold environments.

Ferraretti G, Iannuzzi V, Gentilini D, Calzari L, Sarno S, Moretti E, Dasso MC, Sevini F, Pettener D, Franceschi C, Franceschi ZA, et al., in preparation. Evidence of local adaptation to Chagas disease in the Wichís: a Genome-Wide and Epigenetic perspective.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. Mol Biol Evol 31:1275–1291.

Fontana JM, Dugué B, Capodaglio P. 2024. Prolonged or Repeated Cold Exposure: From Basic Physiological Adjustment to Therapeutic Effects. In: Capodaglio P, editor. Whole-Body Cryostimulation: Clinical Applications. Cham: Springer International Publishing. p. 3–19. Available from: https://doi.org/10.1007/978-3-031-18545-8_1.

Fu W, Akey JM. 2013. Selection and adaptation in the human genome. Annu Rev Genomics Hum Genet 14:467–489.

Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admettla A, Pattini L, Nielsen R. 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet 7.

Gamboa-León R, Ramirez-Gonzalez C, Pacheco-Tucuch FS, O'Shea M, Rosecrans K, Pippitt J, Dumonteil E, Buekens P. 2014. Seroprevalence of Trypanosoma cruzi among mothers and children in rural Mayan communities and associated reproductive outcomes. American Journal of Tropical Medicine and Hygiene 91:348–353.

Garg M, Kanojia D, Okamoto R, Jain S, Madan V, Chien W, Sampath A, Ding L-W, Xuan M, Said JW, et al. 2014. Laminin-5γ-2 (LAMC2) Is Highly Expressed in Anaplastic Thyroid Carcinoma and Is Associated With Tumor Progression, Migration, and Invasion by Modulating Signaling of EGFR. J Clin Endocrinol Metab 99:E62–E72. Available from: https://doi.org/10.1210/jc.2013-2994.

Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps. PLoS Genet 11:1–32.

Ghadimi S, Jamee M, Abolhassani H, Parvaneh N, Rezaei N, Delavari S, Sadeghi-Shabestari M, Tabatabaei SR, Fahimzad A, Armin S, et al. 2023. Demographic, clinical, immunological, and molecular features of iranian national cohort of patients with defect in DCLRE1C gene. Allergy, Asthma and Clinical Immunology 19.

Ghosh S, Park CH, Lee J, Lee N, Zhang R, Huesing C, Reijnders D, Sones J, Münzberg H, Redman L, et al. 2021. Maternal cold exposure induces distinct transcriptome changes in the placenta and fetal brown adipose tissue in mice. BMC Genomics 22.

Gittelman RM, Schraiber JG, Vernot B, Mikacenic C, Wurfel MM, Akey JM. 2016. Archaic Hominin Admixture Facilitated Adaptation to Out-of-Africa Environments. Current Biology 26:3375–3382. Available from: https://www.sciencedirect.com/science/article/pii/S0960982216312672.

Gnecchi-Ruscone GA, Abondio P, De Fanti S, Sarno S, Sherpa MG, Sherpa PT, Marinelli G, Natali L, Di Marcello M, Peluzzi D, et al. 2018. Evidence of polygenic adaptation to high altitude from Tibetan and Sherpa genomes. Genome Biol Evol 10:2919–2930.

Gnecchi-Ruscone GA, Sarno S, De Fanti S, Gianvincenzo L, Giuliani C, Boattini A, Bortolini E, Corcia T Di, Mellado CS, Francia TJD, et al. 2019. Dissecting the pre-Columbian genomic ancestry of Native Americans along the Andes–Amazonia divide. Mol Biol Evol 36:1254–1269.

Goh W, Yamamoto SY, Thompson KS, Bryant-Greenwood GD. 2013. Relaxin, its rfeceptor (RXFP1), and insulin-like peptide 4 expression through gestation and in placenta accreta. Reproductive Sciences 20:968–980.

Gouy A, Daub JT, Excoffier L. 2017. Detecting gene subnetworks under selection in biological pathways. Nucleic Acids Res 45.

Gouy A, Excoffier L. 2020. Polygenic Patterns of Adaptive Introgression in Modern Humans Are Mainly Shaped by Response to Pathogens. Mol Biol Evol 37:1420–1433. Available from: https://doi.org/10.1093/molbev/msz306.

Gower G, Picazo PI, Fumagalli M, Racimo F. 2021. Detecting adaptive introgression in human evolution using convolutional neural networks. Elife 10.

Grarup N, Moltke I, Andersen MK, Dalby M, Vitting-Seerup K, Kern T, Mahendran Y, Jørsboe E, Larsen CVL, Dahl-Petersen IK, et al. 2018. Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. Nat Genet 50:172–174.

Green R, Krause J, Briggs A, Rasilla Vives M de la, Fortea Pérez FJ. 2010. A draft sequence of the neandertal genome. Science (1979) 328:710–722.

Guo Y, Liu Z, Wang M. 2021. NFKB1-mediated downregulation of microRNA-106a promotes oxidative stress injury and insulin resistance in mice with gestational hypertension. Cytotechnology 73:115–126.

Guzy RD, Hoyos B, Robin E, Chen H, Liu L, Mansfield KD, Simon MC, Hammerling U, Schumacker PT. 2005. Mitochondrial complex III is required for hypoxia-induced ROS production and cellular oxygen sensing. Cell Metab 1:401–408. Available from: https://www.sciencedirect.com/science/article/pii/S1550413105001397.

Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, Sukernik R, Utermann G, Pritchard JK, Coop G, Di Rienzo A. 2011. Adaptations to climate-mediated selective pressures in humans. PLoS Genet 7.

Hankir MK, Klingenspor M. 2018. Brown adipocyte glucose metabolism: a heated subject. EMBO Rep 19.

Harding RM, Healy E, Ray AJ, Ellis NS, Flanagan N, Todd C, Dixon C, Sajantila A, Jackson IJ, Birch-Machin MA, et al. 2000. Evidence for Variable Selective Pressures at MC1R.

Harris AM, DeGiorgio M. 2020. A likelihood approach for uncovering selective sweep signatures from haplotype data. Mol Biol Evol 37:3023–3046.

He Y, Guo Y, Zheng W, Yue T, Zhang H, Wang B, Feng Z, Ouzhuluobu, Cui C, Liu K, et al. 2023. Polygenic adaptation leads to a higher reproductive fitness of native Tibetans at high altitude. Current Biology 33:4037-4051.e5. Available from: https://www.sciencedirect.com/science/article/pii/S0960982223010643.

Heather LC, Cole MA, Tan JJ, Ambrose LJA, Pope S, Abd-Jamil AH, Carter EE, Dodd MS, Yeoh KK, Schofield CJ, et al. 2012. Metabolic adaptation to chronic hypoxia in cardiac mitochondria. Basic Res Cardiol 107.

Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, Project 1000 Genomes, Sella G, Przeworski M. 2011. Classic Selective Sweeps Were Rare in Recent Human Evolution. Science (1979) 331:920–924. Available from: https://doi.org/10.1126/science.1198878.

Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowet PA, Bennettt S, Brewster D, Mcmichael AJ, Greenwood BM. Common West African HLA antigens are associated with protection from severe malaria.

Hill WG, Robertson A. Linkage Disequilibrium in Finite Populations.

Hoppeler H, Vogt M, Weibel ER, Flück M. 2003. Response of skeletal muscle mitochonrial to hypoxia. Exp Physiol 88:109–119.

Horscroft JA, Kotwica AO, Laner V, West JA, Hennis PJ, Levett DZH, Howard DJ, Fernandez BO, Burgess SL, Ament Z, et al. 2017. Metabolic basis to sherpa altitude adaptation. Proc Natl Acad Sci U S A 114:6382–6387.

Hosokawa H, Ninomiya H, Kitamura Y, Fujiwara K, Masaki T. 2002. Vascular endothelial cells that express dystroglycan are involved in angiogenesis. J Cell Sci 115:1487–1496. Available from: https://doi.org/10.1242/jcs.115.7.1487.

Hou H, Kang Y, Zeng Y, Li Y, Shang J. 2018. Interleukin-7 augments CD8+ T cells function and promotes viral clearance in chronic hepatitis C virus infection. Cytokine 102:26–33. Available from: https://www.sciencedirect.com/science/article/pii/S104346661730385X.

Hsieh PH, Hallmark B, Watkins J, Karafet TM, Osipova LP, Gutenkunst RN, Hammer MF. 2017. Exome sequencing provides evidence of polygenic adaptation to a fat-rich animal diet in indigenous siberian populations. Mol Biol Evol 34:2913–2926.

Hu H, Petousi N, Glusman G, Yu Y, Bohlender R, Tashi T, Downie JM, Roach JC, Cole AM, Lorenzo FR, et al. 2017. Evolutionary history of Tibetans inferred from whole-genome sequencing. PLoS Genet 13:1–22.

Huan Y, Quan H, Jia B, Hao G, Shi Z, Zhao T, Yuan Y, Yuan F, Dong Y, Liang G. 2023. High-altitude cerebral hypoxia promotes mitochondrial dysfunction and apoptosis of mouse neurons. Front Mol Neurosci 16.

Huang K, Hong Q. 2024. The impact of global warming on obesity. J Popul Econ 37.

Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Somel M, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512:194–197.

Javerzat S, Franco M, Herbert J, Platonova N, Peille AL, Pantesco V, De Vos J, Assou S, Bicknell R, Bikfalvi A, et al. 2009. Correlating global gene regulation to angiogenesis in the developing chick extra-embryonic vascular system. PLoS One 4.

Jeong C, Witonsky DB, Basnyat B, Neupane M, Beall CM, Childs G, Craig SR, Novembre J, Di Rienzo A. 2018. Detecting past and ongoing natural selection among ethnically Tibetan women at high altitude in Nepal. PLoS Genet 14:e1007650. Available from: https://doi.org/10.1371/journal.pgen.1007650.

Jiao S, Zhang Y, Yang X, Wang J, Li Z. 2025. Alternative Splicing Analysis Reveals Adrenergic Signaling as a Novel Target for Protein Arginine Methyltransferase 5 (PRMT5) in the Heart. Int J Mol Sci 26.

Jiménez-Kaufmann A, Chong AY, Cortés A, Quinto-Cortés CD, Fernandez-Valverde SL, Ferreyra-Reyes L, Cruz-Hervert LP, Medina-Muñoz SG, Sohail M, Palma-Martinez MJ, et al. 2022. Imputation Performance in Latin American Populations: Improving Rare Variants Representation With the Inclusion of Native American Genomes. Front Genet 12.

Jing W, Liu J, Liu M. Global Trends and Regional Differences in Hepatitis C Virus Infection Prevalence and Implications for Prevention-Worldwide, 1990−2017. Available from: http://weekly.chinacdc.cn/.

Johnson J, Yang Y, Bian Z, Schena G, Li Y, Zhang X, Eaton DM, Gross P, Angheloiu A, Shaik A, et al. 2023. Systemic Hypoxemia Induces Cardiomyocyte Hypertrophy and Right Ventricular Specific Induction of Proliferation. Circ Res 132:723–740.

Ju D, Mathieson I. 2021. The evolution of skin pigmentation-associated variation in West Eurasia. 118.

Julian CG, Moore LG. 2019. Human genetic adaptation to high altitude: Evidence from the andes. Genes (Basel) 10.

Kalluri R. 2003. Basement membranes: Structure, assembly and role in tumour angiogenesis. Nat Rev Cancer 3:422–433.

Kanchan K, Pati SS, Mohanty S, Mishra SK, Sharma SK, Awasthi S, Venkatesh V, Habib S. 2015. Polymorphisms in host genes encoding NOSII, C-reactive protein, and adhesion molecules thrombospondin and E-selectin are risk factors for Plasmodium falciparum malaria in India. European Journal of Clinical Microbiology and Infectious Diseases 34:2029–2039.

Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Available from: http://www.genome.ad.jp/kegg/.

Keele GR, Prokop JW, He H, Holl K, Littrell J, Deal A, Francic S, Cui L, Gatti DM, Broman KW, et al. 2018. Genetic Fine-Mapping and Identification of Candidate Genes and Variants for Adiposity Traits in Outbred Rats. Obesity 26:213–222.

Kern AD, Schrider DR. 2016. Discoal: Flexible coalescent simulations with selection. Bioinformatics 32:3839–3841.

Koch CA, Sharda P, Patel J, Gubbi S, Bansal R, Bartel MJ. 2021. Climate Change and Obesity. Hormone and Metabolic Research 53:575–587.

Kolář F, Ošťádal B. 2004. Molecular Mechanisms of Cardiac Protection by Adaptation to Chronic Hypoxia. Physiol. Res 53:3–13. Available from: http://www.biomed.cas.cz/physiolres.

Kone A, Diarra B, Cohen K, Diabate S, Kone B, Diakite MT, Diarra H, Sanogo M, Togo ACG, Sarro Y dit S, et al. 2019. Differential HLA allele frequency in Mycobacterium africanum vs Mycobacterium tuberculosis in Mali. HLA 93:24–31.

Konstantinova I, Nikolova G, Ohara-Imaizumi M, Meda P, Kučera T, Zarbalis K, Wurst W, Nagamatsu S, Lammert E. 2007. EphA-Ephrin-A-Mediated β Cell Communication Regulates Insulin Secretion from Pancreatic Islets. Cell 129:359–370. Available from: https://www.sciencedirect.com/science/article/pii/S0092867407003686.

Kosoy R, Ransom M, Chen H, Marconi M, MacCiardi F, Glorioso N, Gregersen PK, Cusi D, Seldin MF. 2011. Evidence for malaria selection of a CR1 haplotype in Sardinia. Genes Immun 12:582–588.

Kranenburg O, Gebbink MFBG, Voest EE. 2004. Stimulation of angiogenesis by Ras proteins. Biochim Biophys Acta Rev Cancer 1654:23–37.

Krock Bryan L, Skuli Nicolas, Simon M. Celeste. 2011. Hypoxia-Induced Angiogenesis: Good and Evil. Genes Cancer 2:1117–1133. Available from: https://journals.sagepub.com/action/showAbstract.

Kuo DS, Labelle-Dumais C, Gould DB. 2012. Col4a1 and col4a2 mutations and disease: Insights into pathogenic mechanisms and potential therapeutic targets. Hum Mol Genet 21.

Kwiatkowski DP. 2005. How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria.

Labbé SM, Caron A, Bakan I, Laplante M, Carpentier AC, Lecomte R, Richard D. 2015. In vivo measurement of energy substrate contribution to cold-induced brown adipose tissue thermogenesis. FASEB Journal 29:2046–2058.

Lamason RL, Mohideen M-AP, Mest JR, Wong AC, Norton -Heather L, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, et al. SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans. Available from: https://www.science.org.

Lara-Riegos JC, Ortiz-López MG, Peña-Espinoza BI, Montúfar-Robles I, Peña-Rico MA, Sánchez-Pozos K, Granados-Silvestre MA, Menjivar M. 2015. Diabetes susceptibility in Mayas: Evidence for the involvement of polymorphisms in HHEX, HNF4α, KCNJ11, PPARγ, CDKN2A/2B, SLC30A8, CDC123/CAMK1D, TCF7L2, ABCA1 and SLC16A11 genes. Gene 565:68–75. Available from: https://www.sciencedirect.com/science/article/pii/S0378111915003716.

Laurent A-R, J. JM, Subhash K, Alasdair M, Klara D, Loren G, Farbod B, Baback G, Ma L, A. PF, et al. 2011. The Shaping of Modern Human Immune Systems by Multiregional Admixture with Archaic Humans. Science (1979) 334:89–94. Available from: https://doi.org/10.1126/science.1209202.

Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. PLoS Genet 8.

Lee MH, Yang HI, Lu SN, Lin YJ, Jen CL, Wong KH, Chan SY, Chen LC, Wang LY, L'Italien G, et al. 2015. Polymorphisms near the IFNL3 Gene Associated with HCV RNA Spontaneous Clearance and Hepatocellular Carcinoma Risk. Sci Rep 5.

Lee P, Bova R, Schofield L, Bryant W, Dieckmann W, Slattery A, Govendir MA, Emmett L, Greenfield JR. 2016. Brown Adipose Tissue Exhibits a Glucose-Responsive Thermogenic Biorhythm in Humans. Cell Metab 23:602–609.

Lei L, Liu M, Ma D, Lei X, Zeng S, Li P, Huang K, Lyu J, Lei Q. 2024. Cardioprotective effects of high-altitude adaptation in cardiac surgical patients: a retrospective cohort study with propensity score matching. Front Cardiovasc Med 11.

Leung L. 2016. Diabetes mellitus and the Aboriginal diabetic initiative in Canada: An update review. J Family Med Prim Care 5:259.

Levy SB, Klimova TM, Zakharova RN, Fedorov AI, Fedorova VI, Baltakhinova ME, Bondy M, Atallah D, Thompson-Vasquez J, Dong K, et al. 2022. Brown adipose tissue thermogenesis among young adults in northeastern Siberia and Midwest United States and its relationship with other biological adaptations to cold climates. American Journal of Human Biology 34.

Levy SB, Leonard WR, Tarskaia LA, Klimova TM, Fedorova VI, Baltakhinova ME, Krivoshapkin VG, Snodgrass JJ. 2013. Seasonal and socioeconomic influences on thyroid function among the Yakut (Sakha) of Eastern Siberia. American Journal of Human Biology 25:814–820.

Lewis CM, Vassos E. 2020. Polygenic risk scores: From research tools to clinical instruments. Genome Med 12.

Li H, Fredriksson L, Li X, Eriksson U. 2003. PDGF-D is a potent transforming and angiogenic growth factor. Oncogene 22:1501–1510.

Li Norman P, van Vugt Mark, Colarelli Stephen M. 2017. The Evolutionary Mismatch Hypothesis: Implications for Psychological Science. Curr Dir Psychol Sci 27:38–44. Available from: https://doi.org/10.1177/0963721417731378.

Li X, Zhou X, Ye Y, Li Y, Li J, Privratsky B, Wu E, Gao H, Huang C, Wu M. 2014. Lyn regulates inflammatory responses in Klebsiella pneumoniae infection via the p38/NF-κB pathway. Eur J Immunol 44:763–773.

Li Y, Huang W, Yu Q, Cheng Y-T, Kong Q-P. 2016. Lower mitochondrial DNA content relates to high-altitude adaptation in Tibetans. Mitochondrial DNA Part A 27:753–757. Available from: https://doi.org/10.3109/19401736.2014.915526.

Li YC, Tian JY, Liu FW, Yang BY, Gu KSY, Rahman ZU, Yang LQ, Chen FH, Dong GH, Kong QP. 2019. Neolithic millet farmers contributed to the permanent settlement of the Tibetan Plateau by adopting barley agriculture. Natl Sci Rev 6:1005–1013.

Liang P, Zu J, Yin J, Li H, Gao L, Cui F, Wang F, Liang X, Zhuang G. 2015. The independent impact of newborn hepatitis B vaccination on reducing HBV prevalence in China, 1992–2006: A mathematical model analysis. J Theor Biol 386:115–121. Available from: https://www.sciencedirect.com/science/article/pii/S0022519315004348.

Liang X, Bi S, Yang W, Wang L, Cui G, Cui F, Zhang Y, Liu J, Gong X, Chen Y, et al. 2009. Epidemiological serosurvey of Hepatitis B in China—Declining HBV prevalence due to Hepatitis B vaccination. Vaccine 27:6550–6557. Available from: https://www.sciencedirect.com/science/article/pii/S0264410X0901233X.

Lin Z, Lu Y, Yu G, Teng H, Wang B, Yang Y, Li Q, Sun Z, Xu S, Wang W, et al. 2023. Genome-wide DNA methylation landscape of four Chinese populations and epigenetic variation linked to Tibetan high-altitude adaptation. Sci China Life Sci 66:2354–2369.

Lindo J, Haas R, Hofman C, Apata M, Moraga M, Verdugo RA, Watson JT, Llave CV, Witonsky D, Beall C, et al. 2018. A N T H R O P O L O G Y The genetic prehistory of the Andean highlands 7000 years BP though European contact. Available from: https://www.science.org.

Little SC, Curran J, Makara MA, Kline CF, Ho HT, Xu Z, Wu X, Polina I, Musa H, Meadows AM, et al. 2015. Protein phosphatase 2A regulatory subunit B56α limits phosphatase activity in the heart. Sci Signal 8.

Lorca RA, Houck JA, Laurent LC, Matarazzo CJ, Baker K, Horii M, Nelson KK, Bales ES, Euser AG, Parast MM, et al. 2021. High altitude regulates the expression of AMPK pathways in human placenta. Placenta 104:267–276.

De Lorenzo C, Greco A, Fiorentino TV, Mannino GC, Hribal ML. 2013. Variants of insulin-signaling inhibitor genes in type 2 diabetes and related metabolic abnormalities. Int J Genomics 2013.

Lorenzo C, Serrano-Rios M, Martinez-Larrad MT, Gabriel R, Williams K, Gonzalez-Villalpando C, Stern MP, Hazuda HP, Haffner SM. Was the Historic Contribution of Spain to the Mexican Gene Pool Partially Responsible for the Higher Prevalence of Type 2 Diabetes in Mexican-Origin Populations? Available from: http://diabetesjournals.org/care/article-pdf/24/12/2059/587520/dc1201002059.pdf.

Ma Y, Zhou X. 2021. Genetic prediction of complex traits with polygenic scores: a statistical review. Trends in Genetics 37:995–1011.

Machado SA, Pasquarelli-do-Nascimento G, da Silva DSS, Farias GR, de Oliveira Santos I, Baptista LB, Magalhães KG. 2022. Browning of the white adipose tissue regulation: new insights into nutritional and metabolic relevance in health and diseases. Nutr Metab (Lond) 19.

Mallet RT, Manukhina EB, Ruelas SS, Caffrey JL, Downey HF. 2018. Cardioprotection by intermittent hypoxia conditioning: evidence, mechanisms, and therapeutic potential. Am J Physiol Heart Circ Physiol 315:216–232. Available from: www.ajpheart.org.

Mallov S. Cold effects in rat: plasma and adipose tissue free fatty acids and adipose lipase'.

Manolache A, Selescu T, Maier GL, Mentel M, Ionescu AE, Neacsu C, Babes A, Szedlacsek SE. 2020. Regulation of TRPM8 channel activity by Src-mediated tyrosine phosphorylation. J Cell Physiol 235:5192–5203.

Marcher A-B, Loft A, Nielsen R, Vihervaara T, Madsen JGS, Sysi-Aho M, Ekroos K, Mandrup S. 2015. RNA-Seq and Mass-Spectrometry-Based Lipidomics Reveal Extensive Changes of Glycerolipid Pathways in Brown Adipose Tissue in Response to Cold. Cell Rep 13:2000–2013. Available from: https://www.sciencedirect.com/science/article/pii/S2211124715012589.

Marnetto D, Huerta-Sánchez E. 2017. Haplostrips: revealing population structure through haplotype visualization. Methods Ecol Evol 8:1389–1392.

McClellan J, King M-C. 2010. Genetic Heterogeneity in Human Disease. Cell 141:210–217. Available from: https://www.sciencedirect.com/science/article/pii/S009286741000320X.

Medina-Torres I, Vázquez-Chagoyán JC, Rodríguez-Vivas RI, De Oca-Jiménez RM. 2010. Risk factors associated with triatomines and its infection with Trypanosoma cruzi in rural communities from the southern region of the State of Mexico, Mexico. American Journal of Tropical Medicine and Hygiene 82:49–54.

Midgley GF, Bond WJ. 2015. Future of African terrestrial biodiversity and ecosystems under anthropogenic climate change. Nat Clim Chang 5:823–829. Available from: https://doi.org/10.1038/nclimate2753.

Miller JM, Wang Y V. 2022. Ostrich eggshell beads reveal 50,000-year-old social network in Africa. Nature 601:234–239.

Miyazaki T, Kanou Y, Murata Y, Ohmori S, Niwa T, Maeda K, Yamamura H, Seo H. 1996. Molecular Cloning of a Novel Thyroid Hormone-responsive Gene, ZAKI-4, in Human Skin Fibroblasts*. Journal of Biological Chemistry 271:14567–14571. Available from: https://www.sciencedirect.com/science/article/pii/S0021925818468378.

Molina-Moya B, Gomgnimbou MK, Spinasse L, Obasanya J, Oladimeji O, Dacombe R, Edwards T, Daragon XO, Lawson L, Abdurrahman ST, et al. 2018. Mycobacterium tuberculosis complex genotypes circulating in Nigeria based on spoligotyping obtained from Ziehl-Neelsen stained slides extracted DNA. PLoS Negl Trop Dis 12.

Morikawa Y, Heallen T, Leach J, Xiao Y, Martin JF. 2017. Dystrophin-glycoprotein complex sequesters Yap to inhibit cardiomyocyte proliferation. Nature 547:227–231.

Mughal MR, DeGiorgio M. 2019. Localizing and classifying adaptive targets with trend filtered regression. Mol Biol Evol 36:252–270.

Murray AJ, Horscroft JA. 2016. Mitochondrial function at extreme high altitude. Journal of Physiology 594:1137–1149.

Nakamura M, Sadoshima J. 2018. Mechanisms of physiological and pathological cardiac hypertrophy. Nat Rev Cardiol 15:387–407. Available from: https://doi.org/10.1038/s41569-018-0007-y.

Nde PN, Lima MF, Johnson CA, Pratap S, Villalta F. 2012. Regulation and use of the extracellular matrix by Trypanosoma cruzi during early infection. Front Immunol 3.

Nelson MP, Metz AE, Li S, Lowell CA, Steele C. 2009. The absence of Hck, Fgr, and lyn tyrosine kinases augments lung innate immune responses to Pneumocystis murina. Infect Immun 77:1790–1797.

Novembre J, Han E. 2012. Human population structure and the adaptive response to pathogen-induced selection pressures. Philosophical Transactions of the Royal Society B: Biological Sciences 367:878–886.

Nowak MA, Boerlijst MC, Cooke J, Smith JM. 1997. Evolution of genetic redundancy. Nature 388:167–171. Available from: https://doi.org/10.1038/40618.

Nunes MCP, Beaton A, Acquatella H, Bern C, Bolger AF, Echeverría LE, Dutra WO, Gascon J, Morillo CA, Oliveira-Filho J, et al. 2018. Chagas Cardiomyopathy: An Update of Current Clinical Knowledge and Management: A Scientific Statement From the American Heart Association. Circulation 138:e169–e209.

Odera S, Mureithi M, Aballa A, Onyango N, Kazungu S, Ogolla S, Kaiyare G, Anzala O, Oyugi J. 2022. Association between human leukocyte antigen class II (HLA-DRB and-DQB) alleles and outcome of exposure to Mycobacterium tuberculosis: a cross-sectional study in Nairobi, Kenya. Pan African Medical Journal 41.

Oefner CM, Sharkey A, Gardner L, Critchley H, Oyen M, Moffett A. 2015. Collagen type IV at the fetal-maternal interface. Placenta 36:59–68.

Ojeda-Granados C, Abondio P, Setti A, Sarno S, Gnecchi-Ruscone GA, González-Orozco E, De Fanti S, Jiménez-Kaufmann A, Rangel-Villalobos H, Moreno-Estrada A, et al. 2022. Dietary, Cultural, and Pathogens-Related Selective Pressures Shaped Differential Adaptive Evolution among Native Mexican Populations. Mol Biol Evol 39.

Okada H, Tsuzuki T, Murata H. 2018. Decidualization of the human endometrium. Reprod Med Biol 17:220–227.

Oliveira-Cortez A, Melo AC, Chaves VE, Condino-Neto A, Camargos P. 2016. Do HLA class II genes protect against pulmonary tuberculosis? A systematic review and meta-analysis. European Journal of Clinical Microbiology and Infectious Diseases 35:1567–1580.

Omi K, Ohashi J, Patarapotikul J, Hananantachai H, Naka I, Looareesuwan S, Tokunaga K. 2003. CD36 Polymorphism Is Associated with Protection from Cerebral Malaria.

Oppenheimer S. 2012. Out-of-Africa, the peopling of continents and islands: Tracing uniparental gene trees across the map. Philosophical Transactions of the Royal Society B: Biological Sciences 367:770–784.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS Genet 2:2074–2093.

Pedregosa FABIANPEDREGOSA F, Michel V, Grisel OLIVIERGRISEL O, Blondel M, Prettenhofer P, Weiss R, Vanderplas J, Cournapeau D, Pedregosa F, Varoquaux G, et al. 2011. Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. Available from: http://scikit-learn.sourceforge.net.

Pei YF, Liu J, Cheng J, Wu WD, Liu XQ. 2019. Silencing of LAMC2 Reverses Epithelial-Mesenchymal Transition and Inhibits Angiogenesis in Cholangiocarcinoma via Inactivation of the Epidermal Growth Factor Receptor Signaling Pathway. American Journal of Pathology 189:1637–1653.

Peng Q, Qin Y, Chen Z, Deng Y, Xu J, Li S, Qin X. 2013. InflamCorrelation between interleukin-23 receptor gene polymorphisms and risk of hepatitis B virus infection in patients. Mol Med Rep 8:613–620.

Peng Y, Yang Z, Zhang H, Cui C, Qi X, Luo X, Tao X, Wu T, Ouzhuluobu, Basang, et al. 2011. Genetic variations in tibetan populations and high-altitude adaptation at the Himalayas. Mol Biol Evol 28:1075–1081.

Pitulko V V, Nikolsky PA, Girya EY, Basilyan AE, Tumskoy VE, Koulakov SA, Astakhov SN, Pavlova EY, Anisimov MA. The Yana RHS Site: Humans in the Arctic Before the Last Glacial Maximum. Available from: https://www.science.org.

Plaisier M. 2011. Decidualisation and angiogenesis. Best Pract Res Clin Obstet Gynaecol 25:259–271. Available from: https://www.sciencedirect.com/science/article/pii/S1521693410001343.

Porto AG, Brun F, Severini GM, Losurdo P, Fabris E, Taylor MRG, Mestroni L, Sinagra G. 2016. Clinical Spectrum of PRKAG2 Syndrome. Circ Arrhythm Electrophysiol 9:e003121. Available from: https://doi.org/10.1161/CIRCEP.115.003121.

Pritchard JK, Pickrell JK, Coop G. 2010. The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. Current Biology 20.

Pritchard JK, Di Rienzo A. 2010. Adaptation - Not by sweeps alone. Nat Rev Genet 11:665–667.

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, De Filippo C, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505:43–49.

Racimo F, Berg JJ, Pickrell JK. 2018. Detecting polygenic adaptation in admixture graphs. Genetics 208:1565–1584.

Racimo F, Marnetto D, Huerta-Sánchez E. 2017. Signatures of archaic adaptive introgression in present-day human populations. Mol Biol Evol 34:296–317.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. Nature 461:489–494.

Ribeiro MO, Carvalho SD, Schultz JJ, Chiellini G, Scanlan TS, Bianco AC, Brent GA. 2001. Thyroid hormone–sympathetic interaction and adaptive thermogenesis are thyroid hormone receptor isoform–specific. Journal of Clinical Investigation 108:97–105.

Rito T, Vieira D, Silva M, Conde-Sousa E, Pereira L, Mellars P, Richards MB, Soares P. 2019. A dispersal of Homo sapiens from southern to eastern Africa immediately preceded the out-of-Africa migration. Sci Rep 9.

Roberts DD, Sherwood JA, Spitalnik SL, Panton LJ, Howard RJ, Dixit VM, Frazier WA, Miller LH, Ginsburg V. 1985. Thrombospondin binds falciparum malaria parasitized erythrocytes and may mediate cytoadherence. Nature 318:64–66. Available from: https://doi.org/10.1038/318064a0.

Robinson MR, Hemani G, Medina-Gomez C, Mezzavilla M, Esko T, Shakhbazov K, Powell JE, Vinkhuyzen A, Berndt SI, Gustafsson S, et al. 2015. Population genetic differentiation of height and body mass index across Europe. Nat Genet 47:1357–1361.

Roth S, Bergmann H, Jaeger M, Yeroslaviz A, Neumann K, Koenig P-A, Prazeres da Costa C, Vanes L, Kumar V, Johnson M, et al. 2016. Vav Proteins Are Key Regulators of Card9 Signaling for Innate Antifungal Immunity. Cell Rep 17:2572–2583. Available from: https://www.sciencedirect.com/science/article/pii/S2211124716315741.

S. ST, Yingzhong Y, D. HC, Haixia Y, Ga Q, J. WD, Zhenzhong B, R. LF, Jinchuan X, B. JL, et al. 2010. Genetic Evidence for High-Altitude Adaptation in Tibet. Science (1979) 329:72–75. Available from: https://doi.org/10.1126/science.1189406.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449:913–918.

Sachdeva H, Barton NH. 2018. Replicability of introgression under linked, polygenic selection. Genetics 210:1411–1427.

Saleh M, Mathison JC, Wolinski MK, Bensinger SJ, Fitzgerald P, Droin N, Ulevitch RJ, Green DR, Nicholson DW. 2006. Enhanced bacterial clearance and sepsis resistance in caspase-12-deficient mice. Nature 440:1064–1068.

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507:354–357.

Sarkar A, Tindle C, Pranadinata RF, Reed S, Eckmann L, Stappenbeck TS, Ernst PB, Das S. 2017. ELMO1 Regulates Autophagy Induction and Bacterial Clearance during Enteric Infection. Journal of Infectious Diseases 216:1655–1666.

Sato H, Nagashima K, Ogura M, Sato Y, Tahara Y, Ogura K, Yamano G, Sugizaki K, Fujita N, Tatsuoka H, et al. 2016. Src regulates insulin secretion and glucose metabolism by influencing subcellular localization of glucokinase in pancreatic β-cells. J Diabetes Investig 7:171–178.

Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, et al. 2022. Database resources of the national center for biotechnology information. Nucleic Acids Res 50:D20–D26.

Sazzini M, Abondio P, Sarno S, Gnecchi-Ruscone GA, Ragno M, Giuliani C, De Fanti S, Ojeda-Granados C, Boattini A, Marquis J, et al. 2020. Genomic history of the Italian population recapitulates key evolutionary dynamics of both Continental and Southern Europeans. BMC Biol 18.

Sazzini M, Schiavo G, De Fanti S, Martelli PL, Casadio R, Luiselli D. 2014. Searching for signatures of cold adaptations in modern and archaic humans: Hints from the brown adipose tissue genes. Heredity (Edinb) 113:259–267.

Scheinfeldt LB, Tishkoff SA. 2013. Recent human adaptation: Genomic approaches, interpretation and insights. Nat Rev Genet 14:692–702.

Schork NJ, Murray SS, Frazer KA, Topol EJ. 2009. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 19:212–219.

Schrauwen P, van Marken Lichtenbelt WD, Spiegelman BM. 2015. The future of brown adipose tissues in the treatment of type 2 diabetes. Diabetologia 58:1704–1707.

Schwenk R, Banania G, Epstein J, Kim Y, Peters B, Belmonte M, Ganeshan H, Huang J, Reyes S, Stryhn A, et al. 2013. Ex vivo tetramer staining and cell surface phenotyping for early activation markers CD38 and HLA-DR to enumerate and characterize malaria antigen-specific CD8+ T-cells induced in human volunteers immunized with a Plasmodium falciparum adenovirus-vectored malaria vaccine expressing AMA1. Malar J 12.

Scruggs SB, Wang D, Ping P. 2016. PRKCE gene encoding protein kinase C-epsilon—Dual roles at sarcomeres and mitochondria in cardiomyocytes. Gene 590:90–96.

Seo JY, You SW, Shin J-G, Kim Y, Park SG, Won H-H, Kang NG. 2022. GWAS Identifies Multiple Genetic Loci for Skin Color in Korean Women. Journal of Investigative Dermatology 142:1077–1084. Available from: https://www.sciencedirect.com/science/article/pii/S0022202X21023034.

Setter D, Mousset S, Cheng X, Nielsen R, DeGiorgio M, Hermisson J. 2020. VolcanoFinder: Genomic scans for adaptive introgression. Available from: http://dx.doi.org/10.1371/journal.pgen.1008867.

Shore AM, Karamitri A, Kemp P, Speakman JR, Graham NS, Lomax MA. 2013. Cold-Induced Changes in Gene Expression in Brown Adipose Tissue, White Adipose Tissue and Liver. PLoS One 8.

Siegler BH, Altvater M, Thon JN, Neuhaus C, Arens C, Uhle F, Lichtenstern C, Weigand MA, Weiterer S. 2021. Postoperative abdominal sepsis induces selective and persistent changes in CTCF binding within the MHC-II region of human monocytes. PLoS One 16.

Silvie O, Greco C, Franetich JF, Dubart-Kupperschmitt A, Hannoun L, van Gemert GJ, Sauerwein RW, Levy S, Boucheix C, Rubinstein E, et al. 2006. Expression of human CD81 differently affects host cell susceptibility to malaria sporozoites depending on the Plasmodium species. Cell Microbiol 8:1134–1146.

Silvie O, Rubinstein E, Franetich JF, Prenant M, Belnoue E, Rénia L, Hannoun L, Elings W, Levy S, Boucheix C, et al. 2003. Hepatocyte CD81 is required for Plasmodium falciparum and Plasmodium yoelii sporozoite infectivity. Nat Med 9:93–96.

Simmons KJ, Nde PN, Kleshchenko YY, Lima MF, Villalta F. 2006. Stable RNA interference of host thrombospondin-1 blocks Trypanosoma cruzi infection. FEBS Lett 580:2365–2370.

Simon-Assmann P, Orend G, Mammadova-Bach E, Spenlé C, Lefebvre O. 2011. Role of laminins in physiological and pathological angiogenesis. International Journal of Developmental Biology 55:455–465.

Sinha S, Qidwai T, Kanchan K, Anand P, Jha GN, Pati SS, Mohanty S, Mishra SK, Tyagi PK, Sharma SK, et al. 2008. Variations in host genes encoding adhesion molecules and susceptibility to falciparum malaria in India. Malar J 7.

Sitras V, Paulssen RH, Grønaas H, Leirvik J, Hanssen TA, Vårtun Å, Acharya G. 2009. Differential Placental Gene Expression in Severe Preeclampsia. Placenta 30:424–433. Available from: https://www.sciencedirect.com/science/article/pii/S0143400409000332.

Smolinska MJ, Page TH, Urbaniak AM, Mutch BE, Horwood NJ. 2011. Hck Tyrosine Kinase Regulates TLR4-Induced TNF and IL-6 Production via AP-1. The Journal of Immunology 187:6043–6051.

Sohail M, Palma-Martínez MJ, Chong AY, Quinto-Cortés CD, Barberena-Jonas C, Medina-Muñoz SG, Ragsdale A, Delgado-Sánchez G, Cruz-Hervert LP, Ferreyra-Reyes L, et al. 2023. Mexican Biobank advances population and medical genomics of diverse ancestries. Nature 622:775–783.

Speakman JR. 2008. Thrifty genes for obesity, an attractive but flawed idea, and an alternative perspective: The “drifty gene” hypothesis. Int J Obes 32:1611–1617.

Stern DL. 2013. The genetic causes of convergent evolution. Nat Rev Genet 14:751–764.

Stringer C. 2016. The origin and evolution of homo sapiens. Philosophical Transactions of the Royal Society B: Biological Sciences 371.

Stringer C, Galway-Witham J. 2017. On the origin of our species. Nature 546:212–214. Available from: https://doi.org/10.1038/546212a.

Sun XY, Hayashi Y, Xu S, Kanou Y, Takagishi Y, Tang YP, Murata Y. 2011. Inactivation of the Rcan2 Gene in mice ameliorates the age- and diet-induced obesity by causing a reduction in food intake. PLoS One 6.

Tajima F. 1983. EVOLUTIONARY RELATIONSHIP OF DNA SEQUENCES IN FINITE POPULATIONS. Genetics 105:437–460. Available from: https://doi.org/10.1093/genetics/105.2.437.

Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. 2019. Benefits and limitations of genome-wide association studies. Nat Rev Genet 20:467–484.

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, et al. 2007. Beringian standstill and spread of native American founders. PLoS One 2.

Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet 49:303–309.

Tian Y, Peng B, Fu X. 2018. New ADCY3 Variants Dance in Obesity Etiology. Trends in Endocrinology and Metabolism 29:361–363.

Tong J, Niu Y, Chen Z-J, Zhang C. 2020. Comparison of the transcriptional profile in the decidua of early-onset and late-onset pre-eclampsia. Journal of Obstetrics and Gynaecology Research 46:1055–1066. Available from: https://doi.org/10.1111/jog.14257.

Tong X, Chen L, Liu S, Yan Z, Peng S, Zhang Y, Fan H. 2015. Polymorphisms in HLA-DRB1 Gene and the Risk of Tuberculosis: A Meta-analysis of 31 Studies. Lung 193:309–318.

Tougan T, Edula JR, Morita M, Takashima E, Honma H, Tsuboi T, Horii T. 2020. The malaria parasite Plasmodium falciparum in red blood cells selectively takes up serum proteins that affect host pathogenicity. Malar J 19.

Tougan T, Edula JR, Takashima E, Morita M, Shinohara M, Shinohara A, Tsuboi T, Horii T. 2018. Molecular Camouflage of Plasmodium falciparum Merozoites by Binding of Host Vitronectin to P47 Fragment of SERA5. Sci Rep 8.

Tronchere H, Cinato M, Timotin A, Guitou L, Villedieu C, Thibault H, Baetz D, Payrastre B, Valet P, Parini A, et al. 2017. Inhibition of PIKfyve prevents myocardial apoptosis and hypertrophy through activation of SIRT3 in obese mice. EMBO Mol Med 9:770–785.

Tsibulnikov S, Maslov L, Voronkov N, Oeltgen P. 2000. Thyroid hormones and the mechanisms of adaptation to cold. Available from: https://doi.org/10.1007/s42000-020-00200-2.

Tsipis CP, Sun X, Xu K, LaManna JC. 2014. Hypoxia-Induced Angiogenesis and Capillary Density Determination. In: Milner R, editor. Cerebral Angiogenesis: Methods and Protocols. New York, NY: Springer New York. p. 69–80. Available from: https://doi.org/10.1007/978-1-4939-0320-7_6.

Váazquez J, Franco E, Reyes G, Meza I. 1995. Characterization of adhesion plates induced by the interaction of Entamoeba histolytica trophozoites with fibronectin. Cell Motil 32:37–45. Available from: https://doi.org/10.1002/cm.970320105.

Vallerand AL, Perusse F, Bukowiecki LJ. Cold exposure potentiates the effect of insulin on in vivo glucose uptake.

Vichaiwong K, Purohit S, An D, Toyoda T, Jessen N, Hirshman MF, Goodyear LJ. 2010. Contraction regulates site-specific phosphorylation of TBC1D1 in skeletal muscle. Biochemical Journal 431:311–320.

De Villartay JP, Lim A, Al-Mousa H, Dupont S, Déchanet-Merville J, Coumau-Gatbois E, Gougeon ML, Lemainque A, Eidenschenk C, Jouanguy E, et al. 2005. A novel immunodeficiency associated with hypomorphic RAG1 mutations and CMV infection. Journal of Clinical Investigation 115:3291–3299.

Visscher PM, Yengo L, Cox NJ, Wray NR. 2021. Discovery and implications of polygenicity of common diseases. Science (1979) 373:1468–1473. Available from: https://doi.org/10.1126/science.abi8206.

Vollstedt S, Franchini M, Hefti HP, Odermatt B, O'Keeffe M, Alber G, Glanzmann B, Riesen M, Ackermann M, Suter M. 2003. Flt3 ligand-treated neonatal mice have increased innate immunity against intracellular pathogens and efficiently control virus infections. Journal of Experimental Medicine 197:575–584.

Wang Q, Zhou J, Zhang B, Tian Z, Tang J, Zheng Y, Huang Z, Tian Y, Jia Z, Tang Y, et al. 2013. Hepatitis B Virus Induces IL-23 Production in Antigen Presenting Cells and Causes Liver Damage via the IL-23/IL-17 Axis. PLoS Pathog 9.

Wang R, Liu W, Zhao J, Liu L, Li S, Duan Y, Huo Y. 2023. Overexpressed LAMC2 promotes trophoblast over-invasion through the PI3K/Akt/MMP2/9 pathway in placenta accreta spectrum. Journal of Obstetrics and Gynaecology Research 49:548–559. Available from: https://doi.org/10.1111/jog.15493.

Wang WYS, Barratt BJ, Clayton DG, Todd JA. 2005. Genome-wide association studies: Theoretical and practical concerns. Nat Rev Genet 6:109–118.

Wang X, Wahl R. 2014. Responses of the insulin signaling pathways in the brown adipose tissue of rats following cold exposure. PLoS One 9.

Weiss DJ, Dzianach PA, Saddler A, Lubinda J, Browne A, McPhail M, Rumisha SF, Sanna F, Gelaw Y, Kiss JB, et al. 2025. Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum and Plasmodium vivax malaria, 2000–22: a spatial and temporal modelling study. The Lancet 405:979–990.

Willerslev E, Meltzer DJ. 2021. Peopling of the Americas as inferred from ancient genomics. Nature 594:356–364.

Xu G, Luo Y, Wu W, Liu X, Yu X, Bao Y, He X, Yu J, Li Y, Yang J, et al. 2022. The Evolution of Acquired Resistance to BRAFV600E kinase inhibitor Is Sustained by IGF1-Driven Tumor Vascular Remodeling. Journal of Investigative Dermatology 142:445–458.

Xu S, Li S, Yang Y, Tan J, Lou H, Jin W, Yang L, Pan X, Wang J, Shen Y, et al. 2011. A genome-wide search for signals of high-altitude adaptation in tibetans. Mol Biol Evol 28:1003–1011.

Xu Y, Liu Y, Pan S, Liu L, Liu J, Zhai X, Shen H, Hu Z. 2013. IL-23R polymorphisms, HBV infection, and risk of hepatocellular carcinoma in a high-risk Chinese population. J Gastroenterol 48:125–131.

Xu Z, You W, Zhou Y, Chen W, Wang Y, Shan T. 2019. Cold-induced lipid dynamics and transcriptional programs in white adipose tissue. BMC Biol 17.

Xue Y, Daly A, Yngvadottir B, Liu M, Coop G, Kim Y, Sabeti P, Chen Y, Stalker J, Huckle E, et al. 2006. Spread of an Inactive Form of Caspase-12 in Humans Is Due to Recent Positive Selection. Available from: www.ajhg.org.

Yang J, Jin ZB, Chen J, Huang XF, Li XM, Liang YB, Mao JY, Chen X, Zheng Z, Bakshi A, et al. 2017. Genetic signatures of high-altitude adaptation in Tibetans. Proc Natl Acad Sci U S A 114:4189–4194.

Yang Z, Shi H, Ma P, Zhao S, Kong Q, Bian T, Gong C, Zhao Q, Liu Y, Qi X, et al. 2018. Darwinian Positive Selection on the Pleiotropic Effects of KITLG Explain Skin Pigmentation and Winter Temperature Adaptation in Eurasians. Mol Biol Evol 35:2272–2283. Available from: https://doi.org/10.1093/molbev/msy136.

Yau WW, Yen PM. 2020. Thermogenesis in adipose tissue activated by thyroid hormone. Int J Mol Sci 21.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329:75–78.

Yu L, Zhang Y, Li W, Mao J, Li Y, Wang H, Li C, Yang L, He W, Jia Yanjun, et al. 2024. Fluoxetine Successfully Treats Intracranial Enterovirus E18 Infection in a Patient with CD79a Deficiency Arising from Segmental Uniparental Disomy of Chromosome 19. J Clin Immunol 44.

Yuan X, Zhang J, Ma TT, Zhuang RJ, Lei BB, Wang L, Cheng BF, Wang M, Yang HJ. 2021. Expression regulation of cold-inducible protein RBM3 by FAK/Src signaling for neuroprotection against rotenone under mild hypothermia. Biochem Biophys Res Commun 534:240–247.

Zekri Y, Flamant F, Gauthier K. 2021. Central vs. Peripheral action of thyroid hormone in adaptive thermogenesis: A burning topic. Cells 10.

Zhang B li, Xu R liang, Zhang Jing, Zhao X xian, Wu H, Ma L ping, Hu J qiang, Zhang Jian liang, Ye Z, Zheng X, et al. 2013. Identification and functional analysis of a novel PRKAG2 mutation responsible for Chinese PRKAG2 cardiac syndrome reveal an important role of non-CBS domains in regulating the AMPK pathway. J Cardiol 62:241–248.

Zhang X, Kim B, Singh A, Sankararaman S, Durvasula A, Lohmueller KE. 2023. MaLAdapt Reveals Novel Targets of Adaptive Introgression From Neanderthals and Denisovans in Worldwide Human Populations. Mol Biol Evol 40.

Zhang X, Witt KE, Bañuelos MM, Ko A, Yuan K, Xu S, Nielsen R, Huerta-Sanchez E. 2021. The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans. Proc Natl Acad Sci U S A 118:1–9.

Zhao Wentao, Ouyang C, Zhang L, Wang J, Zhang J, Zhang Y, Huang C, Xiao Q, Jiang B, Lin F, et al. 2024. The proto-oncogene tyrosine kinase c-SRC facilitates glioblastoma progression by remodeling fatty acid synthesis. Nature Communications 15.

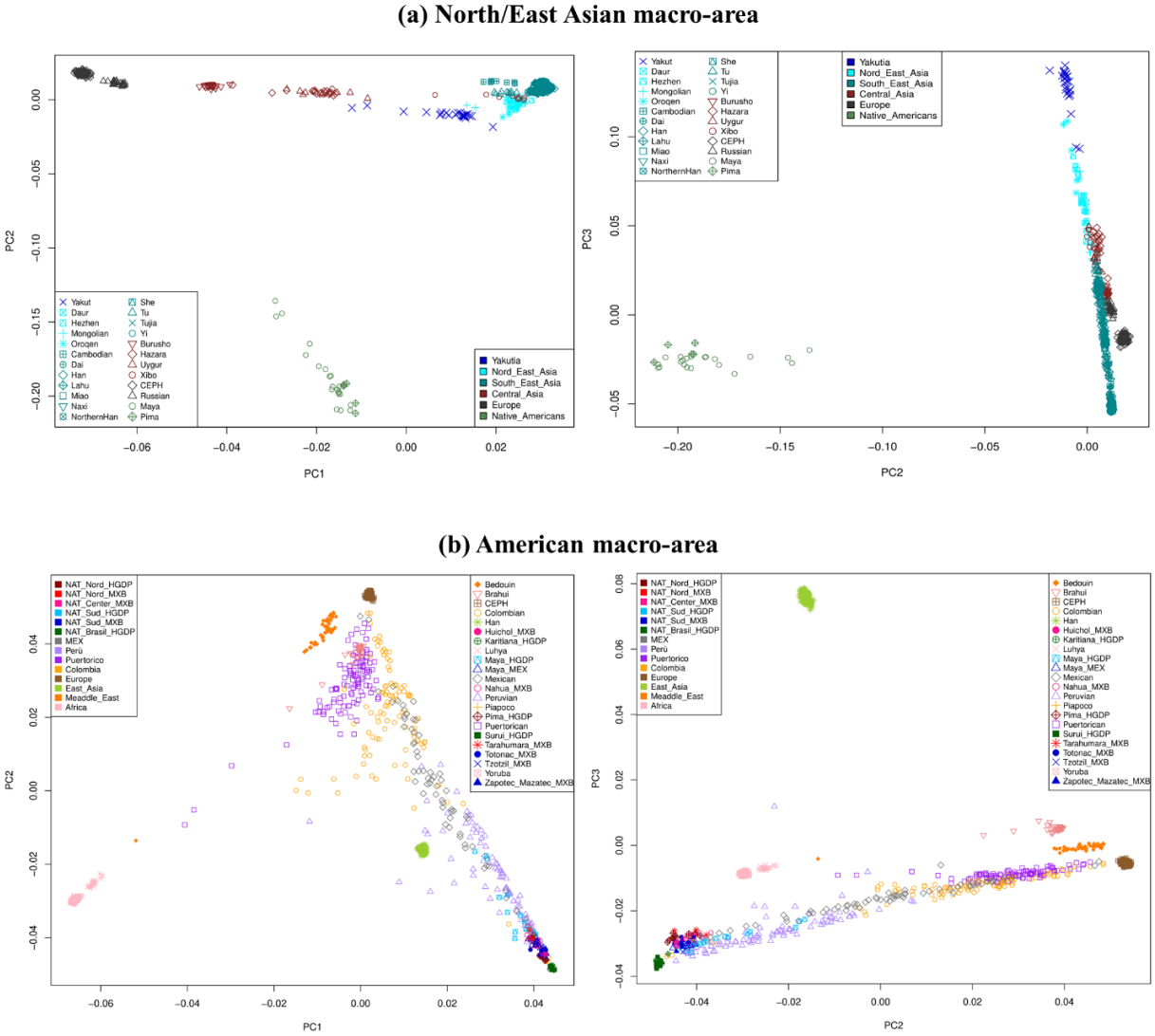
Zheng W, He Y, Guo Y, Yue T, Zhang H, Li J, Zhou B, Zeng X, Li L, Wang B, et al. 2023. Large-scale genome sequencing redefines the genetic footprints of high-altitude adaptation in Tibetans. Genome Biol 24.

Zou L, Yan S, Guan X, Pan Y, Qu X. 2013. Hypermethylation of the PRKCZ gene in type 2 diabetes mellitus. J Diabetes Res 2013.

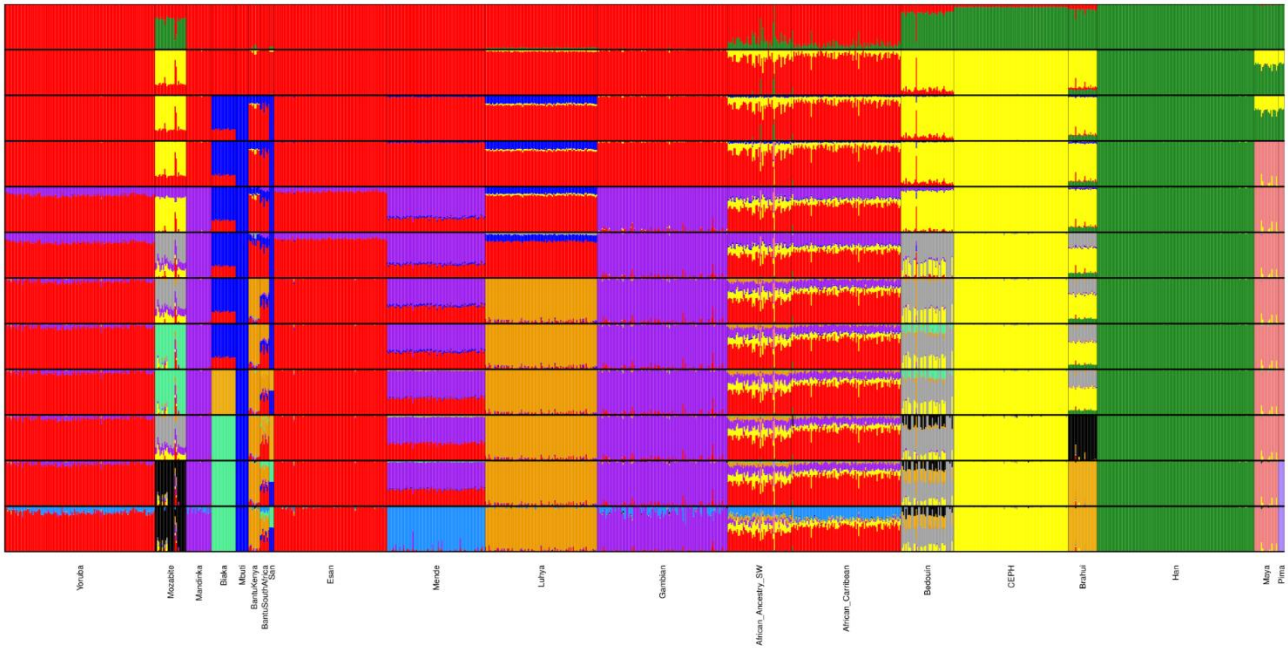
8. ACKNOWLEDGMENTS

We would like to thank John Lindo for having kindly shared Aymara WGS data, Choongwon Jeong and Anna Di Rienzo for having kindly shared Tibetan WGS data and Andrés Moreno-Estrada for having kindly shared WGS data of Native Indigenous human populations from Mexico. Without these collaborations and the subsequent analyses of the shared data, several key aspects of the present thesis would not have been properly investigated.

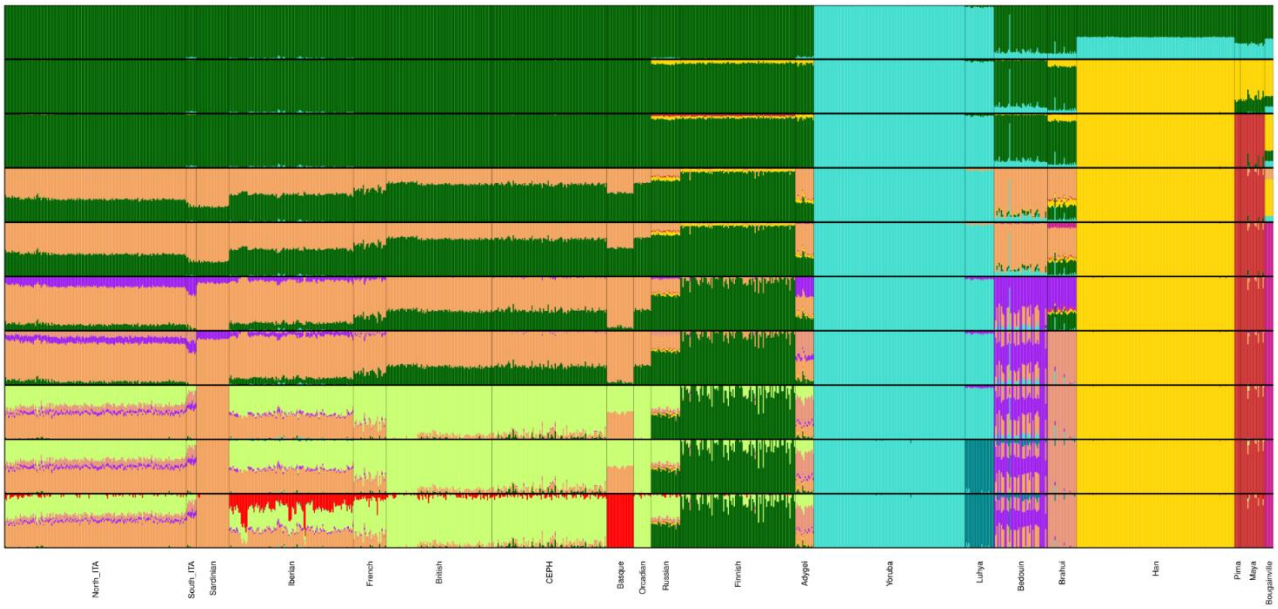
9. APPENDIX



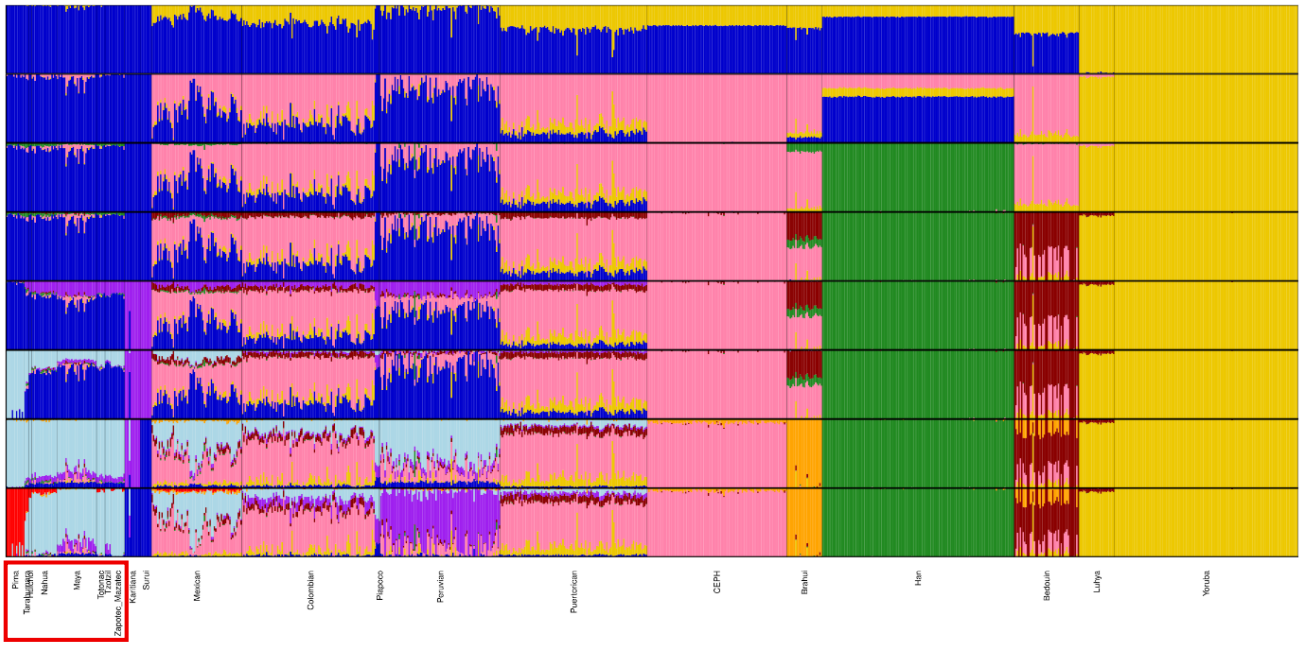
Supplementary Figure S1. PCA plots obtained for (a) North/East Asian and (b) America macro-areas. For all the plots in the figure, from left to right, are displayed the intersections between PC1 vs PC2 and PC2 vs PC3. **(a)** PC1 describes the gradient of variation observable between Asian populations in the dataset; PC2 instead explains how Native American groups differentiate with respect to Asian populations; PC3 enables to appreciate the outlier position occupied by Yakut individuals with respect to most groups of European and Asian ancestry. The low resolution of the PCA analysis do not allow the identification of well delineated East Asian clusters thus requiring further structure analysis (such as Chromopainter and fineSTRUCTURE) aimed at further elucidating the haplotype affinity between these human groups. **(b)** PC1 describes the genetic differences between African and non-African populations. PC2 displays a gradient of reduced genomic variation among Europeans, admixed Americans (e.g. Colombian), East Asians, and indigenous populations from Central and South America. In the bottom left of PC1 vs PC2 plot, are localized the samples belonging to Native American populations (i.e. NAT). Except for five Mayan individuals plotted in the vicinity of the Peruvian samples, all the individuals belonging to Native American populations cluster in well-defined groups. Samples generated in the study by Jiménez-Kaufmann et al. 2022 are collectively indicated in the legend as NAT_MXB.



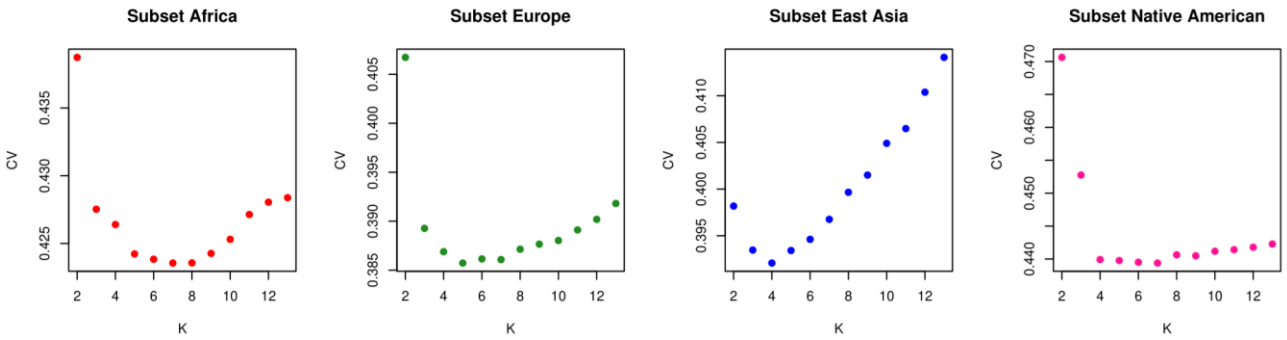
Supplementary Figure S2. ADMIXTURE profiles inferred for the individuals representative of the African macro-area. The lowest cross validation error was obtained for the configuration K=8 reported in the seventh row of the plot. The Yoruba and Esan Nigerian populations are characterized by elevated proportions of the red ancestral component which in the latter group reaches fixation. The West African Mende and Gambian populations present instead greater proportions of the purple ancestral component while the red one is still detectable at low frequencies. The Central/East African Luhya are characterized by great proportions of the orange component and very low frequency of the Yoruba- Esan-like one. The Bantus present similar patters to the Luhya although those from South Africa are characterized by more elevated proportions of the red ancestral component. The Mozabite population present distinctive ADMIXTURE profiles characterized by elevated frequency of the grey ancestral component and moderate frequency of both purple and yellow ancestral components, with the latter describing central Europeans. Such a pattern is analogously found in the Bedouin people. Overall, all the individuals reported in the plot present homogenous ADMIXTURE profiles reflecting their population of origin.



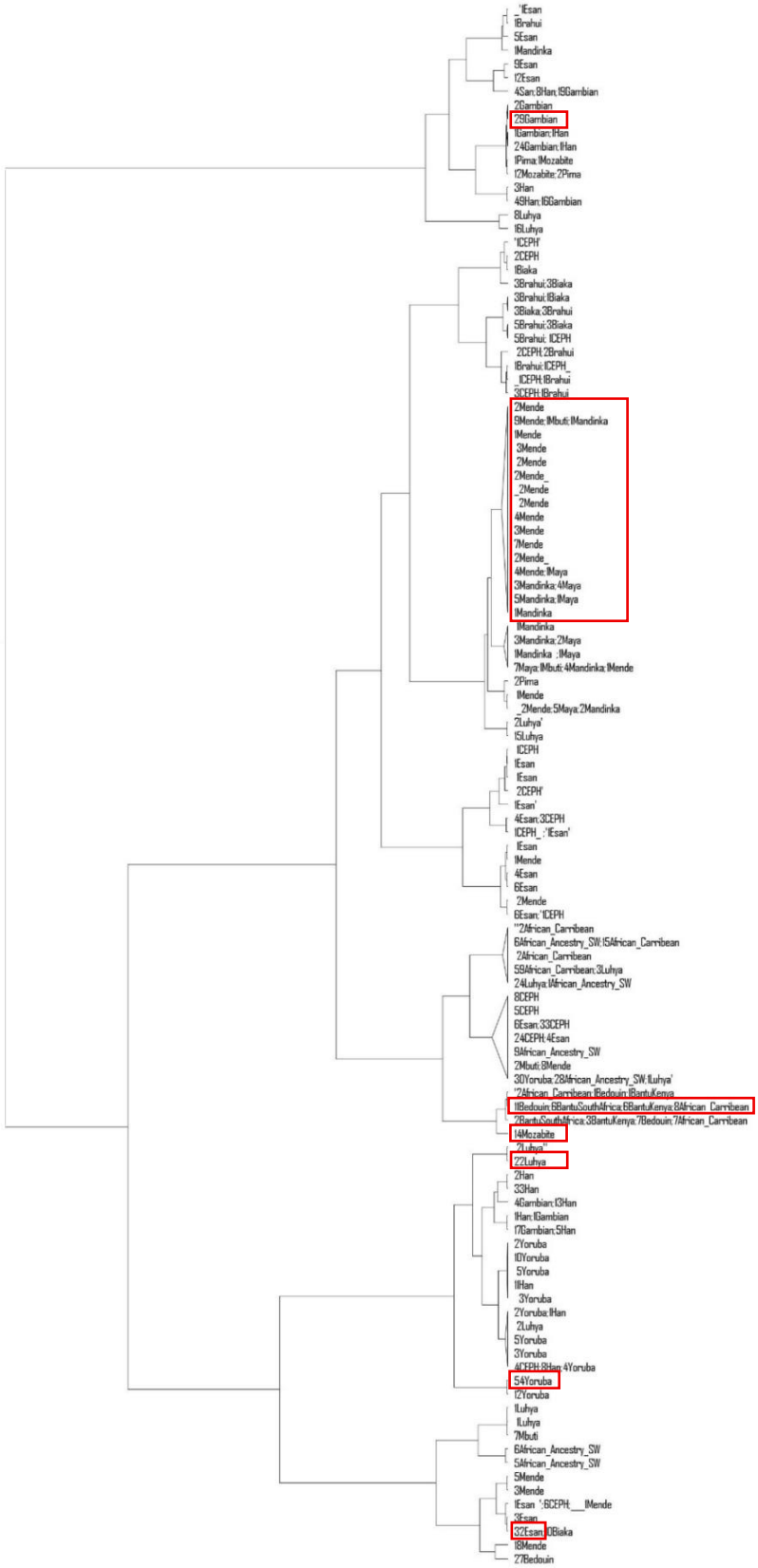
Supplementary Figure S3. ADMIXTURE profiles inferred for the individuals representative of the European macro-area. The configuration associated with the lower cross-validation error is that described by K=5 ancestral components, and is reported graphically in the fourth line of the figure. Such configuration illustrates a high degree of genetic homogeneity among European populations, which exhibit two shared ancestral components represented with pink and dark green colours. The Finnish population displays a distinctive pattern, with the dark green ancestral component significantly prevalent and nearly reaching fixation and also low proportions of the ancestral component characterizing the Han Chinese population (i.e. yellow). Similarly, the Sardinian population exhibits elevated proportions of the pink ancestral component respect to North Italians and to Central (i.e. CEPH) and West European populations (i.e. Iberian and Basque).



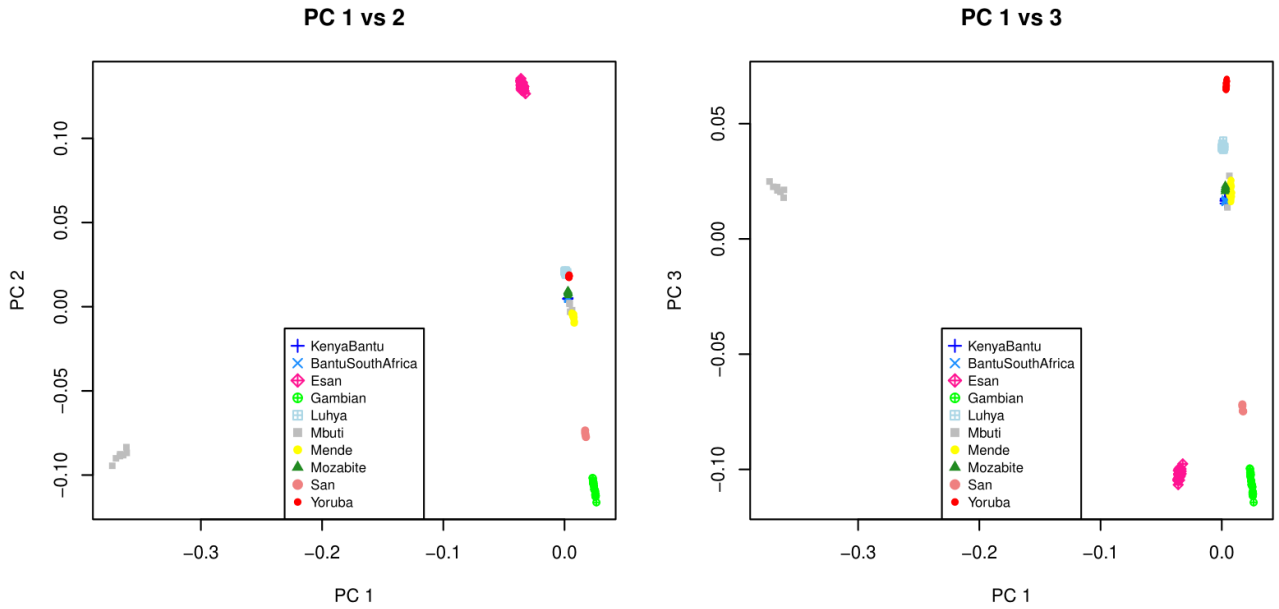
Supplementary Figure S4. ADMIXTURE profiles inferred for the individuals representative of the American macro-area. The ADMIXTURE analysis associated with the lower cross-validation error and thus describing the configuration that better fit the data is described by the configuration K=7 which is displayed in the sixth line of the figure. Admixed Americans (i.e. Mexican, Colombian, Peruvian and Puerto Rican populations) are characterized by consistent proportions of the ancestral component describing Central European individuals (i.e. CEPH, in pink). Indigenous populations from Mexico (i.e. red rectangle) are characterized by elevated proportions of the blue ancestral component, detectable at lower frequency in the admixed samples as well. Only the Pima population deviates from such a pattern, and specifically the individuals belonging to such human group are characterized by a fixed specific ancestral component (i.e. light-blue). Similarly, Surui and Karitiana samples present their own ancestral component describing Native populations from South America.



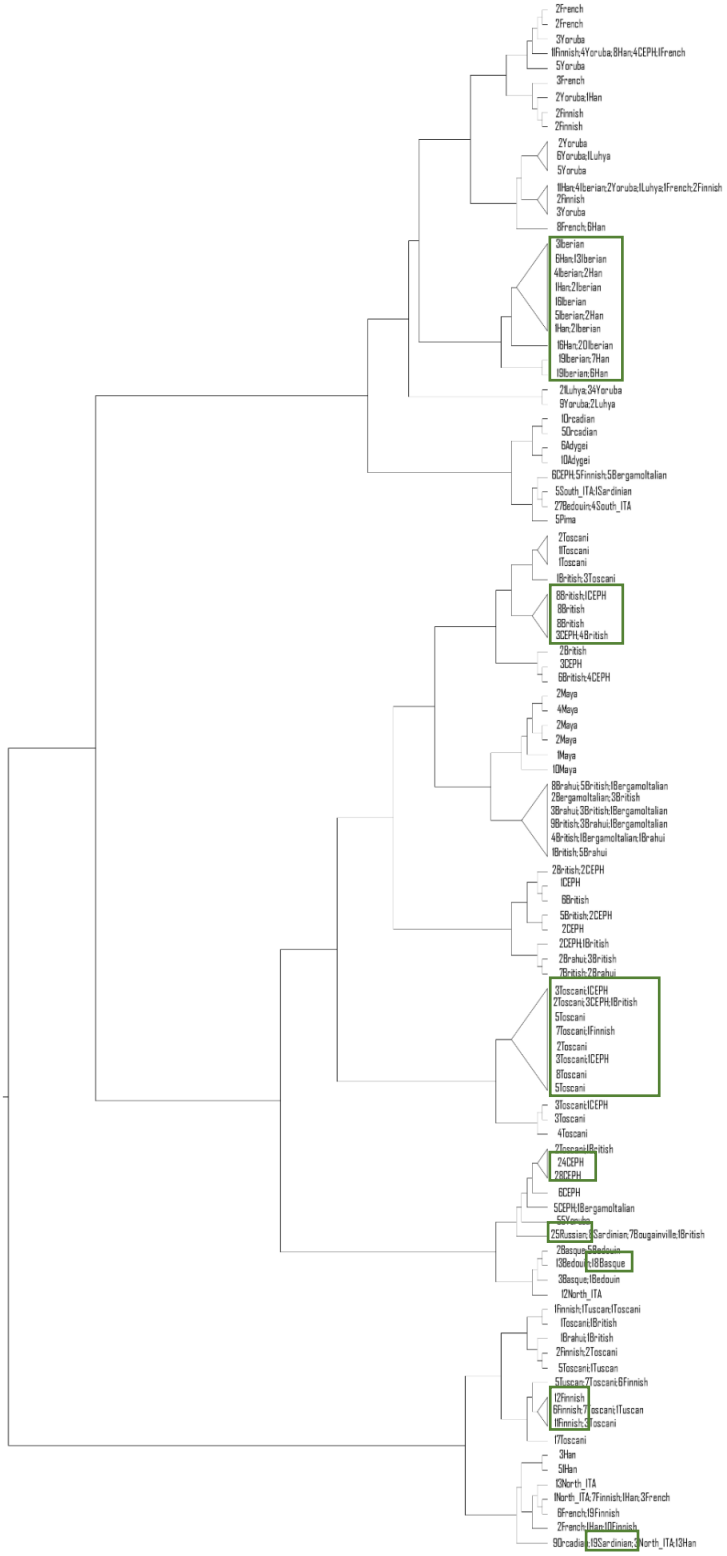
Supplementary Figure S5. Cross-Validation (CV) errors associated to each tested ADMIXTURE configuration. On the x axis are reported all the tested K configurations while in the y axis are displayed CV errors. For both the African (red) and American (pink) macro-areas the lowest CV was obtained for the configuration K=7, while for European (green) and North/East Asian (blue) macro-areas the lowest CV was that of K=5 and K=4, respectively.



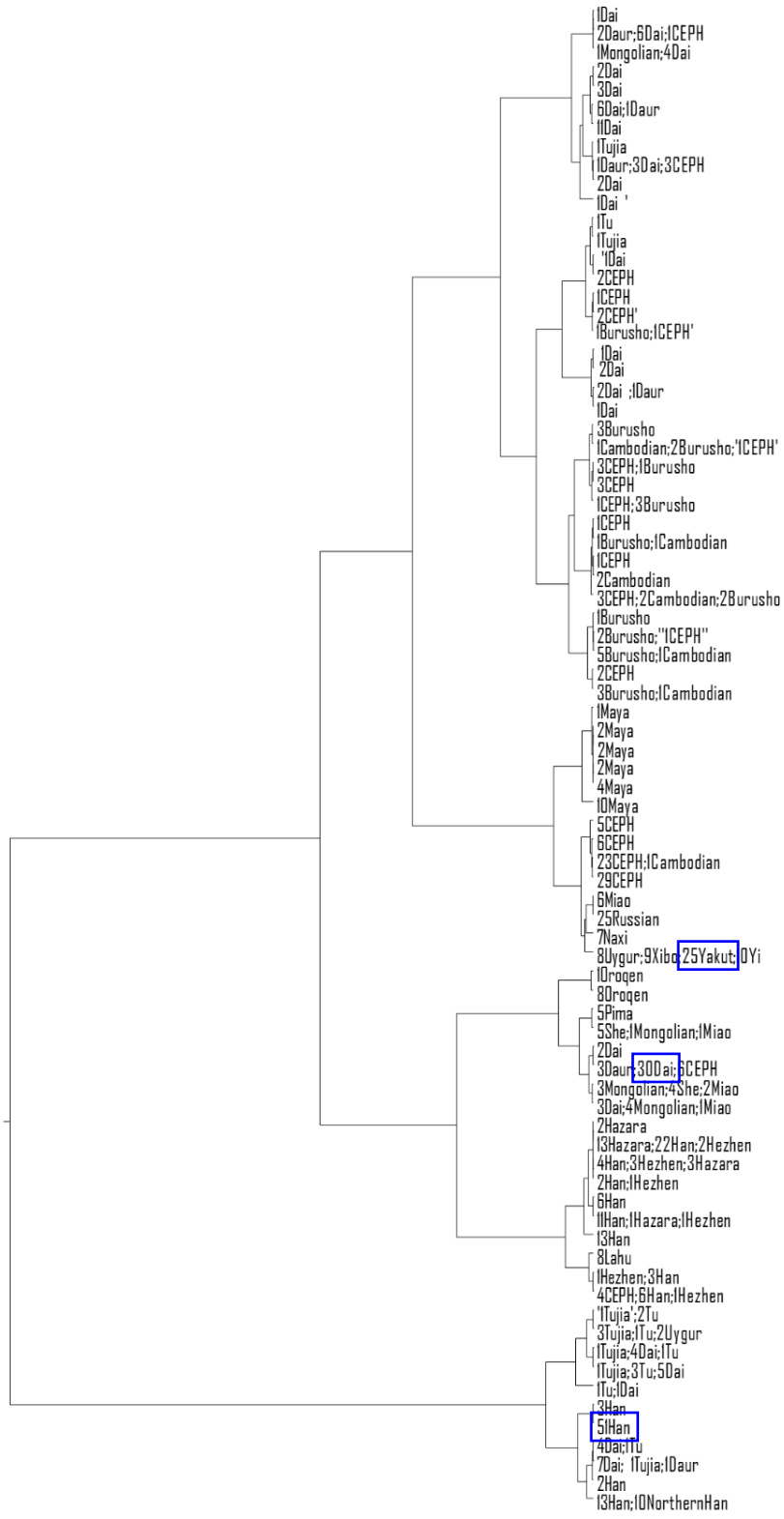
Supplementary Figure S6. Dendrogram resulting from ChomoPainter/fineSTRUCTURE analyses for African macro-area. Individuals are aggregated by population and all the nodes in the plot are characterized by a nodal support value ≥ 0.8. The final groups composed by genetically homogenous individuals of African ancestry and presenting suitable sample size to be subjected to selection analyses are underlined with red rectangles. From top to bottom are reported the Gambian, Mende, Bantu, Mozabite, Luhya, Yoruba and Esan filtered homogeneous populations.



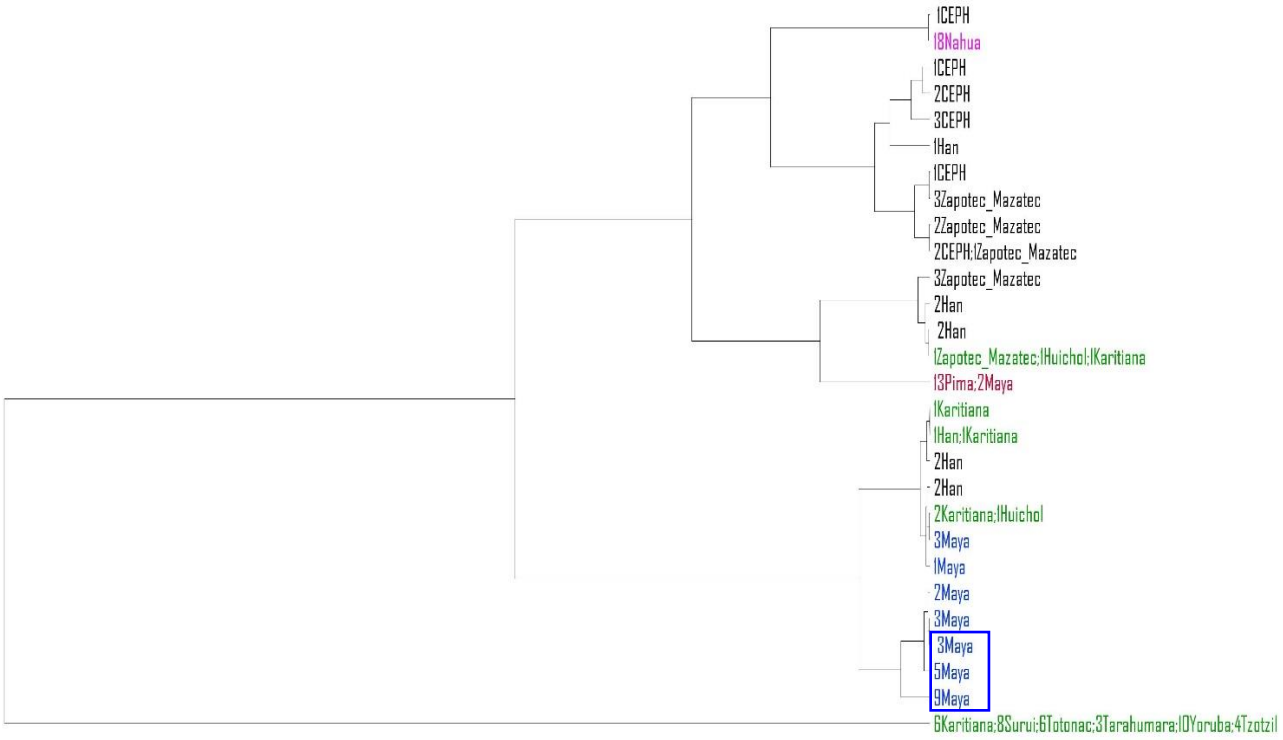
Supplementary Figure S7. PCA analysis performed evaluating the levels of haplotype chunks shared among the individuals representative of the African macro-area. PC1 visualizes the separation between Mbuti population and the rest of the African groups. PC2 describes the differences among Esan, Gambian and San individuals with respect to the macro cluster including Yoruba, Luhya, Mende, Mozabite, Bantu Kenya and Bantu South Africa populations. The individuals belonging to the latter two groups (i.e. blue cross and light-blue *x*) completely overlapped in both the plots thus supporting the unification of these samples into a unique cluster.



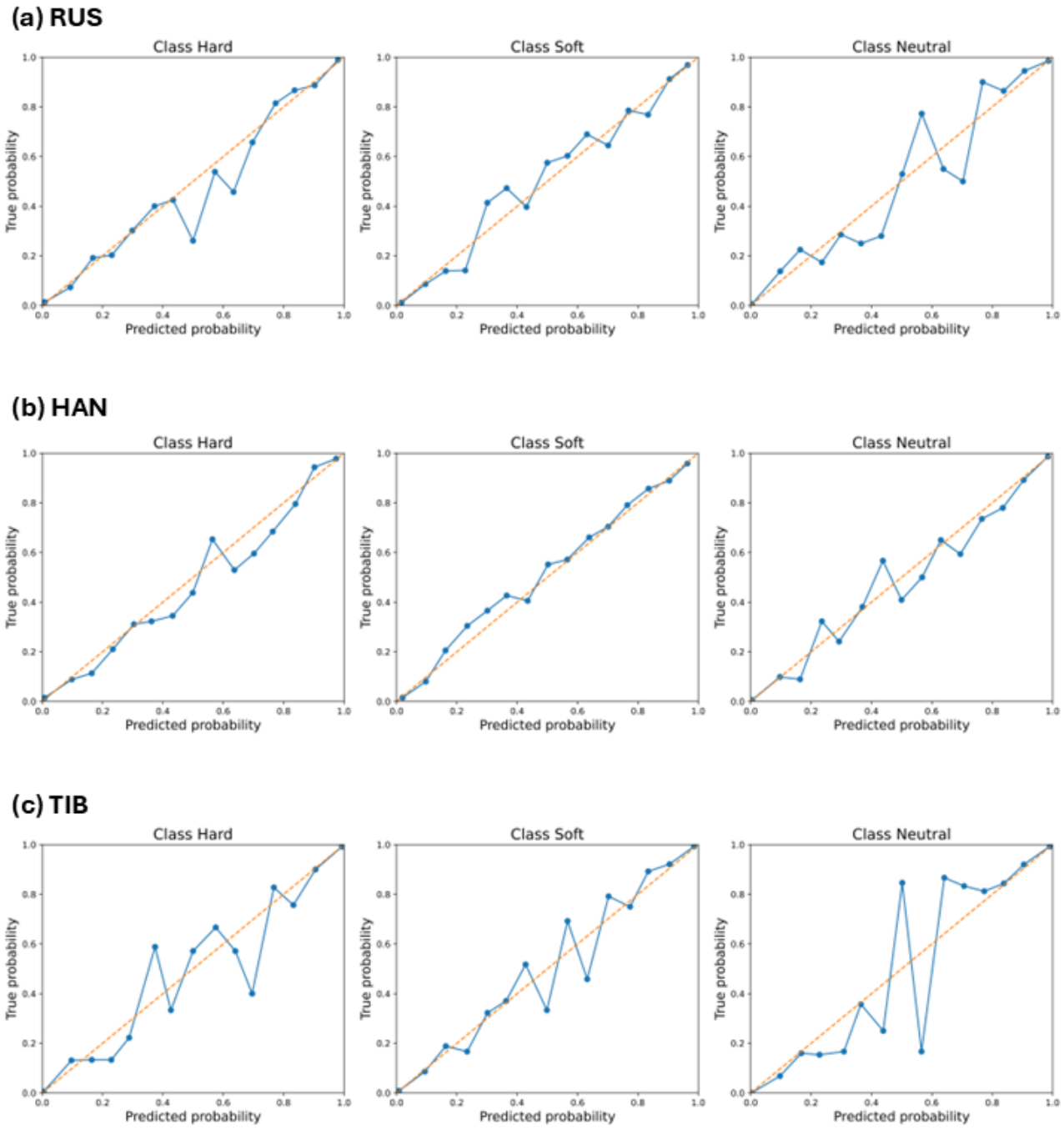
Supplementary Figure S8. Dendrogram resulting from ChomoPainter/fineSTRUCTURE analyses for European macro-area. Individuals are aggregated by population and all the nodes in the plot are characterized by a nodal support value ≥ 0.8. The final groups composed by genetically homogenous individuals of European ancestry and presenting suitable sample size to be subjected to further analyses are underlined with green rectangles. In order, from top to bottom are reported the Iberian, British, Tuscan, CEPH, Russian, Basque, Finnish and Sardinian filtered homogeneous populations. From the reported Iberian group were removed six individuals according to the PCA analysis displayed in Figure 4.5.



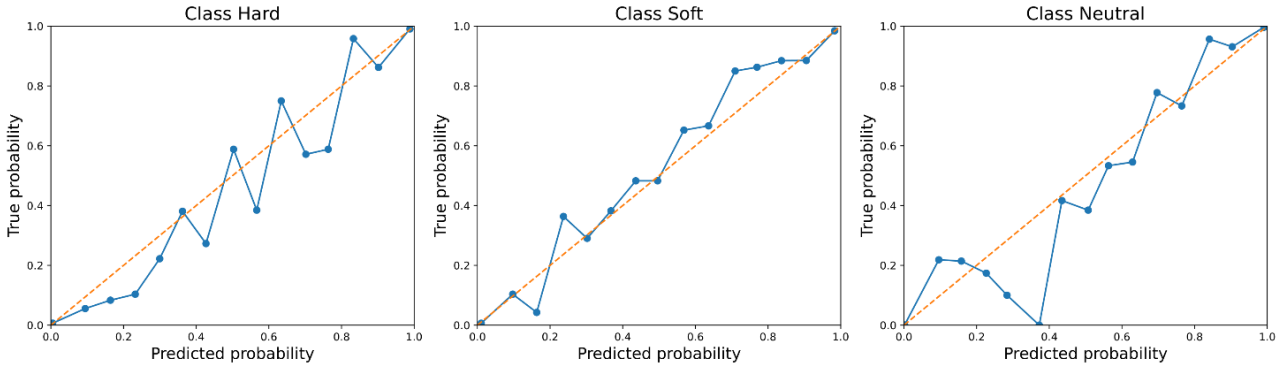
Supplementary Figure S9. Dendrogram resulting from ChomoPainter/fineSTRUCTURE analyses for North/East Asian macro-area. Individuals are aggregated by population and all the nodes in the plot are characterized by a nodal support value ≥ 0.8. The final groups composed by genetically homogenous individuals of North/East Asian ancestry and presenting suitable sample size to be subjected to further analyses are underlined with blue rectangles. From top to bottom are reported the Yakut, Han, and Dai Chinese filtered populations. Particularly, these latter two groups were further filtered by retaining homogeneous samples generated in the context of the same reference database (i.e. HGDP or 1000 Genomes) with the aim of avoiding the introduction of potential bias in the subsequent SMC++ inferences.



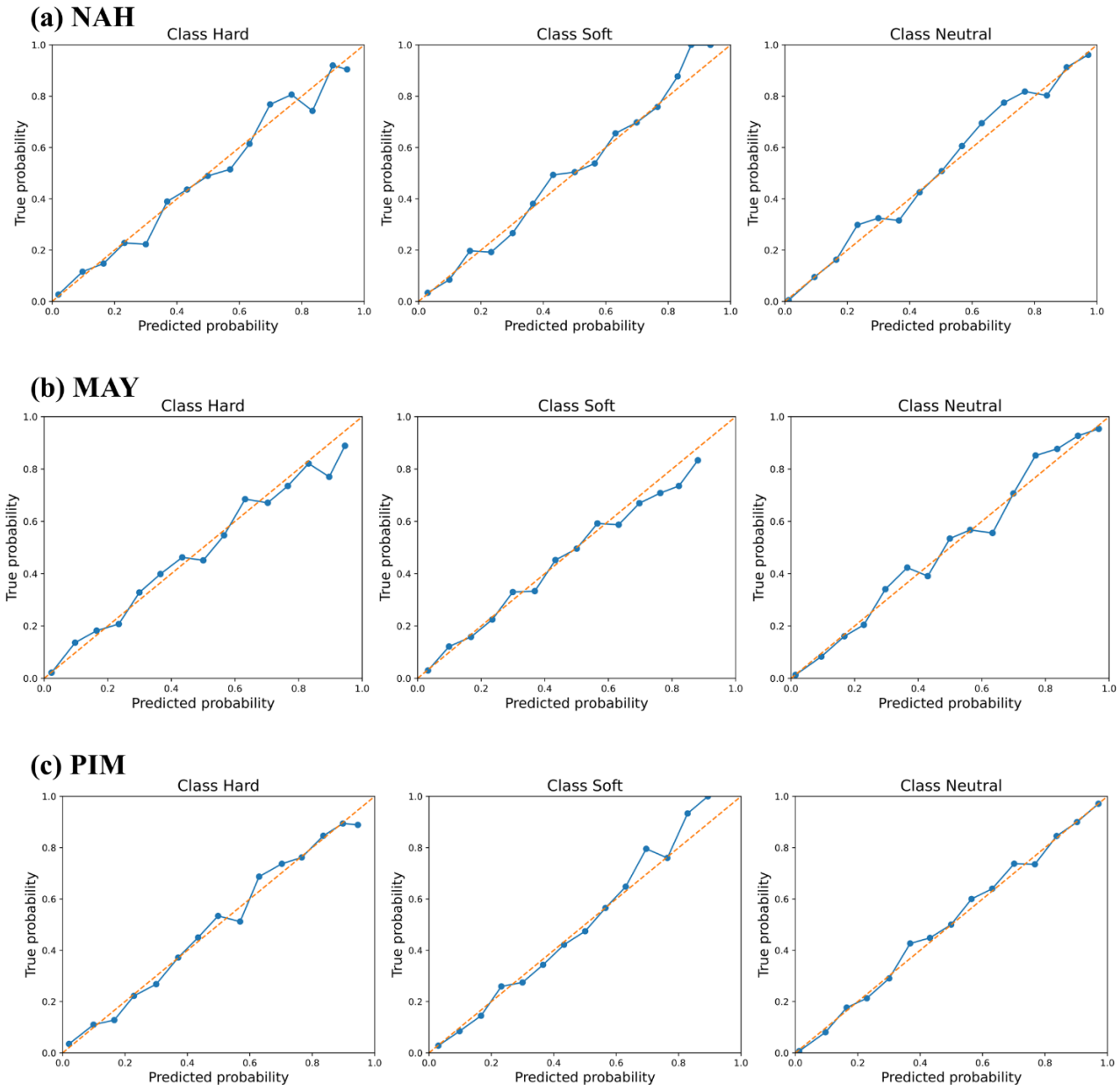
Supplementary Figure S10. Full dendrogram resulting from ChomoPainter/fineSTRUCTURE analyses for American macro-area. Individuals are aggregated by population and all the nodes in the plot are characterized by a nodal support value ≥ 0.8. The final groups composed by genetically homogenous individuals of Native American ancestry and presenting suitable sample size to be subjected to further analyses are: the Nahuas (pink), the Pima (red) and the Mayas (blue). The Maya cluster was further filtered by excluding the individuals presenting proportions of European-like ancestral component exceeding 10%, resulting in the final subset including 17 Mayas (i.e. blue rectangle). Finally, the green colour highlights clusters including Surui and Karitiana populations that however were not analysed in the present study due to their elevated inbreed levels.



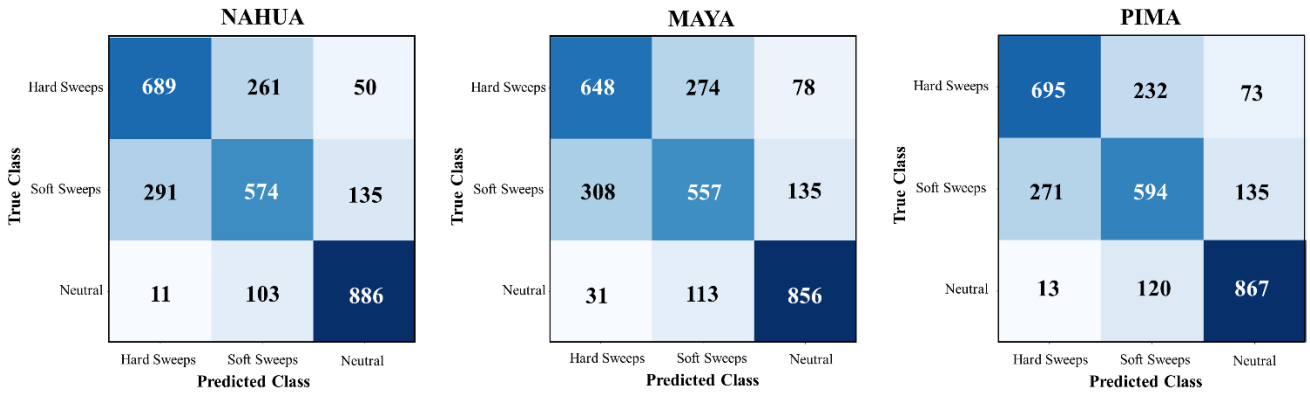
Supplementary Figure S11. Calibration plots obtained for Russian (a), Han Chinse (b) and high-altitude Tibetan (c) populations. Calibration/reliability curves of the *Trendsetter* inferred probabilities obtained after the application of the trained models to the validation set of simulations built for the Russian, Han Chinese and Tibetan individuals. The curves corresponding to each tested class (from left to right: hard sweeps, soft sweeps and neutral scenarios) well fit with the diagonal thus indicating that no further calibration was needed. The curves were built by dividing the values of the class frequency/predicted *Trendsetter* probabilities in 15 bins.



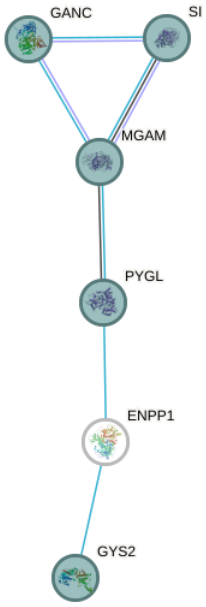
Supplementary Figure S12. Calibration plots obtained for the Yoruba population. Calibration/reliability curves of the *Trendsetter* inferred probabilities obtained after the application of the trained model to the validation set of simulations built for the Yoruba individuals. The curves corresponding to each tested class (from left to right: hard sweeps, soft sweeps and neutral scenarios) well fit with the diagonal with the sole exception of one frequency bin (centred at 0.4) of the neutral class. Taken together, the tends reported indicate that the *Trendsetter* output probabilities are calibrated thus not requiring further calibration. The curves were built by dividing the values of the class frequency/predicted *Trendsetter* probabilities in 15 bins.



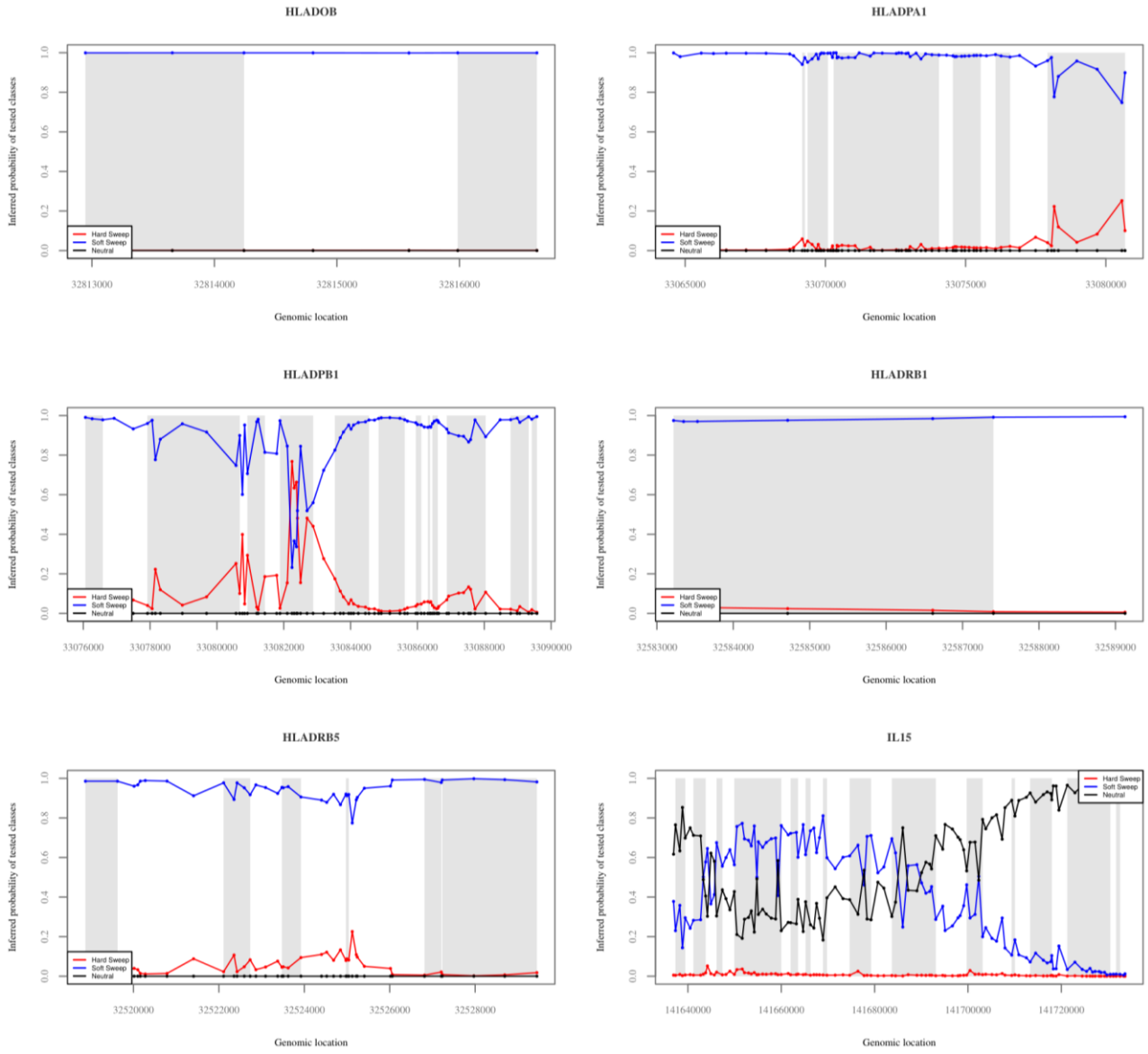
Supplementary Figure S13. Calibration plots obtained for Nahua (a), Maya (b) and Pima (c) populations. Calibration/reliability curves of the *Trendsetter* inferred probabilities obtained after the application of the trained models to the validation set of simulations built for the Nahua, Maya and Pima individuals. The curves corresponding to each tested class (from left to right: hard sweeps, soft sweeps and neutral scenarios) well fit with the diagonal thus indicating that no further calibration was needed. The curves were built by dividing the values of the class frequency/predicted *Trendsetter* probabilities in 15 bins.



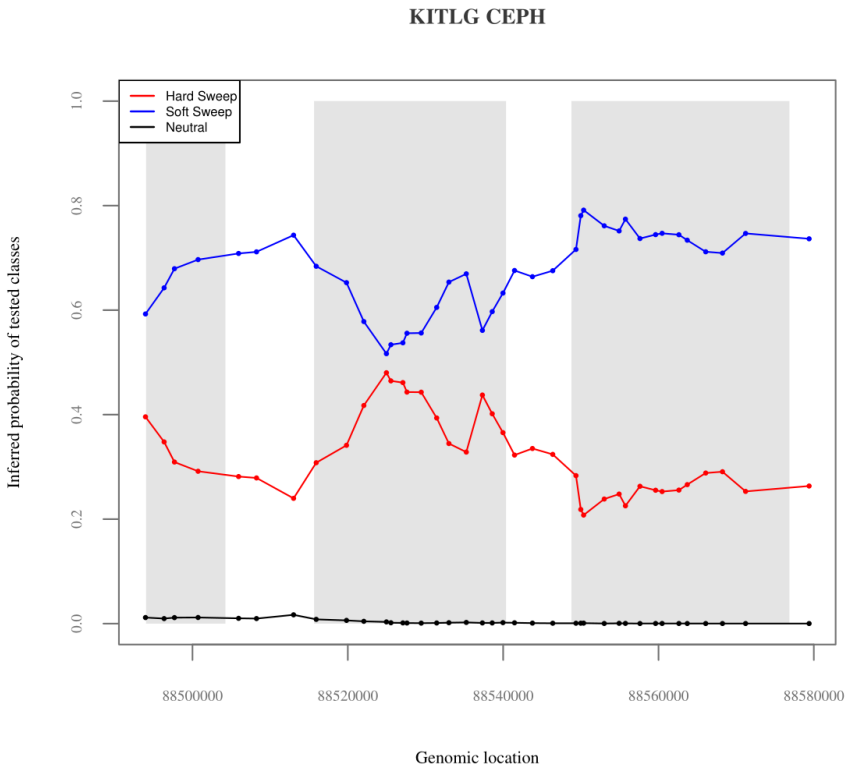
Supplementary Figure S14. Confusion matrices displaying the reliability of the trained *Trendsetter* classifiers obtained for indigenous populations from Mexico. Confusion matrices built relaying on results obtained after the application of the trained *Trendsetter* classifiers on the validation sets of genomic simulations. Along the diagonal are reported the correct classifications assigned by the algorithm, while in the remaining rows/columns are shown the misclassifications. From left to right the figure displays the confusion matrices resulted for Nahua, Maya and Pima human groups. Although these classifiers are less precise in distinguishing between hard and soft selective sweeps classes, they still recognize with a high precision the neutral genomic simulations from the adaptive scenarios.



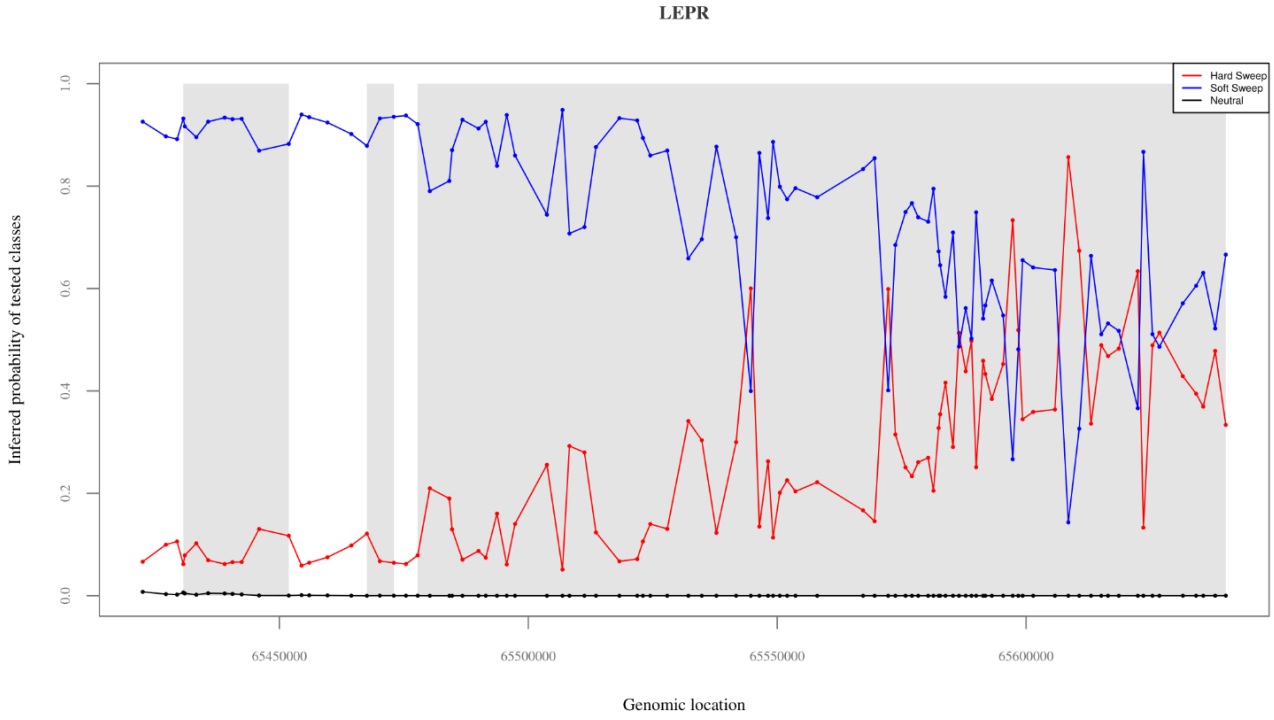
Supplementary Figure S15. Functional network targeted by natural selection in North African Mozabite and belonging to the *Starch and sucrose metabolism* pathway. The reported network was build using the STRING toll (available at https://string-db.org/) and by setting the confidence score to predict gene-gene associations (i.e. edges) at 0.9. The functional associations were inferred by considering co-expression (i.e. black edges), experimental (i.e. pink edges) and database (i.e. light-blue edges) evidence. Genes depicted as see-green circles are enriched in the *Cellular carbohydrate metabolic process* of the Gene Ontology database with a false discovery rate of 4.91e-08. *GANC*, *SI* and *MGAM* genes were also enriched in *Disaccharide metabolic process* annotated in the Gene Ontology database (FDR = 7.94e-08).



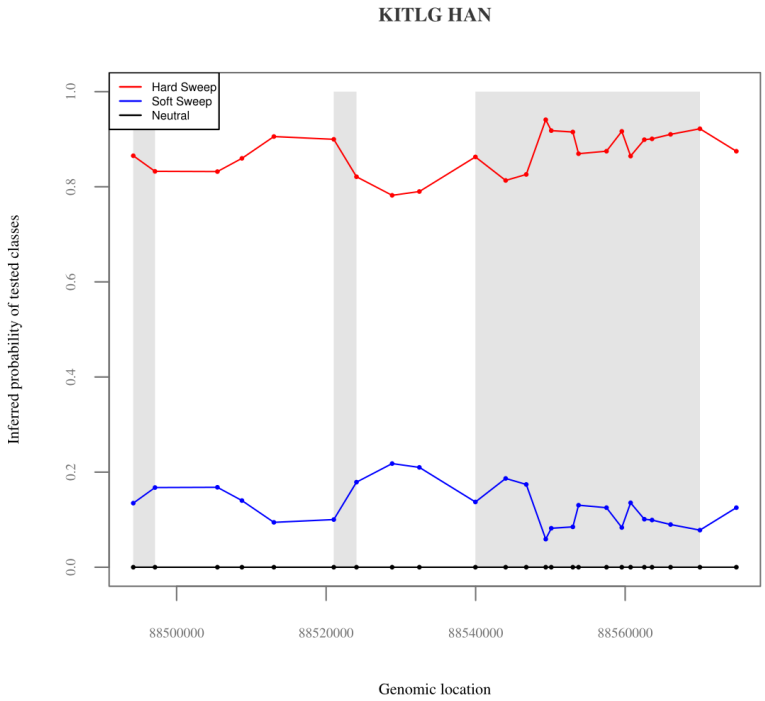
Supplementary Figure S16. *Trendsetter* predicted probabilities at HLA- and *IL15* genes in the Yoruba population. Distribution of the *Trendsetter* output probabilities of being classified as hard sweeps (i.e. red curve), soft sweeps (i.e. blue curve) and neutral regions (i.e. black curve) across the putative adaptive genes HLA- and *IL15*. For all the displayed plots, the x axis reports the positions of the central SNV in each genomic window identified according to the *Trendsetter* approach while on the y axis are shown the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent the genomic regions in the considered genes that present values indicative of the action of natural selection according to the *LASSI* statistic. Probabilities values associated to the soft selective sweeps class are remarkably elevated for these genes, reaching values almost equal to one at all HLA loci investigated. Furthermore, some of the windows falling in the *HLA-DPB1* gene are classified as hard selective sweeps, presenting probability values associated to this class reaching ~76%. The *IL15* gene includes multiple windows classified as soft selective sweeps, falling particularly in the first/middle portions of such a gene.



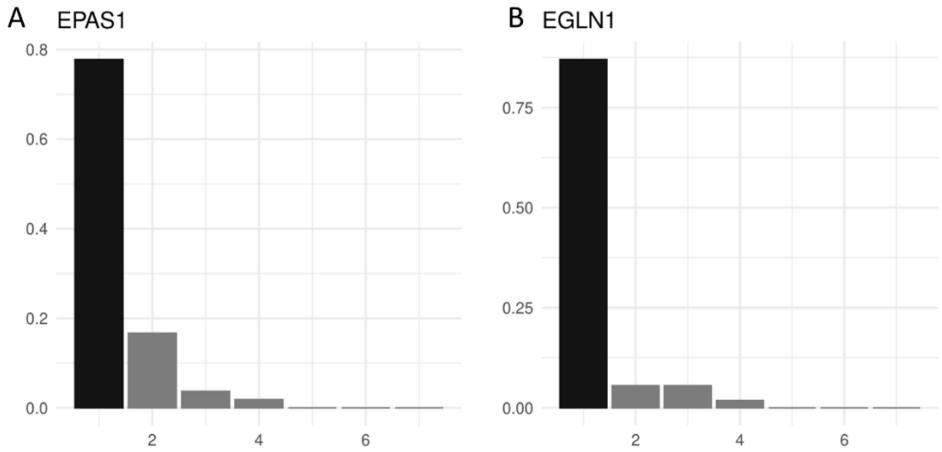
Supplementary Figure S17. *Trendsetter* Predicted probabilities at *KITLG* gene in the CEPH individuals. Distribution of *Trendsetter* output probabilities of being classified as hard sweeps (i.e. red curve), soft sweeps (i.e. blue curve) and neutral regions (i.e. black curve) across the putative adaptive *KITLG* gene. The x axis reports the positions of the central SNV in each genomic window identified according to the *Trendsetter* approach while on the y axis are shown the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent the genomic regions in the considered genes that present values indicative of the action of natural selection according to the *LASSI* statistic. Probabilities values associated to the soft selective sweep class are remarkably elevated in the entire gene locus, reaching values of ~79% in the middle/final portions of the gene. Furthermore, the predicted probabilities associated to the hard selective sweep class assume moderate values with an average of 32%. The neutral-associated probabilities are extremely low in all the genomic windows detected for the *KITLG* gene.



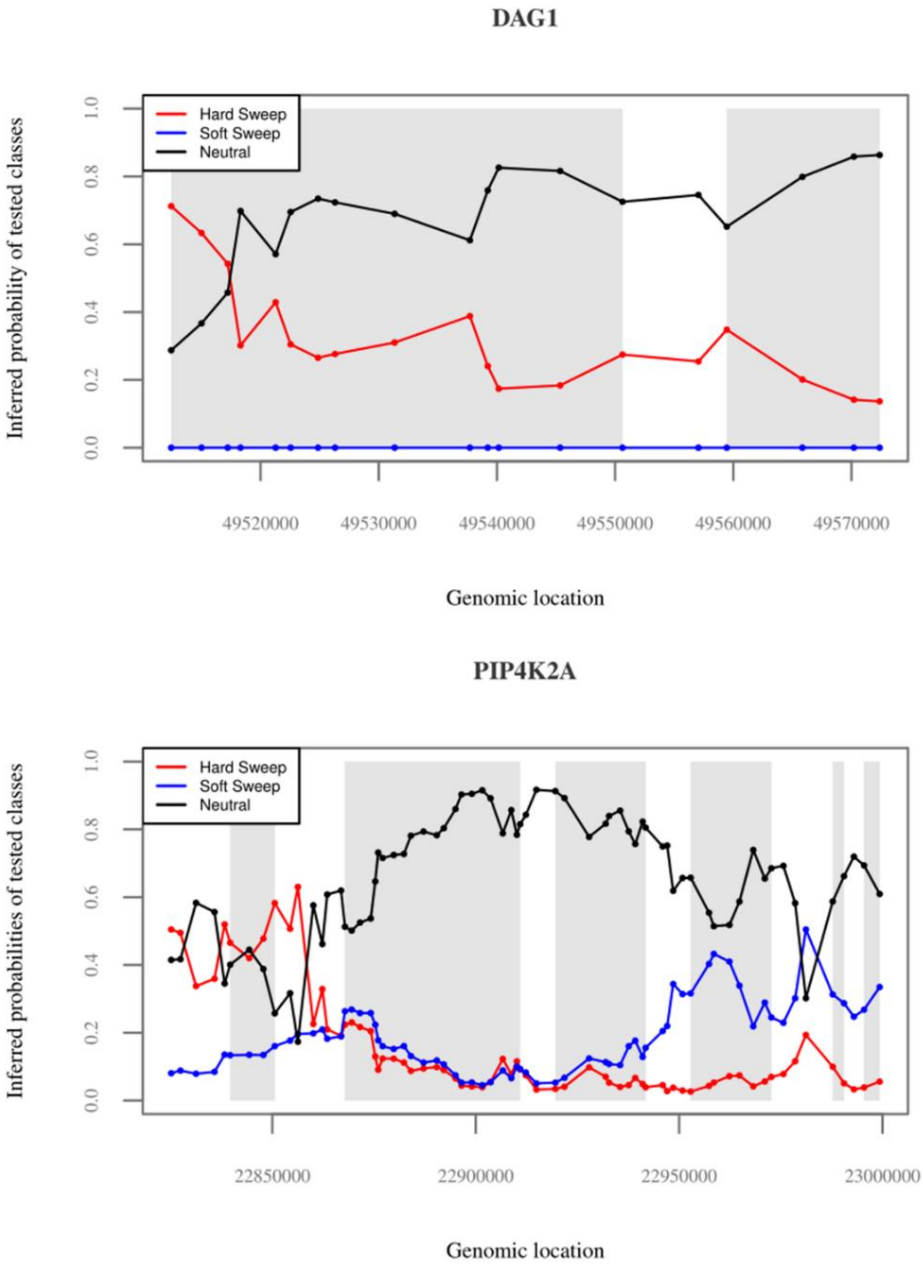
Supplementary Figure S18. *Trendsetter* predicted probabilities at *LEPR* gene in the Han Chinese population. Distribution of the *Trendsetter* output probabilities of being classified as hard sweeps (i.e. red curve), soft sweeps (i.e. blue curve) and neutral regions (i.e. black curve) across the putative adaptive *LEPR* gene. The x axis reports the positions of the central SNV in each genomic window identified according to the *Trendsetter* approach while on the y axis are shown the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent the genomic regions in the considered genes that present values indicative of the action of natural selection according to the *LASSI* statistic. The probabilities values associated to the soft selective sweep class are remarkably elevated in the first/middle portions of the gene, reaching values of ~95%. Furthermore, the predicted probability associated to the hard selective sweep class assume moderate to elevated values in the final part of the gene peaking at ~87%. The neutral-associated probabilities are extremely low in the all the genomic windows detected for the *LEPR* gene.



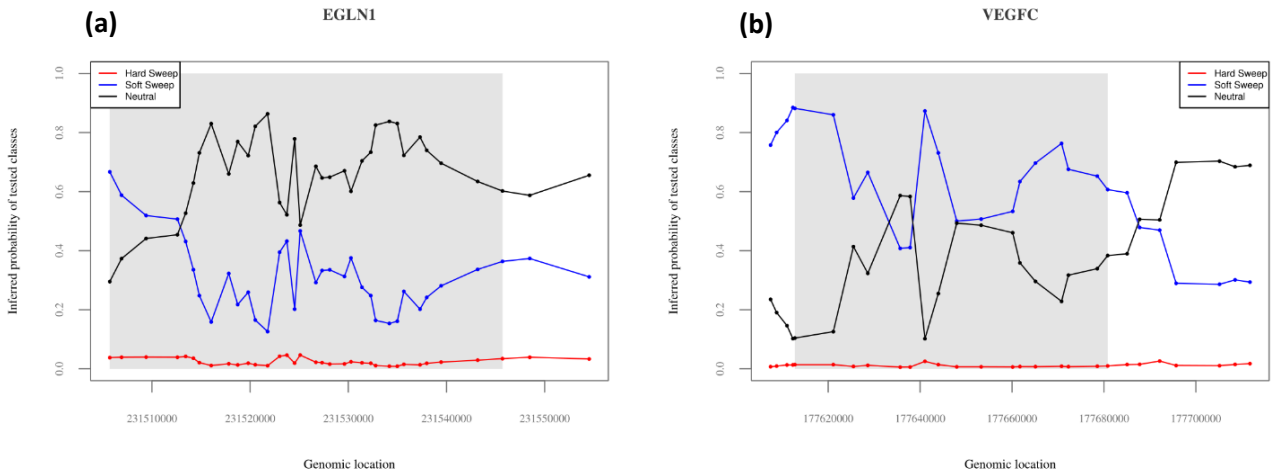
Supplementary Figure S19. *Trendsetter* predicted probabilities obtained for *KITLG* gene in Han Chinese population. Distribution of the *Trendsetter* output probabilities of being classified as hard sweeps (i.e. red curve), soft sweeps (i.e. blue curve) and neutral regions (i.e. black curve) across the putative adaptive *KITLG* gene. The x axis reports the positions of the central SNV in each genomic window identified according to the *Trendsetter* approach while the y axis shows the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent the genomic regions in the considered genes that present values indicative of the action of natural selection according to the *LASSI* statistic. Probabilities values associated to the hard selective sweep class are remarkably elevated in the entire gene locus, reaching values of ~94% in the middle/final portions of the gene. Furthermore, the predicted probabilities associated to the soft selective sweep class assume moderate values with an average of ~13%. The neutral-associated probabilities are extremely low in the all the genomic windows detected for such a locus.



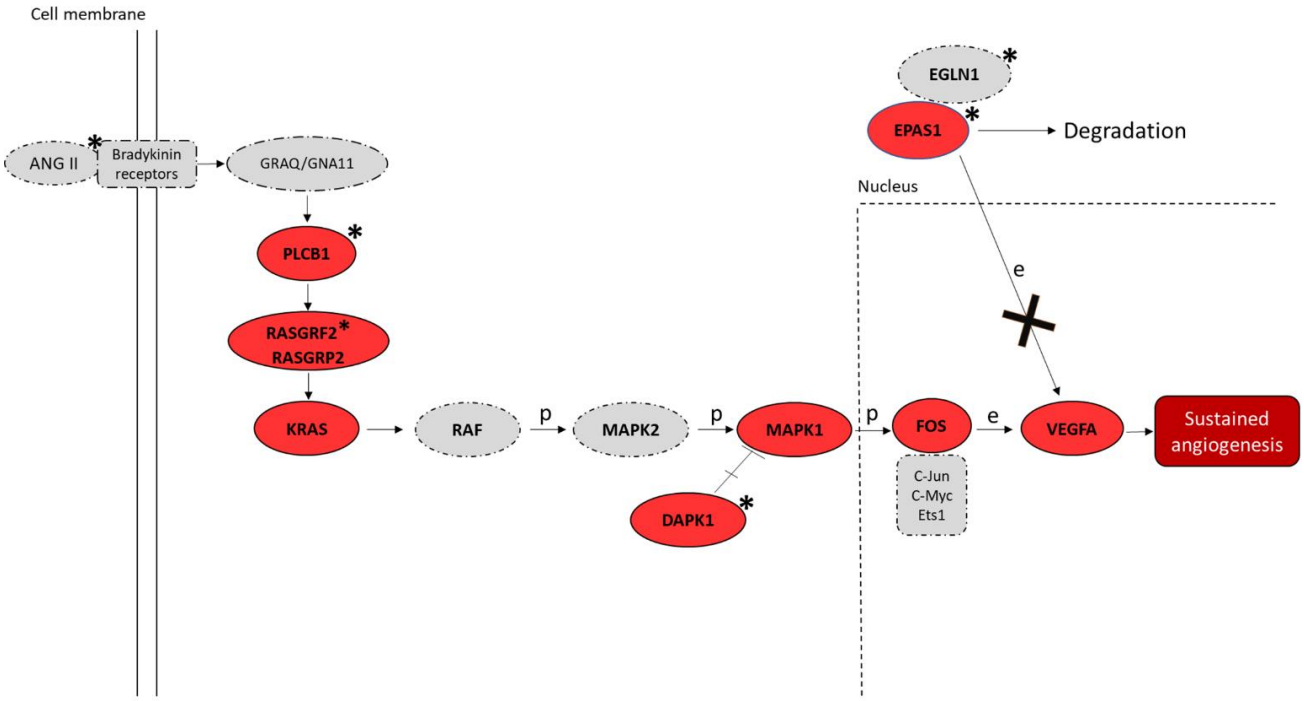
Supplementary Figure S20. Haplotype frequency spectra of the top windows detected as adaptively evolved by LASSI in the *EPAS1* and *EGLN1* genomic regions. Barplots showing haplotype frequency spectra in the genomic windows associated with the highest LASSI *T* value, falling in the top 5% of the obtained distribution and linked to **(A)** *EPAS1* and **(B)** *EGLN1* genes. The x-axis reports the haplotypes detected in the windows, while on y-axes are indicated the frequencies of each haplotype. For these windows the haplotype frequency spectra clearly reflect the pattern of diversity expected under the hard selective sweeps model in which a single predominant haplotype carrying adaptive variants (i.e. sweeping haplotypes represented with the black bars) reaches elevated frequencies in the population. This figure has been taken from Ferraretti et al. (2024).



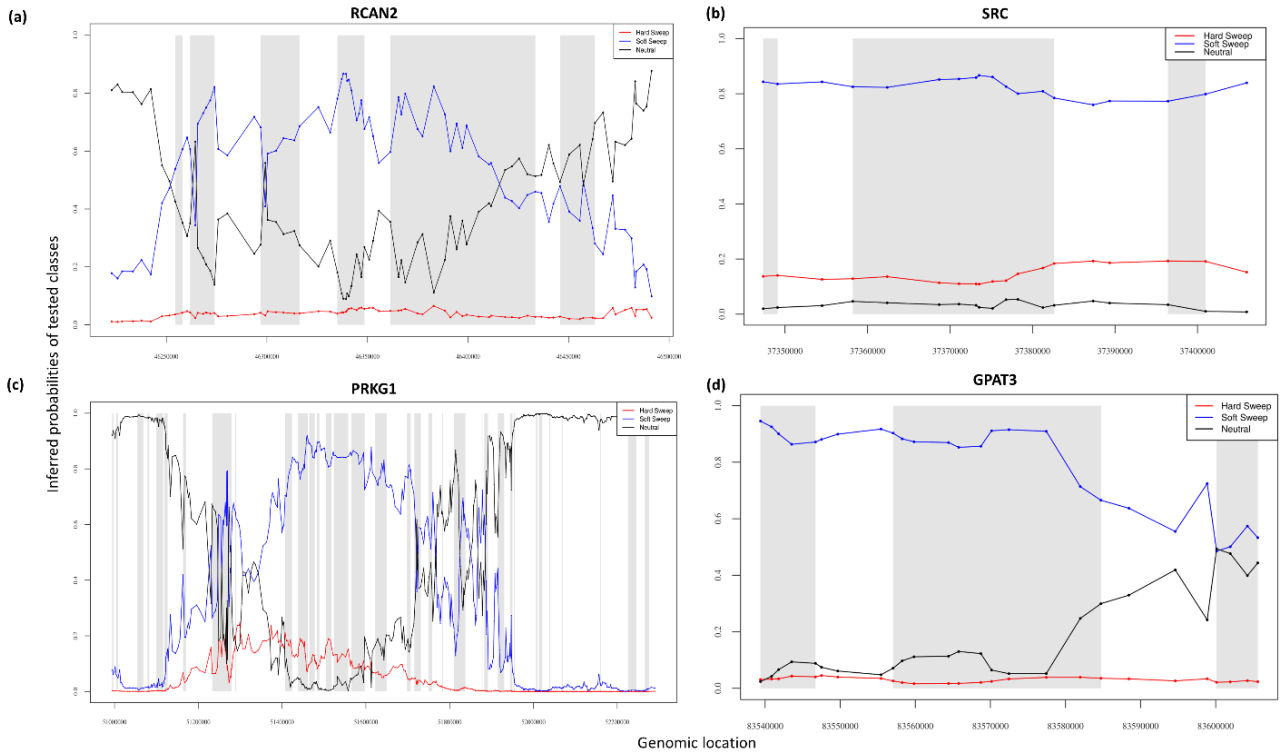
Supplementary Figure S21. *Trendsetter* predicted probabilities at (a) *DAG1* and (b) *PIP4K2A* genes in high-altitude Tibetans. Distribution of the *Trendsetter* output probabilities of being classified as hard sweeps (i.e. red curve), soft sweeps (i.e. blue curve) and neutral regions (i.e. black curve) across *DAG1* and *PIP4K2A* genes. The x axis reports the positions of the central SNV in each genomic window identified according to the *Trendsetter* approach while on the y-axis are shown the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent the genomic regions in the considered genes that present values indicative of the action of natural selection according to the *LASSI* statistic. Although the *Trendsetter* method assigned to the neutral class the majority of the genomic windows detected for such genes, multiple genomic windows located in the first portions of both *DAG1* and *PIP4K2A* are assigned to the hard selective sweep class, overall validating signatures of adaptive evolution pinpointed by LASSI-*signet* methods.



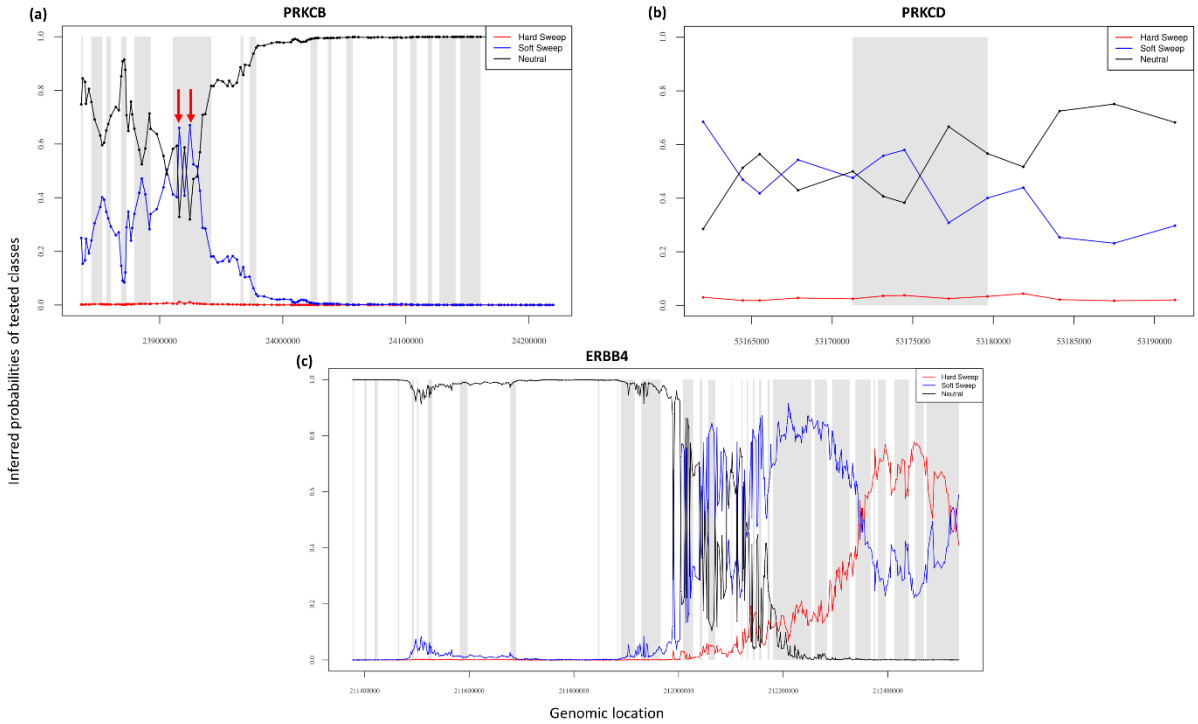
Supplementary Figure S22. *Trendsetter* predicted probabilities at (a) *EGLN1* and (b) *VEGF* genes in high-altitude Tibetans. Distribution of the *Trendsetter* output probabilities of being classified as hard sweeps (i.e. red curve), soft sweeps (i.e. blue curve) and neutral regions (i.e. black curve) across *EGLN1* and *VEGF* genes, previously proposed as genetic targets of natural selection in high-altitude Tibetans. The x axis reports the positions of the central SNV in each genomic window identified according to the *Trendsetter* approach while on the y axis shows the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent the genomic regions in the considered genes that present values indicative of the action of natural selection according to the *LASSI* statistic. **(a)** The *Trendsetter* approach classifies as neutral the majority of the genomic windows detected for the *EGLN1* gene. Nevertheless, four genomic windows located in the first portion of the gene are assigned to the soft selective sweep class, contributing to corroborate previous adaptive evidence advanced for such a locus. **(b)** The probabilities values associated to the soft selective sweep class are remarkably elevated in almost the entire *VEGFC* locus, reaching their greatest values in the initial/middle part of the gene.



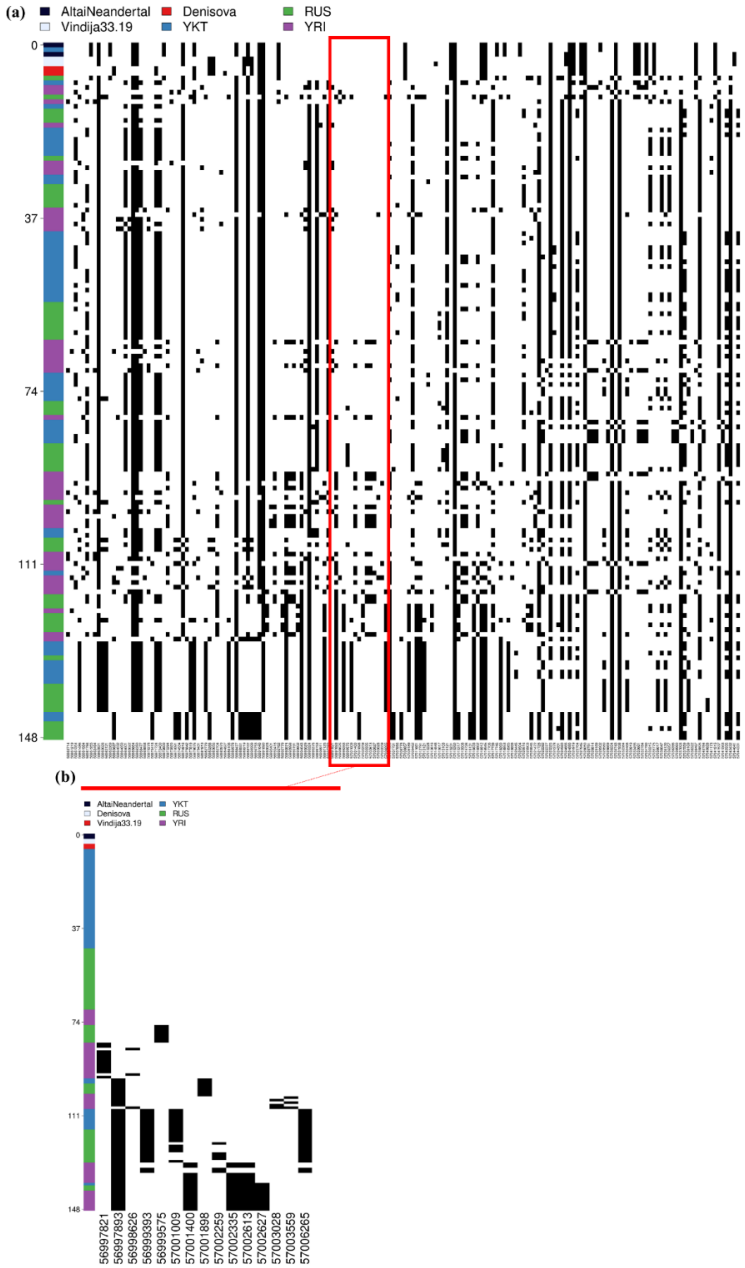
Supplementary Figure S23. Significant gene networks including Denisovan-like derived alleles according to the Signet analysis. Schematic representation of the activation of the RAS/MAPK(ERK) axis after interaction of the bradykinin receptors with their ligands (e.g. ANG II) within the framework of the *Pathways in Cancer* network. Genes supported by both the analyses used to identify archaic adaptive introgression (i.e. VolcanoFinder and *signet*), are highlighted in red and present solid outline. Grey circles with dotted-dashed contour instead indicate genes supported only by the network-based approach, while loci marked with stars are those including genomic windows showing LASSI T statistic within the top 5% of the related distribution. After the interaction between ANG II (active enzyme angiotensin II) and bradykinin receptors, the activation of the Ras protein encoded by *KRAS* mediated by RAS-GTPases (e.g. *RASGRF2*) comports a series of phosphorylation reactions that eventually promotes angiogenesis (Kranenburg et al. 2004). In detail, the phosphorylation of the MAPK1 protein and the prevention of MAPK1-DAPK-1-dependent apoptosis leads to increased MAPK1 activity (Kanehisa and Goto, 2000; Stevens et al. 2007) that causes improved FOS mRNA expression (Monje et al. 2005). FOS, together with other proteins (e.g. Jun), forms the AP-1 transcription factor, which binds to the *VEGF* promoter region upregulating its expression in endothelial cells (Catar et al. 2013) and sustaining angiogenesis when the hypoxia inducible factor -which could be mediated by HIFs activity such as that codified by the *EPAS1* gene- signalling cascade is inhibited (Lorenzo et al. 2014). This figure has been adapted from Ferraretti et al. (2024).



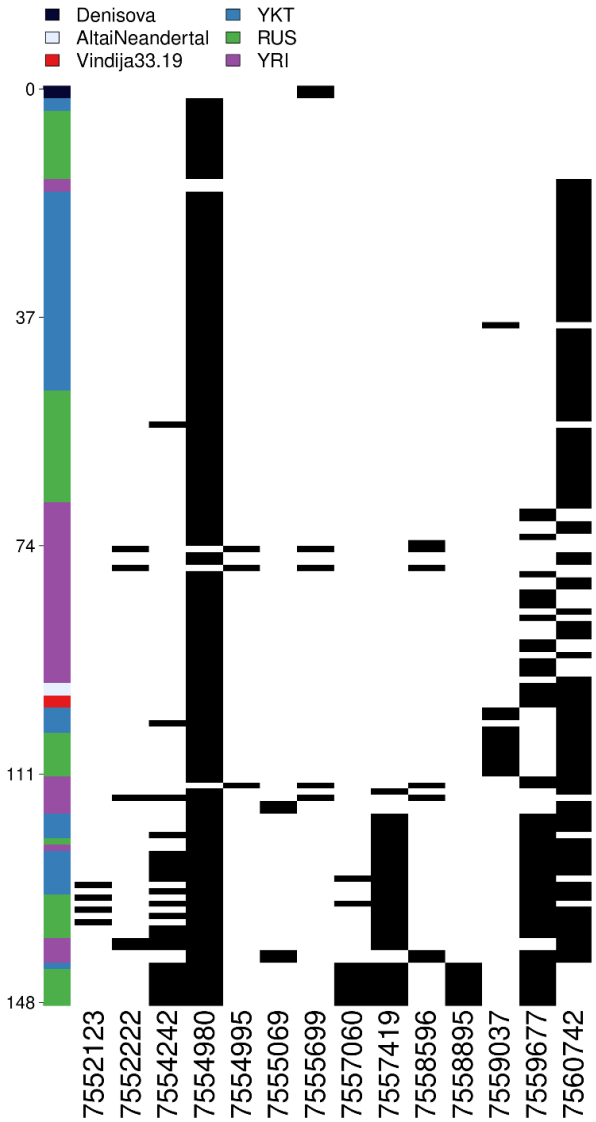
Supplementary Figure S24. *Trendsetter* predicted probabilities at *RCAN2* (a), *SRC* (b), *PRKG1* (c) and *GPAT3* (d) genes in the Yakut population. These genes were included in **(a-b)** *Thyroid hormone signalling* (i.e. *RCAN2* and *SRC*), **(c)** *cGMP-PKG signaling* (i.e. *PRKG1*) and **(d)** *Glycerolipid metabolism* (i.e. *GPAT3*) significant LASSI-*signet* pathways detected for the Yakut population. For all the plots the x axis reports the position of the central SNV in each genomic window identified according to the *Trendsetter* approach and included in these gene. The y axis shows the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent genomic regions in the considered genes that present values indicative of the action of natural selection according to the LASSI statistic. Overall, the predicted probabilities for the soft sweep class are remarkably higher respect to both neutral and hard sweep scenarios, supporting signatures of natural selection ascribable to a polygenic model of adaptation for all these genes.



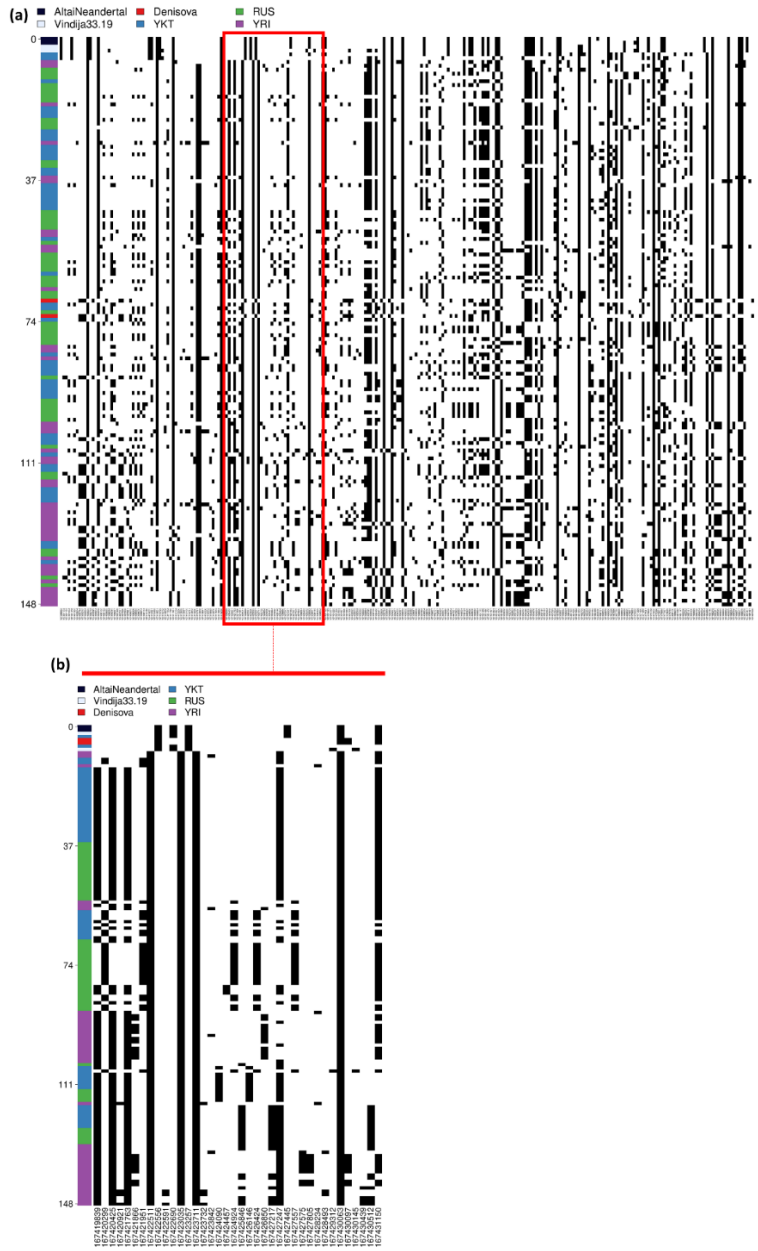
Supplementary Figure S25. *Trendsetter* predicted probabilities at *PRKCB* (a), *PRKCD* (b) and *ERBB4* (d) genes in the Yakut population. These genes were included in **(a-b)** *AGE-RAGE signalling pathway in diabetic complications* (i.e. *PRKCB* and *PRKCD*) and **(c)** *MAPK/PI3K-Akt signalling* (i.e. *ERBB4*) significant LASSI-*signet* pathways detected for the Yakut population. For all the plots the x axis reports the position of the central SNV in each genomic window identified according to the *Trendsetter* approach and included in these gene. The y axis shows the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent genomic regions in the considered genes that present values indicative of the action of natural selection according to the LASSI statistic. **(a)** Overall, neutral output probabilities inferred for the *PRKCB* gene are more elevated respect to those associated to selection events except for the windows indicated with the red arrows and for which the probabilities values of being classified as soft sweeps overcame the neutral ones. **(b)** Output probabilities of being classified as soft sweeps are more elevated respect to the two other classes in the initial/middle portions of the genes. **(c)** Both soft and hard sweeps probabilities are more elevated respect to the neutral one especially in the windows located at the end of the gene. Overall, these results support signals of selection previously identified for all these genes adopting the LASSI-*signet* pipeline.



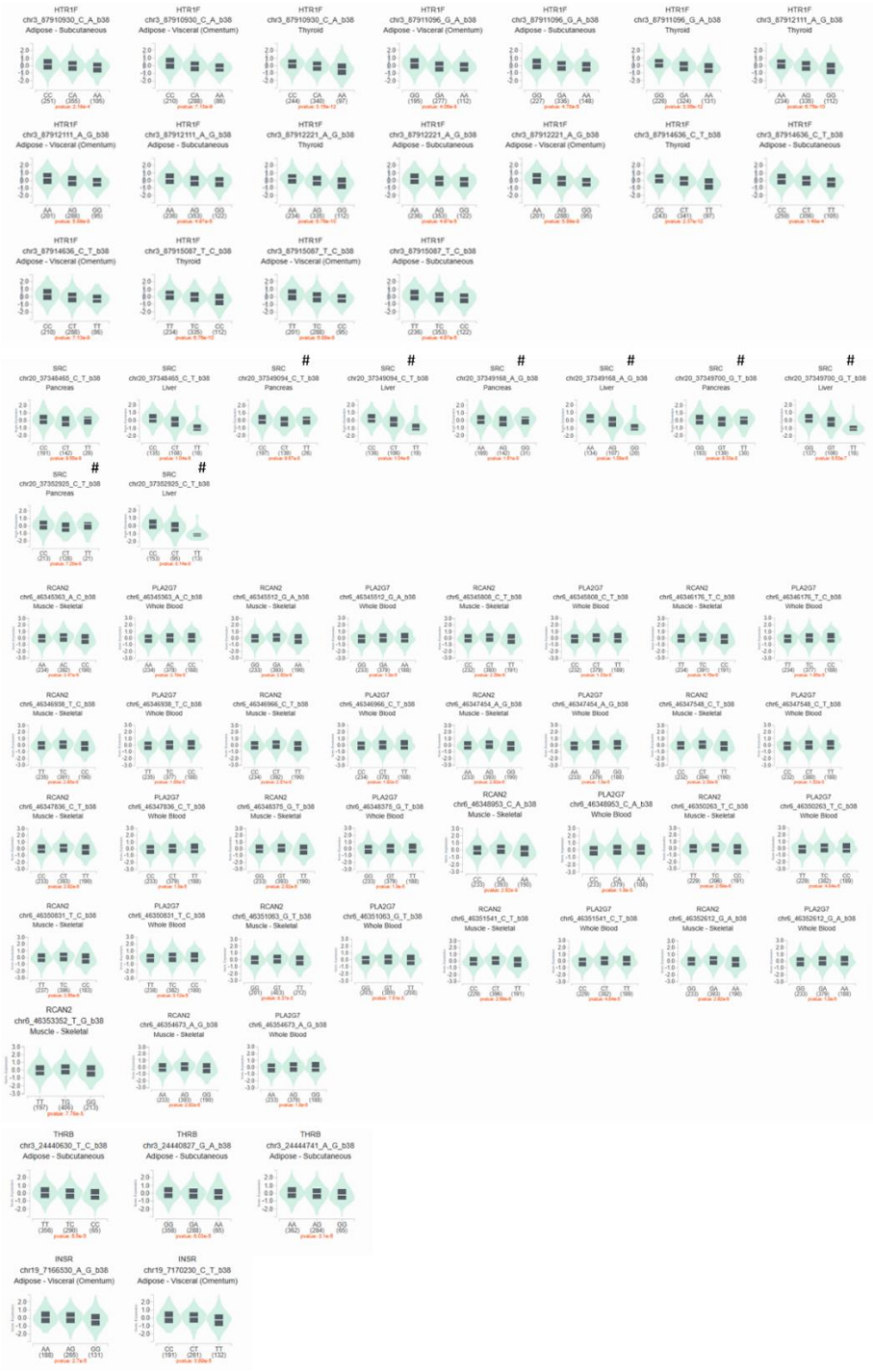
Supplementary Figure S26. Haplotype structure of *PLPP3* gene. (a) Haplostrips plot showing patterns of haplotypic diversity at *PLPP3* gene locus among modern human populations (i.e. Yakut, Russian, Yoruba) and archaic samples (i.e. Denisovan Altai, Neanderthal Altai and Neanderthal Vindija). Single haplotypes (reported in rows) are ranked with respect to the number of differences with the Altai Neanderthal sequence. SNVs are displayed in columns with the ancestral and derived states are reported in white and black colours, respectively. The great majority of the haplotypes belonging to both Yakut (i.e. blue) and Russian (i.e. green) human groups are those presenting the smallest amount of haplotypic pairwise differences respect to all the archaic samples considered, with a single Yakut haplotype being much more similar to one Altai Neanderthal haplotype respect to both Vindija and Denisovan samples. Haplotypes belonging to the Yoruba outgroup population (i.e. purple) are less conserved between each other, clustering mainly in the middle/bottom sections of the plot and thus attesting a remarkable diversity of the Yoruba's *PLPP3* haplotypes respect to the archaic sequences. The red square localizes the position of the overlapped LASSI putative adaptive windows falling in the *PLPP3* gene. **(b-c)** Haplostrips plot build for the overlapped genomic windows presenting a variation pattern indicative of adaptive evolution according to the LASSI statistic in the Yakut population and falling in the *PLPP3* gene. The similarity among archaic samples and Yakut/Russian is even more pronounced in this genomic region, with modern haplotypes identical to all the archaic samples.



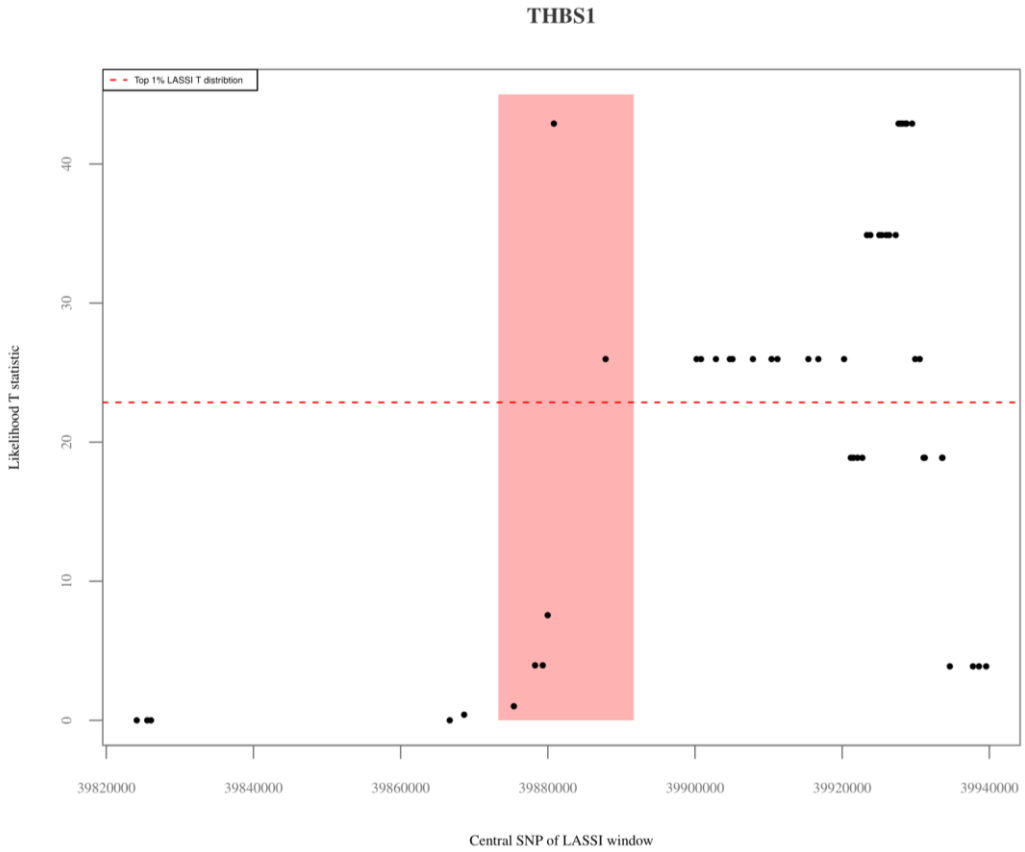
Supplementary Figure S27. Haplotype structure of *ATP1B2* gene. (a) Haplostrips plot showing patterns of haplotypic diversity at *ATP1B2* gene locus among modern human populations (i.e. Yakut, Russian, Yoruba) and archaic samples (i.e. Denisovan Altai, Neanderthal Altai and Neanderthal Vindija). Single haplotypes (reported in rows) are ranked with respect to the number of differences with the Denisovan sequence. SNVs are displayed in columns with the ancestral and derived states reported in white and black colours, respectively. The great majority of the haplotypes belonging to both Yakut (i.e. blue) and Russian (i.e. green) human groups are those presenting the smallest amount of haplotypic pairwise differences respect to the Denisovan archaic samples. The haplotypes belonging to the Yoruba outgroup population (i.e. purple) cluster mainly in the middle/bottom sections of the plot, attesting a remarkable diversity of the Yoruba's *ATP1B2* haplotypes respect to the archaic sequences. The entire genomic region plotted presents a variation pattern indicative of adaptive evolution according to the LASSI statistic in the Yakut population, supporting signals of archaic Denisovan introgression also for such a gene.



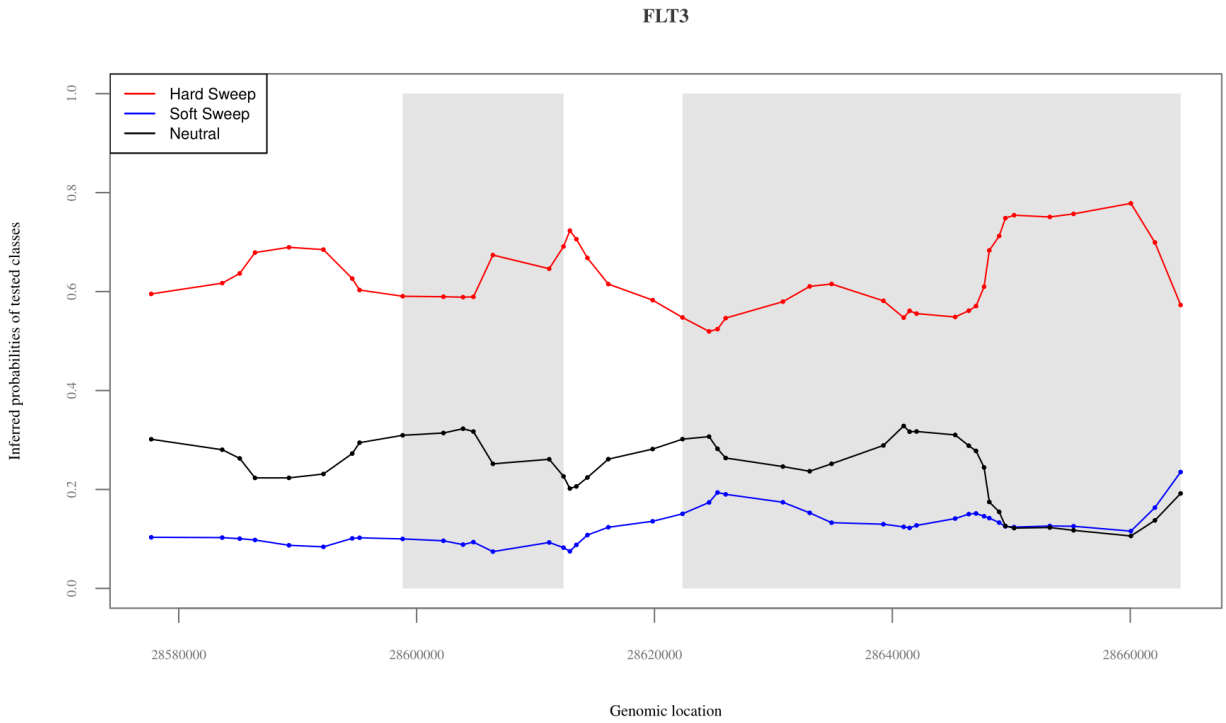
Supplementary Figure S28. Haplotype structure of *CD247* gene. (a) Haplostrips plot showing patterns of haplotypic diversity at *CD247* gene locus among modern human populations (i.e. Yakut, Russian, Yoruba) and archaic samples (i.e. Denisovan Altai, Neanderthal Altai and Neanderthal Vindija). Single haplotypes (reported in rows) are ranked with respect to the number of differences with the Altai Neanderthal sequence. SNVs are displayed in columns with the ancestral and derived states reported in white and black colours, respectively. The great majority of the haplotypes belonging to both Yakut (i.e. blue) and Russian (i.e. green) human groups are those presenting the smallest amount of haplotypic pairwise differences with the Altai Neanderthal (i.e. black) and Vindija Neanderthal (i.e. light blue), even respect to Denisovan (i.e. red) samples. Haplotypes belonging to the Yoruba outgroup population (i.e. purple) cluster mainly at the bottom of the plot, attesting a remarkable diversity of this human group compared to the Neanderthal sequences, showing an opposite haplotypic pattern respect to the Yakut and Russian populations. The red square localizes the position of the overlapped LASSI putative adaptive windows falling in the *CD247* gene. **(b-c)** Haplostrips plot build for LASSI genomic windows presenting a variation pattern indicative of adaptive evolution in the Yakut population and falling in the *CD247* gene. The similarity among Neanderthal Altai haplotypes and Yakut/Russian samples is even more pronounced in this region. In detail, two Yakut haplotypes present lesser differences with respect to the Alta Neanderthal samples even respect to one Vindija haplotype. Such a pattern suggests signatures ascribable to Neanderthal adaptive introgression in *CD247* gene detectable in both Yakut and Russian human groups.



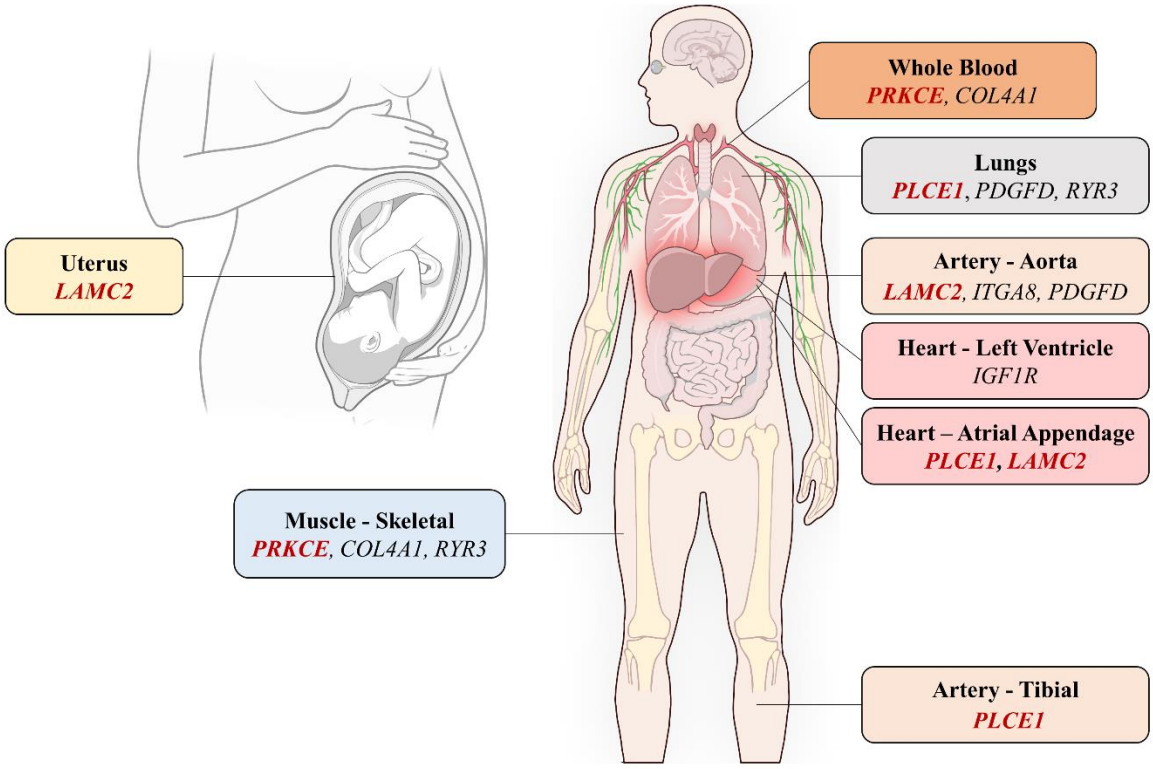
Supplementary Figure S29. Violin plots showing the expression levels of *HTR1F*, *SRC*, *RCAN2*, *THRB*, and *INSR* genes in target tissues depending on genotypes of putative adaptive eQTLs. Candidate adaptive eQTLs at *HTR1F* and *SRC* genes were found to regulate the expression of such genes in white adipose tissues, thyroid, liver and pancreas. Putative adaptive eQTLs found at *INSR* and *THRB* genes have been instead associated in the homozygous state to increased expression of these genes in white adipose tissue. Homozygotes for candidate adaptive eQTLs at *RCAN2* have been shown to decrease expression of such a gene in skeletal muscle, as well as of *PLA2G7* in the blood. The **#** symbol indicates putative adaptive eQTLs on *SRC* gene that do not present significant shifts in frequency between Yakut and control populations after adjusting P-values for multiple tests but only significant nominal P-values. The reported violin plots are downloaded from the GTEx portal (available at https://gtexportal.org/home/).



Supplementary Figure S30. LASSI statistical values across *THBS1* genomic windows in the Maya population. The x axis reports the value of the likelihood *T* statistic calculated by the LASSI approach. The y axis displays the genomic location of SNV located at the centre of each LASSI window. The red dashed line indicates the top 1% threshold of the LASSI distribution obtained. Two top 1% LASSI genomic windows were retrieved for *THBS1* gene (i.e. red rectangle), which particularly present likelihood values peaking at 42.90 and 25.97. Additional 26 top 1% LASSI windows were included in the chromosomal intervals located in the downstream portions of *THBS1* gene.



Supplementary Figure S31. *Trendsetter* predicted probabilities at *FLT3* gene in the Maya population. Distribution of the *Trendsetter* predicted probabilities of being classified as hard sweeps (i.e. red curve), soft sweeps (i.e. blue curve) and neutral regions (i.e. black curve) across the putative adaptive *FLT3* gene. The x axis reports the positions of the central SNV in each genomic window identified according to the *Trendsetter* approach while the y axis shows the values of the *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent the genomic regions in the considered genes that present values indicative of the action of natural selection according to the *LASSI* statistic. Probabilities values associated to the hard selective sweep class are remarkably elevated in the entire gene locus, reaching values of ~80% in the final portions of the gene.



Supplementary Figure S32. Schematic representation of the body districts in which the expression of the *COL4A1*, *IGF1R*, *PDGFD*, *ITGA8*, *RYR3*, *LAMC2*, *PLCE1* and *PRKCE* candidate adaptive genes is modulated by eQTL variants. Genes reported in bold/red are those previously proposed to play an adaptive role also in populations of Tibetan/Sherpa ancestry. Illustration from NIAID NIH BIOART Source (https://bioart.niaid.nih.gov/bioart/519; https://bioart.niaid.nih.gov/bioart/420). The figure has been taken from Ferraretti et al. (2025).

Supplementary Table S1. Composition of the sub-datasets analysed. List of populations with relative sample size and reference study included in the **(a)** African, **(b)** European, **(c)** North/East Asian and **(d)** American macro-areas. Please consult Ferraretti et al. (2024) and Ferraretti et al. (2025) for further detail concerning the structure of the datasets used to investigate the genomic variation of the WGS belonging to high-altitude Tibetans (Jeong et al. 2018) and Aymara (Lindo et al. 2018). Columns: **(1)** Population name; **(2)** Number of individuals retained per population; **(3)** Reference panel.

POPULATION	N. INDIVIDUALS	REFERENCE STUDY
(a) Macroarea Africa		
African_Ancestry_SW	55	Auton et al. 2015
African_Carribean	95	Auton et al. 2015
BantuKenya	10	Bergström et al. 2020
BantuSouthAfrica	8	Bergström et al. 2020
Bedouin	46	Bergström et al. 2020
Biaka	21	Bergström et al. 2020
Brahui	25	Bergström et al. 2020
CEPH	99	Auton et al. 2015
Esan	98	Auton et al. 2015
Gambian	113	Auton et al. 2015
Han	136	Auton et al. 2015; Bergström et al. 2020
Luhya	97	Auton et al. 2015
Mandinka	22	Bergström et al. 2020
Maya	21	Bergström et al. 2020
Mbuti	11	Bergström et al. 2020
Mende	85	Auton et al. 2015
Mozabite	27	Bergström et al. 2020
Pima	5	Bergström et al. 2020
San	4	Bergström et al. 2020
Yoruba	130	Auton et al. 2015; Bergström et al. 2020
(b) Macroarea Europe		
Adygei	16	Bergström et al. 2020
Basque	23	Bergström et al. 2020
Bedouin	46	Bergström et al. 2020

BergamoItalian	12	Bergström et al. 2020
Brahui	25	Bergström et al. 2020
British	91	Auton et al. 2015
CEPH	99	Auton et al. 2015
Finnish	99	Auton et al. 2015
French	28	Bergström et al. 2020
Han	136	Auton et al. 2015; Bergström et al. 2020
Iberian	107	Auton et al. 2015
Maya	21	Bergström et al. 2020
North_ITA	29	Sazzini et al. 2020
Orcadian	15	Bergström et al. 2020
Pima	5	Bergström et al. 2020
Russian	25	Bergström et al. 2020
Bouganville	7	Bergström et al. 2020
Sardinian	28	Bergström et al. 2020
South_ITA	9	Sazzini et al. 2020
Toscani	115	Auton et al. 2015; Bergström et al. 2020
Yoruba	130	Auton et al. 2015; Bergström et al. 2020
Luhya	25	Auton et al. 2015
(c) Macroarea North/East Asia		
Burusho	24	Bergström et al. 2020
Cambodian	9	Bergström et al. 2020
CEPH	99	Auton et al. 2015
Dai	102	Bergström et al. 2020, Auton et al. 2015
Daur	9	Bergström et al. 2020
Han	136	Bergström et al. 2020, Auton et al. 2015
Hazara	19	Bergström et al. 2020
Hezhen	9	Bergström et al. 2020
Lahu	8	Bergström et al. 2020
Maya	21	Bergström et al. 2020
Miao	10	Bergström et al. 2020

Mongolian	9	Bergström et al. 2020
Naxi	7	Bergström et al. 2020
Northern Han	10	Bergström et al. 2020
Oroqen	9	Bergström et al. 2020
Pima	5	Bergström et al. 2020
Russian	25	Bergström et al. 2020
She	9	Bergström et al. 2020
Tu	10	Bergström et al. 2020
Tujia	9	Bergström et al. 2020
Uygur	10	Bergström et al. 2020
Xibo	9	Bergström et al. 2020
Yakut	25	Bergström et al. 2020
Yi	10	Bergström et al. 2020
(d) Macroarea America		
Puertorican	104	Auton et al. 2015
Colombian	97	Auton et al. 2015
Peruvian	85	Auton et al. 2015
Brahui	25	Bergström et al. 2020
Bedouin	45	Bergström et al. 2020
Piapoco	1	Mallick et al. 2016
Han	136	Bergström et al. 2020, Auton et al. 2015
Surui	8	Bergström et al. 2020
Maya	28	Bergström et al. 2020; Jiménez-Kaufmann et al. 2022
Yoruba	130	Bergström et al. 2020, Auton et al. 2015
Karitiana	11	Bergström et al. 2020
Pima	13	Bergström et al. 2020
Nahua	18	Jiménez-Kaufmann et al. 2022
Totonac	6	Jiménez-Kaufmann et al. 2022
Zapotec_Mazatec	10	Jiménez-Kaufmann et al. 2022
Tarahumara	3	Jiménez-Kaufmann et al. 2022

Huichol	2	Jiménez-Kaufmann et al. 2022
Tzotzil	4	Jiménez-Kaufmann et al. 2022
CEPH	99	Auton et al. 2015
Luhya	25	Auton et al. 2015
Mexican	64	Auton et al. 2015

Supplementary Table S2. Accuracy values of the trained *Trendsetter* classifiers. Columns: **(1)** Population labels; **(2)** Accuracy values.

Population	Accuracy
YRI	0.95
CEPH	0.92
RUS	0.89
HAN	0.87
YKT	0.87
TIB	0.95
NAH	0.72
PIM	0.72
MAY	0.69

Supplementary Table S3. Tables reporting the metrices used to evaluate the reliability of the *Trendsetter* classifiers obtained for **(a)** Yoruba, **(b)** CEPH, **(c)** Russian, **(d)** Han Chinese, **(e)** Yakut, **(f)** high-altitude Tibetan, **(g)** Nahua, **(h)** Maya, and **(i)** Pima populations. Columns: **(1)** Population label and tested class; **(2)** Precision; **(3)** Recall; and **(4)** f1 score values. The last row shows the values of the weighted average obtained for the f1 score.

(a) YRI	**Precision**	**Recall**	**f1 score**
Hard Sweeps	0.96	0.96	0.96
Soft Sweeps	0.94	0.92	0.93
Neutral	0.96	0.98	0.97
Weighted Average	.	.	0.95

(b) CEPH	**Precision**	**Recall**	**f1 score**
Hard Sweeps	0.92	0.93	0.92
Soft Sweeps	0.89	0.86	0.88
Neutral	0.95	0.96	0.96
Weighted Average	.	.	0.92

(c) RUS	**Precision**	**Recall**	**f1-score**
Hard Sweeps	0.90	0.89	0.90
Soft Sweeps	0.84	0.85	0.84
Neutral	0.94	0.94	0.94
Weighted Average	.	.	0.89

(d) HAN	**Precision**	**Recall**	**f1-score**
Hard Sweeps	0.88	0.89	0.88
Soft Sweeps	0.82	0.8	0.81
Neutral	0.91	0.93	0.92
Weighted Average	.	.	0.87

(e) YKT	**Precision**	**Recall**	**f1 score**
Hard Sweeps	0.86	0.88	0.87
Soft Sweeps	0.82	0.80	0.81
Neutral	0.93	0.94	0.93
Weighted Average	.	.	0.87

(f) TIB	**Precision**	**Recall**	**f1 score**
Hard Sweeps	0.95	0.95	0.95
Soft Sweeps	0.93	0.92	0.92
Neutral	0.96	0.98	0.97
Weighted Average	.	.	0.95

(g) NAH	Precision	Recall	f1 score
Hard Sweeps	0.7	0.69	0.69
Soft Sweeps	0.61	0.57	0.59
Neutral	0.83	0.89	0.86
Weighted Average	.	.	0.71

(h) MAY	Precision	Recall	f1 score
Hard Sweeps	0.66	0.65	0.65
Soft Sweeps	0.59	0.56	0.57
Neutral	0.8	0.86	0.83
Weighted Average	.	.	0.68

(i) PIM	Precision	Recall	f1 score
Hard Sweeps	0.71	0.69	0.7
Soft Sweeps	0.63	0.59	0.61
Neutral	0.81	0.87	0.84
Weighted Average			0.72

Supplementary Table S4. LASSI-*signet* results obtained for (a) African, (b) European, (c) North/East Asian and (d) American macro-areas. Columns: **(1)** Label of the ethnic group analysed; **(2)** Pathways encompassing the significant networks; **(3)** Number of genes included in each network; **(4)** Score calculated by the *signet* approach for each network; **(5)** P-values; **(6)** List of genes included in each network. The genes further supported by the *Trendsetter* approach are reported in bold.

Dataset	Pathway	Subnetwork size	Subnetwork Score	P. value	Subnetwork genes
(a) Macroarea Africa					
BNT	Focal adhesion	24	10.32	0.005	COL2A1,COL4A2,COL4A3,COL4A4,COL9A1,LAMB4,ITGA11,FLNB,FLNC,LAMA1,PARVB,ITGA6,ITGA5,ITGB3,LAMA2,LAMC1,RELN,THBS4,ACTG1,TNR,VCL,VWF,TLN2,ITGA8
ESN	Focal adhesion	30	8.30	0.007	COL4A1,COL4A2,COL4A3,COL6A3,COL9A2,FYN,SHC2,GRB2,TNC,IBSP,ITGA9,ITGB1,ITGB3,ITGB5,LAMA3,LAMA4,LAMA5,MET,PRKCA,RELN,PTK2,SRC,THBS3,THBS4,VTN,VWF,CAPN2,TLN2,ITGA8,CAV2
ESN	Chemokine signaling pathway	8	6.62	0.031	PLCB1,GNB1,GNG4,GNG7,GNGT1,PIK3CD,PLCB4,PREX1
GMB	ECM-receptor interaction	35	9.91	0.005	COL1A2,COL4A1,COL4A3,COL4A4,COL9A1,COL9A2,COL9A3,COL6A6,DAG1,LAMB4,ITGA11,SV2C,FN1,COL6A5,TNC,ITGA6,ITGA1,ITGA2,ITGA3,ITGA9,ITGB3,LAMA2,LAMA4,LAMA5,LAMB1,LAMC1,LAMC2,GP6,THBS4,TNR,TNXB,VTN,ITGA10,ITGA8,SV2B
GMB	JAK-STAT signaling pathway	13	7.58	0.012	EGFR,GHR,IFNAR2,IFNGR2,IL10RB,IL15,JAK3,LEPR,LIFR,IL22,PDGFB,PDGFRA,PRLR
GMB	AMPK signaling pathway	11	6.78	0.027	CFTR,MLYCD,GYS2,LIPE,PRKAG2,PFKFB3,PRKAA2,RPTOR,TSC1,CAB39L,CD36
MND	Focal adhesion	25	8.66	0.008	COL4A1,COL4A2,COL4A3,COL4A4,COL9A1,COL6A6,ITGA11,FN1,PARVB,TNC,ITGA6,ITGA1,ITGA3,ITGA9,ITGB6,LAMA2,PARVA,RELN,THBS3,TLN1,TNR,VCL,VTN,TLN2,ACTN1
MOZ	Starch and sucrose metabolism	8	7.44	0.022	TREH,GANC,GYS2,ENPP1,GBA3,PYGL,SI,MGAM
YRI	ECM-receptor interaction	27	6.67	0.017	LAMC3,COL2A1,COL4A1,COL4A2,COL4A3,COL4A4,COL6A3,COL9A2,COL6A6,DAG1,ITGA11,SV2C,COL6A5,ITGA1,ITGA2B,ITGB1,ITGB8,LAMA2,LAMA3,LAMA4,LAMB2,LAMB3,**LAMC2**,THBS2,THBS4,**VTN**,SV2B
YRI	JAK-STAT signaling pathway	22	6.07	0.034	IL22RA2,CSF2RB,CTF1,IFNLR1,EGFR,GHR,IL19,IFNA6,IFNAR2,IFNGR2,IL7,IL10RB,**IL15**,JAK1,JAK2,JAK3,**LEPR**,PDGFB,PDGFRA,PIK3R2,PRLR,STAT1
YRI	Tuberculosis	6	5.82	0.047	**HLA-DOB,HLA-DPA1,HLA-DPB1,HLA-DRB1,HLA-DRB5**,NFYC

(b) Macroarea Europe					
BRT	Sphingolipid metabolism	13	6.20	0.049	ACER1,CERS3,CERS6,GALC,SGMS1,ACER2,ASAH1,ACER3,SMPD4,ASAH2,CERK,PLPP3,SPHK1
BSQ	MAPK signaling pathway	19	7.94	0.010	ERBB4,EREG,FGF1,FGF5,FGF10,FLT3,ANGPT1,ANGPT2,GRB2,IGF1R,MET,KITLG,NTRK1,NTRK2,ANGPT4,PDGFRA,TEK,VEGFC,PDGFD
CEPH	MAPK signaling pathway	19	7.94	0.011	**ERBB4**,EREG,**FGF1**,FGF5,FGF10,**FLT3**,ANGPT1,ANGPT2,**GRB2**,IGF1R,MET,**KITLG**,NTRK1,**NTRK2**,ANGPT4,**PDGFRA**,TEK,VEGFC,PDGFD
CEPH	Chemokine signaling pathway	6	7.42	0.017	PRKCZ,**PARD3**,**VAV1**,**VAV2**,PIK3R3,CDC42
FIN	MAPK signaling pathway	12	7.29	0.014	ERBB4,FGFR2,FLT3,ANGPT2,GRB2,IGF1R,INSR,KITLG,NTRK2,PDGFRA,VEGFC,PDGFD
FIN	Purine metabolism	12	6.73	0.025	ADCY1,PDE10A,ADCY3,ADK,GUCY1A1,NME7,PDE11A,PDE1C,PDE3A,PDE7A,PDE9A,ENPP3
FIN	Glycerolipid metabolism	14	6.58	0.030	AGPAT2,MGLL,MOGAT1,PNLIPRP3,MBOAT1,DGKB,DGKG,PNPLA3,DGAT2,GPAT3,PLPP1,PLPP3,DGKI,LPIN2
FIN	Pathogenic Escherichia coli infection	7	6.34	0.041	ABI1,ARPC5,WASF3,CYFIP1,ACTB,ARHGEF1,CDC42
IBR	Glycerolipid metabolism	16	6.19	0.031	MGLL,PNLIPRP3,DGKA,DGKB,DGKG,LCLAT1,LIPC,AGK,MOGAT2,PNPLA3,GPAT3,DGKZ,DGKD,PLPP3,DGKI,LPIN2
IBR	Mucin type O-glycan biosynthesis	15	5.86	0.049	GALNT5,GALNT13,GALNTL5,GALNT1,GALNT2,GALNT3,GALNTL6,GALNT9,ST6GALNAC1,C1GALT1,GALNT11,ST3GAL2,GALNT14,GALNT12,GCNT3
RUS	Purine metabolism	24	9.37	0.006	ADCY2,PDE10A,**ADCY3**,ADCY5,ADCY8,AK8,AK2,FHIT,AK5,PDE7B,AMPD3,NME7,APRT,PDE11A,NT5C3A,PDE1A,PDE1C,PDE3A,PDE4B,PDE4D,PDE6A,PDE6C,ENPP3,RRM2
RUS	Glycerolipid metabolism	15	7.28	0.018	MGLL,MOGAT1,MBOAT1,**DGKB**,DGKG,LCLAT1,LIPC,**PNLIP**,AGPAT3,**MOGAT2**,PNPLA3,**PLPP1**,PLPP3,**DGKI**,LIPG
RUS	Rap1 signaling pathway	14	6.66	0.037	EGF,**FGF5**,FGF10,PLCB1,ANGPT1,ANGPT2,INSR,PDGFB,**PDGFRA**,PLCB4,PRKCA,PRKCB,VEGFC,PDGFD
SAR	Purine metabolism	25	8.11	0.009	ADA,ADCY2,PDE10A,ADCY3,ADCY5,ADCY8,ADK,AK8,AK9,FHIT,PDE7B,GMPR,LOC390877,NME4,PDE11A,NT5C3A,PDE1C,PDE3A,PDE4B,PDE4D,PDE6C,PDE9A,ENPP1,RRM1,ENTPD5
SAR	Chemokine signaling pathway	7	7.20	0.018	PTK2B,PLCB1,GNAI1,GNG4,LYN,PLCB4,PRKCB
SAR	Ras signaling pathway	9	6.82	0.029	EFNA5,FLT1,FLT3,INSR,KITLG,PDGFRA,TEK,VEGFC,PDGFD
SAR	Glycerophospholipid metabolism	8	6.51	0.040	PLB1,DGKB,DGKG,PLD1,LPCAT2,CHPT1,PLA2G6,DGKI
SAR	Oxytocin signaling pathway	8	6.50	0.041	ADCY5,EGFR,GNAI1,ITPR1,PLCB4,PRKCA,PRKCB,CD38
SAR	Glycerolipid metabolism	4	6.20	0.057	DGKB,DGKG,PNPLA3,DGKI
TUS	Purine metabolism	30	9.23	0.006	ADCY2,PDE10A,ADCY3,ADCY5,ADCY8,ADK,AK8,ADCY4,AK4,AK9,FHIT,AK5,PDE7B,AMPD3,GUCY1A2,GUCY2C,NME7,NME4,NT5C3A,PDE1C,PDE3A,PDE4B,PDE4D,PDE6A,PDE6C,PDE9A,ENPP1,PKLR,PDE5A,ENTPD3
TUS	MAPK signaling pathway	22	6.93	0.018	EGF,ERBB2,ERBB4,EREG,FGF1,FGF9,FGFR2,FLT1,FLT3,ANGPT2,GRB2,IGF1R,INSR,KDR,KITLG,NTF3,NTRK2,PDGFB,PDGFRA,PDGFC,TEK,PDGFD
TUS	Glycerolipid metabolism	9	6.80	0.021	MGLL,PNLIPRP3,DGKB,DGKG,LPL,PNPLA3,DGAT2,PLPP3,DGKI
TUS	Sphingolipid metabolism	14	6.53	0.031	ACER1,CERS3,CERS6,GALC,SGMS1,GBA1,ACER2,ASAH1,ACER3,SMPD3,SMPD4,CERK,SGPP1,PLPP3

(c) Macroarea North/East Asia					
DAI	Pentose and glucuronate interconversions	10	7.61	0.011	UGT2A1,UGT1A10,UGT1A8,UGT1A7,UGT1A6,UGT1A5,UGT1A9,UGT2A2,UGT2B4,KL
DAI	Rap1 signaling pathway	11	7.10	0.016	EGFR,FLT1,PLCB1,ANGPT1,KDR,KITLG,PDGFB,PDGFRA,PLCB4,PRKCB,PDGFD
DAI	Glycerophospholipid metabolism	9	6.64	0.028	MBOAT1,DGKB,DGKG,PISD,PLD1,LPCAT1,PLA2G6,PLA2G4C,DGKI
HAN	JAK-STAT signaling pathway	10	7.06	0.019	**IL23R**,**IFNL3**,IFNAR2,IFNGR2,**IL7**,JAK1,LEP,**LEPR**,PDGFRA,PRLR
HAN	PI3K-Akt signaling pathway	13	6.13	0.052	**ERBB4**,**FGF2**,FLT1,FLT3,FLT4,ANGPT1,INSR,**IRS1**,KDR,**KITLG**,NTRK2,PDGFRA,PDGFD
TIB	Phosphatidylinositol signaling system	19	9.15	0.005	DGKB,DGKG,**PIKFYVE**,PLCB1,IMPA2,INPP1,INPP5A,INPP5D,ITPK1,PLCE1,PIK3C2G,**PIP4K2A**,PLCB4,PLCG2,PIP5K1B,INPP4B,SYNJ2,PLCZ1,DGKI
TIB	ECM-receptor interaction	18	7.37	0.011	COL4A1,COL4A2,COL4A3,COL4A4,COL6A3,COL6A6,**DAG1**,LAMB4,ITGA11,SV2C,LAMA1,ITGA2,ITGA3,ITGA9,ITGB4,LAMA2,LAMC1,LAMC2
TIB	Adrenergic signaling in cardiomyocytes	10	6.52	0.028	PPP2R1A,PPP2R2A,PPP2R2B,PPP2R3A,**PPP2R5C**,CACNA2D3,RYR2,CACNA1C,CACNA1S,CAMK2D
YKT	cGMP-PKG signaling pathway	21	10.21	0.004	KCNMB2,IRAG1,CREB1,PLCB1,ITPR1,ITPR2,KCNMA1,ATP1B2,ATP2B4,PLCB4,PRKCE,**PRKG1**,PRKG2,CREB3L2,SLC8A1,SLC8A3,TRPC6,CACNA1C,CACNA1D,GTF2IRD1,CREB5
YKT	Rap1 signaling pathway	12	7.95	0.010	CRKL,EFNA5,EGFR,PLCB1,ANGPT2,HGF,INSR,PDGFRA,PLCB4,PRKCB,VEGFC,PDGFD
YKT	MAPK signaling pathway	16	7.10	0.020	EFNA5,EGFR,ERBB3,**ERBB4**,FLT1,FLT3,ANGPT1,ANGPT2,GRB2,HGF,INSR,NTRK1,PDGFB,PDGFRA,VEGFC,PDGFD
YKT	PI3K-Akt signaling pathway	17	7.10	0.020	EFNA5,EGFR,ERBB3,**ERBB4**,FLT1,**FLT3**,ANGPT1,ANGPT2,GRB2,HGF,INSR,NTRK1,PDGFB,PDGFRA,SOS1,VEGFC,PDGFD
YKT	Glycerolipid metabolism	9	7.04	0.021	MGLL,MBOAT1,DGKB,DGKG,LIPC,**GPAT3**,PLPP3,DGKI,LIPG
YKT	AGE-RAGE signaling pathway in diabetic complications	10	6.87	0.026	PLCB1,**NFKB1**,PLCE1,PLCB4,PLCG2,**PRKCB**,**PRKCD**,PRKCE,MAPK10,MAPK12
YKT	Serotonergic synapse	11	6.37	0.045	PLCB1,GNAI1,GNG7,GNGT1,**HTR1F**,ITPR1,ITPR2,KCNJ6,PLCB4,GNG2,GNG12
YKT	Thyroid hormone signaling pathway	7	6.35	0.047	**RCAN2**,HIF1A,ATP1B2,PIK3R2,RXRA,**SRC**,THRB

(d) Macroarea America					
MAY	Acute myeloid leukemia	6	6.61	0.043	**FLT3**,PIK3CA,SPI1,STAT5A,STAT5B,PIK3R3
MAY	Axon guidance	8	6.53	0.046	**NFATC3**,**PLCG2**,WNT4,**PPP3CA**,PPP3CB,FZD3,PIK3R3,SSH2
MAY	Insulin resistance	8	6.50	0.048	IRS1,PRKCD,**PRKCE**,PRKCQ,**PRKCZ**,MAPK10,**PTPRF**,PIK3R3
MAY	Chemokine signaling pathway	6	6.33	0.058	DOCK2,PTK2B,**HCK**,LYN,PIK3R3,ELMO1
MAY	Phospholipase D signaling pathway	8	7.35	0.020	**EGFR**,GAB1,CYTH4,**RHOA**,PDGFRA,PDGFD,PIK3R3,GAB2
NAH	Chemokine signaling pathway	9	7.70	0.013	GNG3,**GNG4**,**GNG7**,GNGT1,LYN,PLCB4,**PRKCZ**,PIK3R3,ELMO1
NAH	Focal adhesion	9.00	6.404	0.049	**VAV3**, EGFR, IGF1R, PDGFRA, SHC3, **VAV2**, VEGFC, **PDGFD**, **PIK3R3**
PIM	Chemokine signaling pathway	9	8.02	0.010	**PTK2B**,PLCB1,**GNGT1**,**HCK**,PLCB4,GNG2,**PRKCB**,**GNG12**,PIK3R3
PIM	Fc gamma R-mediated phagocytosis	7	7.58	0.014	DOCK2,**GSN**,PAK1,**PLD1**,RAC2,PIP5K1B,PIK3R3
AYM *Ferraretti et al., 2025	Apelin signaling pathway	16	9.21	0.005	ADCY1,ADCY2,ADCY3,PLCB1,GNAQ,GNG4,GNG7,GNGT1,ITPR1,ITPR2,PRKAG2,PLCB4,GNG2,PRKCE,RYR2,RYR3
AYM *Ferraretti et al., 2025	Rap1 signaling pathway	18	7.01	0.019	CRKL,CSF1,EGFR,FGF1,FGF10,PLCB1,GNAQ,ANGPT1,ANGPT2,IGF1R,INSR,KDR,PDGFRA,PLCB4,PRKCB,TEK,PDGFD,FGF17
AYM *Ferraretti et al., 2025	Vascular smooth muscle contraction	8	6.94	0.021	PLCB1,GNAQ,ITPR1,ITPR2,PLCB4,PRKACA,PRKCE,PRKCH
AYM *Ferraretti et al., 2025	Focal adhesion	16	6.29	0.044	COL4A1,COL4A2,COL4A3,COL4A4,COL9A1,FLNB,TNC,ITGA9,ITGB3,LAMC1,LAMC2,RELN,PARVG,VWF,ITGA8,ACTN1

Supplementary Table S5. Table reporting the genes included in significant LASSI-*signet* networks for both Yakut and control Russian populations and whose expression was altered in several mice tissues after cold exposure. **(1)** Gene name; **(2)** Chromosome; **(3)** NCBI ID; **(4)** Tissue in which the expression of each gene was altered after cold exposure; **(5)** Type of expression alteration; **(6)** Reference study; **(7)** Genes supported also in the control population as involved in the modulation of polygenic adaptations; **(8)** Genes showing variation patterns indicative of archaic introgression according to the Sprime statistic.

Gene	Tissue	Type of alteration	Reference study	Supported in controls	Putative introgressed segments
ADCY3	Mice BAT	Up regulated	Shore et al. 2013	√	√
CAMK2D	Mice Liver	Up regulated	Shore et al. 2013	.	√
CLDN10	Mice Placenta	Down regulated	Gosh et al. 2021	.	.
CLDN11	Mice Liver	Up regulated	Shore et al. 2013	.	.
EGFR	Mice Adipose tissue	Down regulated	Shore et al. 2013	.	√
GPAT3	Mice BAT	Up regulated	Labbé et al. 2015	√	.
GTF2IRD1	Mice BAT	Up regulated	Shore et al. 2013	.	.
IMPA2	Mice BAT	Up regulated	Shore et al. 2013	.	√
INSR	Mice Adipose tissue, mice BAT	Up regulated	Shore et al. 2013; Wang & Wahl 2014	√	.
KCNN2	Mice Liver	Down regulated	Shore et al. 2013	.	√
NTRK1	Mice BAT	Up regulated	Shore et al. 2013	.	.
PDE4D	Mice Liver	Up regulated	Shore et al. 2013	√	√
PDGFRA	Mice Liver	Down regulated	Shore et al. 2013	√	.
PLCB1	Mice BAT	Down regulated	Shore et al. 2013	√	.
SLC8A1	Mice BAT	Down regulated	Shore et al. 2013	.	.

Supplementary Table S6. Table reporting the haplotype frequencies inferred for candidate adaptive genes in the Yakut and Russian populations. **(1)** Gene name; **(2)** Population in which the haplotype frequencies are calculated (Yakut - YKT; Russian - RUS); **(3)** Start position of the inferred haplotype block; **(4)** End position of the inferred haplotype block; **(5)** Haplotype allele composition; **(6)** Haplotype frequencies. In red are indicated candidate adaptive eQTLs.

Gene	Population	Start position	End position	Inferred haplotypes	Haplotype frequency
RCAN2	YKT	46345363	46361060	AGCTTCACCGACTTTCGGAGT	0.02
RCAN2	YKT	46345363	46361060	AGCTTCACCGACTTGCGTATT	0.19
RCAN2	YKT	46345363	46361060	CATCCTGTTTGACCTTAGGGG	0.79
RCAN2	RUS	46345363	46361060	AGCTTCACCGACTTTCGGAGT	0.02
RCAN2	RUS	46345363	46361060	AGCTTCACCGACTTGCGTAGG	0.08
RCAN2	RUS	46345363	46361060	AGCTTCACCGACTTGCGTATT	0.42
RCAN2	RUS	46345363	46361060	CATCCTGTTTGACCTTAGGGG	0.48
HTR1F	YKT	87910930	87918173	CAAAGCTCGC	0.02
HTR1F	YKT	87910930	87918173	AAGGGTCCGC	0.02
HTR1F	YKT	87910930	87918173	AAGGATCCGC	0.04
HTR1F	YKT	87910930	87918173	CGAAGCTCGC	0.92
HTR1F	RUS	87910930	87918173	AAGGGTCCGC	0.10
HTR1F	RUS	87910930	87918173	AAGGATCCGC	0.10
HTR1F	RUS	87910930	87918173	AAGGGTCAAT	0.13
HTR1F	RUS	87910930	87918173	CGAAGCTCGC	0.67
THRB	YKT	24440630	24449645	TGCCACGTGCAAG	0.02
THRB	YKT	24440630	24449645	CAGCGTATGCAAG	0.02
THRB	YKT	24440630	24449645	TGCTACGCCCGGG	0.14
THRB	YKT	24440630	24449645	TGCCACGCCAAAA	0.22
THRB	YKT	24440630	24449645	TGGCATATGCAAG	0.6
THRB	RUS	24440630	24449645	TGGCGTATGCAAG	0.02
THRB	RUS	24440630	24449645	TGCCACGCCAAAA	0.04
THRB	RUS	24440630	24449645	TGCTACGCCCGGG	0.04
THRB	RUS	24440630	24449645	CAGCGTATGCAAG	0.26
THRB	RUS	24440630	24449645	TGGCATATGCAAG	0.64
SRC	YKT	37340671	37354179	TTAATTGTCTC	0.02
SRC	YKT	37340671	37354179	CCGACCAGCCA	0.12
SRC	YKT	37340671	37354179	TCGGCCAGCCC	0.86
SRC	RUS	37340671	37354179	TTGGCCAGTCC	0.02
SRC	RUS	37340671	37354179	CTAATTATTTA	0.02
SRC	RUS	37340671	37354179	CTAATTGTCCC	0.04
SRC	RUS	37340671	37354179	CCGACCAGCCA	0.10
SRC	RUS	37340671	37354179	CTAATTGTTTA	0.15
SRC	RUS	37340671	37354179	TCGGCCAGCCC	0.67
INSR	YKT	7169466	7178430	TGTCGGCCCCAGCT	0.02
INSR	YKT	7169466	7178430	TGTCGGCCCCAGCC	0.02
INSR	YKT	7169466	7178430	TGTCGGCCCCGATT	0.04
INSR	YKT	7169466	7178430	TGCCAGCCCCGATT	0.92

INSR	RUS	7169466	7178430	TGTCGGCCCCGGCC	0.02
INSR	RUS	7169466	7178430	CATAGATTTGGATT	0.12
INSR	RUS	7169466	7178430	TGTCGGCCCCAGCC	0.2
INSR	RUS	7169466	7178430	TGCCAGCCCCGATT	0.66
INSR	YKT	7166530	7169335	GACTTAGCCAC	0.02
INSR	YKT	7166530	7169335	GGAGCGAATGT	0.08
INSR	YKT	7166530	7169335	AACTTAGCCAC	0.9
INSR	RUS	7166530	7169335	GACTTAGCCAC	0.16
INSR	RUS	7166530	7169335	GGAGCGAATGT	0.22
INSR	RUS	7166530	7169335	AACTTAGCCAC	0.62

Supplementary Table S7. List of putative sepsis-related genes supported by the *Trendsetter* approach. Columns: **(1)** Gene name; **(2)** Chromosome; **(3)** Gene start; **(4)** Gene end; Number of windows falling in the gene and assigned to the hard and/or soft selective sweep classes according to the *Trendsetter* approach in **(5)** Yoruba; **(6)** CEPH; **(7)** Han Chinese and **(8)** Maya populations; **(9)** Reference studies supporting the functional link of the considered genes with sepsis.

Gene	Chr	Start	End	YRI	CEPH	HAN	MAY	Reference
CASP12	11	104885718	104898670	.	1	.	.	Saleh et al. 2006; Xue et al. 2006
CD79A	19	41877279	41881372	3	.	.	1	Lang et al. 2024
CTCF	16	67561831	67639189	.	10	14	1	Siegler et al. 2021
DCLRE1C	10	14897359	14954432	.	.	.	9	Ghadimi et al. 2023
LRBA	4	150264435	151015755	.	103	3	26	Burns et al. 2014
MAD1L1	7	1815787	2233243	.	14	4	.	Chen et al. 2024
RAG1	11	36510372	36593156	.	7	6	.	de Villartay et al. 2005
RAG2	11	36575574	36598279	.	1	1	.	de Villartay et al. 2005

RESEARCH ARTICLE

Archaic introgression contributed to shape the adaptive modulation of angiogenesis and cardiovascular traits in human high-altitude populations from the Himalayas

Giulia Ferraretti[1†], Paolo Abondio[2†], Marta Alberti[1], Agnese Dezi[3], Phurba T Sherpa[4], Paolo Cocco[5], Massimiliano Tiriticco[5], Marco Di Marcello[5], Guido Alberto Gnecchi-Ruscone[6], Luca Natali[5,7], Angela Corcelli[8], Giorgio Marinelli[5], Davide Peluzzi[5], Stefania Sarno[1†], Marco Sazzini[1,9*†]

[1]Laboratory of Molecular Anthropology and Centre for Genome Biology, Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy; [2]Department of Cultural Heritage, Ravenna Campus, University of Bologna, Bologna, Italy; [3]Department of Emergency and Organ Transplantation, University of Bari Aldo Moro, Bari Aldo Moro, Italy; [4]Mount Everest Summiters Club, Rolwaling, Dolakha, Nepal; [5]Explora Nunaat International, Montorio al Vomano, Teramo, Italy; [6]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; [7]Italian Institute of Human Paleontology, Rome, Italy; [8]Department of Basic Medical Science, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy; [9]Interdepartmental Centre Alma Mater Research Institute on Global Changes and Climate Change, University of Bologna, Bologna, Italy

***For correspondence:** marco.sazzini2@unibo.it

[†]These authors contributed equally to this work

Competing interest: The authors declare that no competing interests exist.

Funding: See page 16

Preprint posted
25 May 2023
Sent for Review
21 June 2023
Reviewed preprint posted
19 December 2023
Reviewed preprint revised
17 October 2024
Version of Record published
08 November 2024

Reviewing Editor: Emilia Huerta-Sanchez, Brown University, United States

eLife Assessment

This study presents **valuable** findings on what networks of genes were impacted by introgression from Denisovans, to identify the biological functions involved in high-altitude adaptation in Tibet. This study applies **solid** and previously validated methodology to identify genes with signatures of both introgression and positive selection. This paper would be of interest to population geneticists, anthropologists, and scientists studying the genetic basis underlying high-altitude adaptation.

Abstract It is well established that several *Homo sapiens* populations experienced admixture with extinct human species during their evolutionary history. Sometimes, such a gene flow could have played a role in modulating their capability to cope with a variety of selective pressures, thus resulting in archaic adaptive introgression events. A paradigmatic example of this evolutionary mechanism is offered by the *EPAS1* gene, whose most frequent haplotype in Himalayan highlanders was proved to reduce their susceptibility to chronic mountain sickness and to be introduced in the gene pool of their ancestors by admixture with Denisovans. In this study, we aimed at further expanding the investigation of the impact of archaic introgression on more complex adaptive responses to hypobaric hypoxia evolved by populations of Tibetan/Sherpa ancestry, which have been plausibly mediated by soft selective sweeps and/or polygenic adaptations rather

than by hard selective sweeps. For this purpose, we used a combination of composite-likelihood and gene network-based methods to detect adaptive loci in introgressed chromosomal segments from Tibetan WGS data and to shortlist those presenting Denisovan-like derived alleles that participate to the same functional pathways and are absent in populations of African ancestry, which are supposed to do not have experienced Denisovan admixture. According to this approach, we identified multiple genes putatively involved in archaic introgression events and that, especially as regards *TBC1D1*, *RASGRF2*, *PRKAG2*, and *KRAS*, have plausibly contributed to shape the adaptive modulation of angiogenesis and of certain cardiovascular traits in high-altitude Himalayan peoples. These findings provided unprecedented evidence about the complexity of the adaptive phenotype evolved by these human groups to cope with challenges imposed by hypobaric hypoxia, offering new insights into the tangled interplay of genetic determinants that mediates the physiological adjustments crucial for human adaptation to the high-altitude environment.

Introduction

The scientific community currently agrees that the *Homo sapiens* species experienced admixture with extinct Hominins since traces of such inbreeding events are still detectable in the genomes of modern humans (**Gouy and Excoffier, 2020**). In fact, people belonging to non-African populations show 1–2% of Neanderthal ancestry (**Green et al., 2010**; **Prüfer et al., 2014**), while Melanesians and East-Asians present 3% and 0.2% of Denisovan ancestry, respectively (**Reich et al., 2010**; **Meyer et al., 2013**; **Prüfer et al., 2014**; **Racimo et al., 2017**). Despite evidence supporting selection against introgressed alleles has been collected (**Simonti et al., 2016**; **Racimo et al., 2017**; **McArthur et al., 2021**), some of the genomic segments showing signatures ascribable to archaic introgression were also proved to have been targeted by natural selection in modern human populations, thus providing examples for the occurrence of adaptive introgression (AI) events (**Racimo et al., 2017**).

So far, several studies have indeed identified introgressed archaic alleles at high frequency in human genes involved in metabolism or in the response to environmental conditions, such as temperature, sunlight, and altitude (**Prüfer et al., 2014**; **Vernot and Akey, 2014**; **Sankararaman et al., 2014**; **Huerta-Sánchez et al., 2014**; **Gittelman et al., 2016**; **Racimo et al., 2017**; **Enard and Petrov, 2018**; **Dannemann and Racimo, 2018**). Moreover, some genes that play a role in immune responses to pathogens are found to be characterized by a similar pattern of variability (**Laurent et al., 2011**; **Enard and Petrov, 2018**) and certain Neanderthal alleles have been shown to be associated with down-regulation of gene expression in brain and testes (**McCoy et al., 2017**; **Racimo et al., 2017**; **Dannemann and Racimo, 2018**). These works collectively attest how genetic variants introduced in the human gene pool by admixture with archaic species can significantly impact our biology by possibly comporting modifications in the modulation of several functional pathways. In particular, the high frequency of some archaic alleles in protein-coding and/or regulatory genomic regions suggests a possible adaptive role for Neanderthal and/or Denisovan variants, pointing to a further evolutionary mechanism having potentially contributed to the processes of human biological adaptation to different environmental and cultural settings. By introducing new alleles in the gene pool of a given population, admixture in fact provides a very rapid opportunity for natural selection to act on it (**Huerta-Sánchez et al., 2014**; **Jeong et al., 2014**; **Racimo et al., 2015**; **Hamid et al., 2021**) and this is supposed to have likely occurred during the evolutionary history of *H. sapiens*, particularly after the last Out of Africa migration in the late Pleistocene (**Sugden, 2018**; **Vahdati et al., 2022**). According to this view, gene flow from extinct Hominin species could have facilitated the adaptation of *H. sapiens* populations to peculiar Eurasian environments.

For instance, a Denisovan origin of the adaptive *EPAS1* haplotype, which confers reduced susceptibility to chronic mountain sickness to Tibetan and Sherpa highlanders (**Beall, 2007**; **Bigham et al., 2010**; **Yi et al., 2010**; **Peng et al., 2011**; **Xu et al., 2011**) is well established (**Huerta-Sánchez et al., 2014**; **Zhang et al., 2021**). However, the hard selective sweep experienced in high-altitude Himalayan populations by the *EPAS1* introgressed haplotype has been demonstrated to account only for an indirect aspect of their adaptive phenotype, which does not explain most of the physiological adjustments they evolved to cope with hypobaric hypoxia (**Gnecchi-Ruscone et al., 2018**). Therefore, how far gene flow between Denisovans and the ancestors of Tibetan/Sherpa peoples facilitated the evolution of other key adaptive traits of these populations remains to be elucidated.

To fill this gap, and to overcome the main limitation of most approaches currently used to test for AI (i.e., inferring archaic introgression and the action of natural selection separately by means of different algorithms, which increases the risk of obtaining biased results due to confounding variables), we assembled a dataset of whole-genome sequences (WGSs) from 27 individuals of Tibetan ancestry living at high altitude (***Cho et al., 2017***; ***Jeong et al., 2018***) and we analysed it using a composite-likelihood method specifically developed to detect AI events at once (***Setter et al., 2020***). Notably, this method was designed to recognize AI mediated by subtle selective events (as those involved in polygenic adaptation) and/or soft selective sweeps, which represent the evolutionary mechanisms that are supposed to have played a more relevant role than hard selective sweeps during the adaptive history of human groups characterized by particularly small effective population size, such as Tibetans and Sherpa (***Gnecchi-Ruscone et al., 2018***). Coupled with validation of the identified putative adaptive introgressed loci through (1) the assessment of the composition of gene networks made up of functionally related DNA segments presenting archaic derived alleles that are absent in human groups which are supposed to do not have experienced Denisovan admixture, such as African ones, (2) the confirmation that natural selection targeted these genomic regions in populations of Tibetan ancestry, and (3) the quantification of genetic distance between modern and archaic haplotypes, such an approach provided new evidence about the biological functions that have mediated high-altitude adaptation in Himalayan populations and that have been favourably shaped by admixture of their ancestors with Denisovans.

Results

Spatial distribution of genomic variation and ancestry components of Tibetan samples

After quality control (QC) filtering of the available WGS data, we obtained a dataset made up of 27 individuals of Tibetan ancestry characterized for 6,921,628 single-nucleotide variants (SNVs). To assess whether this dataset represents a reliable proxy for the genomic variation observable in high-altitude Himalayan populations, we merged it with genome-wide genotyping data for 1086 individuals of East-Asian ancestry belonging to both low- and high-altitude groups (***Gnecchi-Ruscone et al., 2017***; ***Landini et al., 2021***). We thus obtained an extended dataset including 231,947 SNVs (***Supplementary file 1a***), which was used to perform population structure analyses.

Results from ADMIXTURE and principal components analysis (PCA) were found to be concordant with those described in previous studies (***Jeong et al., 2014***; ***Gnecchi-Ruscone et al., 2017***; ***Gnecchi-Ruscone et al., 2018***; ***Yang et al., 2021***). According to the ADMIXTURE model showing the best predictive accuracy ($K = 7$) (***Figure 1—figure supplement 1***), the examined WGS exhibited a predominant genetic component that was appreciably represented also in other populations speaking Tibeto-Burman languages, such as Tu, Yizu, Naxi, Lahu, and Sherpa (***Figure 1A*** and ***Figure 1—figure supplement 2***). Such a component reached an average proportion of around 78% in individuals of Tibetan ancestry from Nepal included in the extended dataset, as well as of more than 80% in the subjects under investigation, who live in the Nepalese regions of Mustang and Ghorka (***Figure 1A, B***). This suggests that after their relatively recent migration in Nepalese high-altitude valleys, these communities might have experienced a higher degree of isolation and genetic drift with respect to populations that are still settled on the Tibetan Plateau, in which the same ancestry fraction did not exceed 64% (***Figure 1A, B***). Nevertheless, the overall ADMIXTURE profile of the considered WGS appeared to be quite comparable to those inferred according to genome-wide genotyping data for other Tibetan populations (***Figure 1A, B***, ***Figure 1—figure supplement 2***). Similarly, PCA pointed to the expected divergence of Tibetan and Sherpa high-altitude groups from the cline of genomic variation of East-Asian lowland populations (***Abdulla et al., 2009***; ***Jeong et al., 2014***; ***Gnecchi-Ruscone et al., 2017***; ***Zhang et al., 2017***; ***Wang et al., 2022***; ***Figure 1C***). Remarkably, the WGS under investigation clustered within the bulk of genome-wide data generated for other groups from the Tibetan Plateau, thus supporting their representativeness as concerns the overall genetic background of high-altitude Himalayan populations.

Detecting putative AI signatures in Tibetan genomes

To identify genomic regions showing signatures putatively ascribable to AI events, we scanned Tibetan WGS with the *VolcanoFinder* algorithm and we computed the composite likelihood ratio (LR)

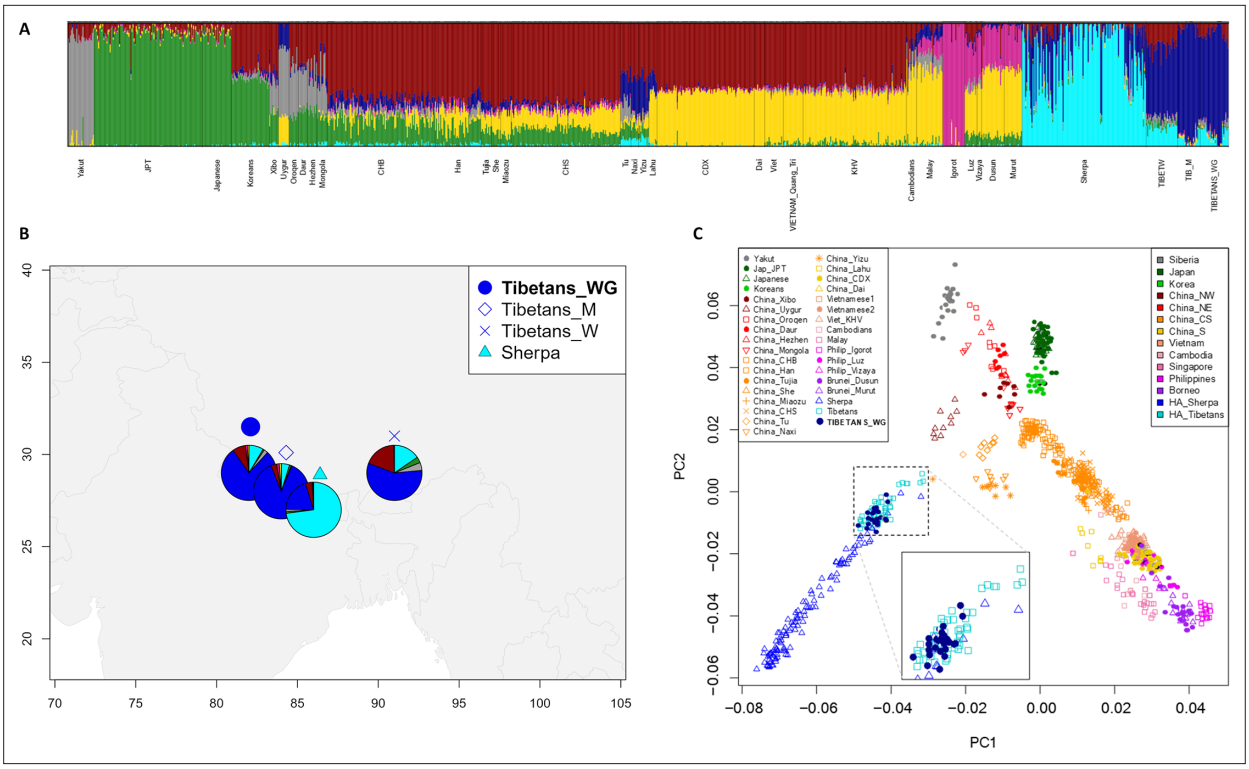


Figure 1. Population structure analyses performed on the extended dataset including Tibetan, Sherpa, and lowland East-Asian individuals. (**A**) Admixture analysis showed the best predictive accuracy when seven ($K = 7$) population clusters were tested. Populations included in the dataset are labelled according to population names and acronyms reported in ***Supplementary file 1a***. (**B**) Map showing geographic location and admixture proportions at $K = 7$ of the high-altitude groups included in the extended dataset. The label *Tibetans_WG* indicates whole-genome sequence data for individuals of Tibetan ancestry analysed in the present study. Additional information about the considered samples (e.g., number of individuals per group, reference study, and used abbreviations) are reported in ***Supplementary file 1a***. (**C**) Principal components analysis (PCA) plot considering PC1 vs PC2 and summarizing genomic divergence between high-altitude Tibetan/Sherpa people and the cline of variation observable for lowland East-Asian populations. The enlarged square displays clustering between Tibetan samples sequenced for the whole genome (i.e., blue dots) and Tibetan samples characterized by genome-wide data (i.e., light-blue squares).

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Scatterplot showing the number of possible population clusters (*K*) tested by the different ADMIXTURE runs performed and the cross-validation (CV) errors associated to them.

Figure supplement 2. Admixture analyses performed on the extended dataset for $K = 2$ to $K = 12$.

and $-\log\alpha$ statistics for each polymorphic site (***Setter et al., 2020***). We then considered the most significant results by focusing on loci showing LR values falling in the positive tail (i.e., top 5%) of the obtained distribution (see Materials and methods, ***Supplementary file 1b***).

According to such an approach, we were first able to recapitulate the AI event previously described for the *EPAS1* gene (***Figure 2—figure supplement 1A***; ***Huerta-Sánchez et al., 2014***; ***Hu et al., 2017***; ***Zhang et al., 2021***). In fact, this chromosomal interval was found to be characterized by a remarkable number of variants ($N = 19$) showing significant LR scores, as well as by high overall values of $-\log\alpha$ (***Figure 2—figure supplement 1A***), suggesting, respectively, the plausible archaic origin of many alleles at this gene and an appreciable action of natural selection on it. Five of these significant SNVs have been already described as Denisovan-like derived alleles at outstanding frequency (i.e., ranging between 0.96 and 1) in Tibetans but not in other modern human populations (***Supplementary file 1c***; ***Zhang et al., 2021***). On the contrary, at the genomic region encompassing *EGLN1* (which we have considered as a negative control for AI, see Materials and methods) we detected high $-\log\alpha$ values coupled with a low number of SNVs ($N = 3$) showing significant LR scores, with only one being remarkably above the adopted significance threshold (***Figure 2—figure supplement 1B*** and ***Supplementary file 1d***). Overall, these findings are concordant with evidence from literature that suggest adaptive evolution of both *EPAS1* and *EGLN1* loci in high-altitude Himalayan populations (***Yang et al.,***

2017; *Liu et al., 2022*), although only the former was proved to have been impacted by archaic introgression (*Huerta-Sánchez et al., 2014*; *Hu et al., 2017*; *Zhang et al., 2021*).

Moreover, we were able to confirm other introgression signatures previously inferred from WGS data for populations of Tibetan ancestry, such as those involving the *PRKCE* gene and the *MIRLET7BHG* long non-coding region, which are located in the overlapping upstream chromosomal intervals, respectively, of *EPAS1* and *PPARA* (*Figure 2—figure supplement 2A, B*). In line with what reported for *EPAS1*, also *PPARA* has been already proposed to play a role in the modulation of high-altitude adaptation of Himalayan human groups (*Simonson et al., 2010*; *Horscroft et al., 2017*; *Zhang et al., 2021*). Interestingly, AI signatures identified by *VolcanoFinder* in the *MIRLET7BHG* locus extended also in the *PPARA* gene, as well in its downstream region (*Figure 2—figure supplement 2A*), supporting the findings described by *Hu et al., 2017*. Finally, we observed patterns comparable to those at *EPAS1* and *PPARA* for 10 additional genomic regions that were differentially pointed out by previous studies as Tibetan and/or Han Chinese DNA segments potentially carrying introgressed Denisovan alleles (*Hu et al., 2017*; *Browning et al., 2018*; *Zhang et al., 2021*; *Figure 2A, B*, *Figure 2—figure supplement 3A*, *Supplementary file 1e*).

Validating genomic regions affected by archaic introgression

To validate signatures of archaic introgression at the candidate AI loci identified with *VolcanoFinder*, we relied on the approach described by *Gouy and Excoffier, 2020*. In detail, we used the *Signet* algorithm to identify networks of genes participating to the same functional pathway and presenting archaic-like (i.e., Denisovan) derived alleles observable in the putative admixed group (i.e., Tibetans) but not in an outgroup of African ancestry (i.e., Yoruba, YRI), by assuming that only Eurasian *H. sapiens* populations experienced Denisovan admixture (see Materials and methods).

After having crosschecked results from the *VolcanoFinder* and *Signet* analyses, we identified six gene networks that turned out to be consistently significant across all the *Signet* runs performed and that included a total of 15 genes pointed by *VolcanoFinder* as candidate AI loci (*Supplementary file 1f*). Four of these loci composed the gene network overall ascribable to the *Pathways in Cancer* biological functions, which included also the *EPAS1* positive control for AI (*Figure 3A*). Most of the other genes supported by both the analyses were instead observed in significant networks belonging to the *Ras signalling* and *AMPK signalling* pathways (*Supplementary file 1f*).

Interestingly, gene networks belonging to the *Pathways in Cancer* and *Ras signalling pathway* appeared to be tightly related from a functional perspective because included oncogenes that promote the initiation and progression of tumour growth by stimulating cell proliferation and angiogenesis (*Kranenburg et al., 2004*). In particular, according to the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database *KRAS* and *RASGRF2* genes from the *Ras signalling* network were found to contribute also to the *Pathways in Cancer* functions (*Figure 3A*), especially by interacting with the identified candidate introgressed loci *PLCB1*, *RASGRP2*, *DAPK1*, *MAPK1*, *FOS*, and *VEGFA* to modulate the *VEGF signalling pathway*, which is activated in hypoxic conditions and induces the transcription of genes that promote angiogenesis (*Figure 3A*; *Maxwell and Ratcliffe, 2002*; *Kranenburg et al., 2004*).

Also, the *AMPK signalling pathway* is known to be activated in different cell types by stresses such as deprivation of oxygen and/or glucose, leading to the inhibition of energy-consuming biosynthetic pathways (e.g., protein and glycogen synthesis) and to the activation of ATP-producing catabolic pathways, such as fatty acid oxidation and glycolysis (*Kanehisa and Goto, 2000*; *Chen et al., 2018*; *Dengler, 2020*).

No significant gene networks involving the *EGLN1* genomic region considered as a negative control for AI were instead reconstructed with the *Signet* approach (*Supplementary file 1f*), suggesting that the very few variants at this locus that showed *VolcanoFinder* LR scores above the adopted significant threshold might represent false positive results (*Figure 2—figure supplement 1B* and *Supplementary file 1d*).

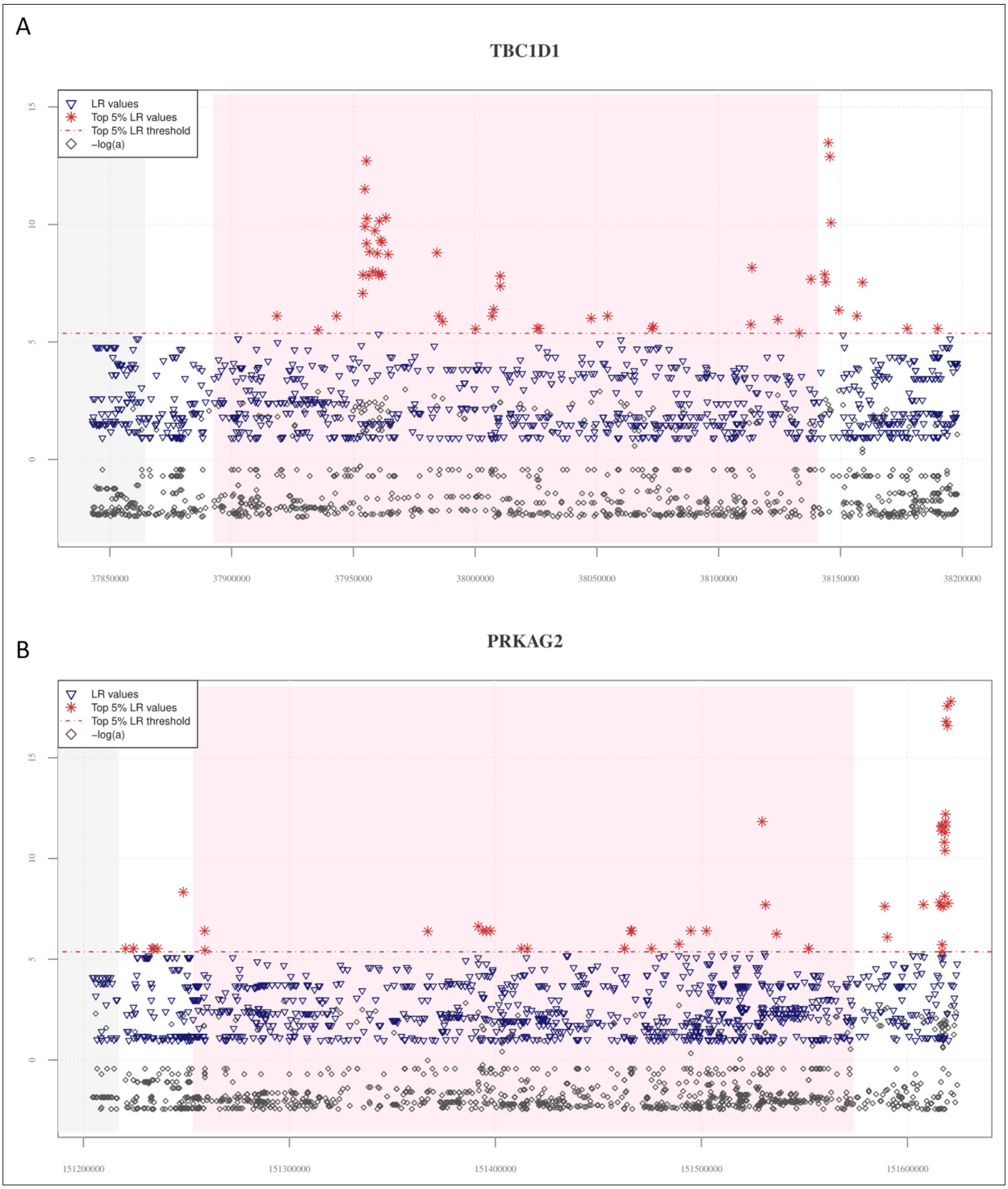


Figure 2. Distribution of *VolcanoFinder* statistics suggestive of putative adaptive introgressed loci across the *TBC1D1* and *PRKAG2* genomic regions. On the *x*-axis are reported genomic positions of each single-nucleotide variant (SNV), while on the *y*-axis are displayed the related statistics obtained. Pink background indicates the chromosomal interval occupied by the considered genes, while the grey background identifies those genes (i.e., *PGM2* in the *TBC1D1* downstream genomic region and the *RHEB* gene in the upstream *PRKAG2* region) possibly involved in regulatory transcription mechanisms. The dashed red line identifies the threshold set to filter for significant likelihood ratio (LR) values (i.e., top 5% of LR values). For both these genomic regions, the distribution of LR and −logα are concordant with those observed at the *EPAS1* positive control for adaptive introgression (AI). (**A**) A total of 50 significant LR values (red stars) and −logα (grey diamonds) values resulted collectively elevated in both the *TBC1D1* gene and its downstream genomic regions. A remarkable concentration of significant LR values characterizing 19 SNVs was especially observable in the first portion

Figure 2 continued on next page

Figure 2 continued

of the gene. (**B**) The entire *PRKAG2* genomic region was found to comprise 46 SNVs showing significant LR values, with the greatest peaks being located in the downstream region associated to such gene. Peaks detected for the LR statistic are accompanied by peaks of −logα values.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Distribution of *VolcanoFinder* statistics across the *EPAS1* and *EGLN1* positive and negative controls for adaptive introgression.

Figure supplement 2. Distribution of *VolcanoFinder* statistics across *MIRLET7BHG*, *PPARA*, and *PRKCE* genes.

Figure supplement 3. Distribution of *VolcanoFinder* statistics across the *RASGRF2* candidate adaptive introgression (AI) gene.

Figure supplement 4. Distribution of *VolcanoFinder* statistics across the *KRAS* candidate adaptive introgression (AI) gene.

Shortlisting introgressed genomic regions characterized by adaptive evolution

To further shortlist the most robust candidate genes involved in AI events, we applied the *LASSI* algorithm to phased Tibetan WGS data with the aim of searching for genomic signatures ascribable to the action of natural selection (***Harris and DeGiorgio, 2020***) (see Materials and methods).

This enabled us to confirm the strong selective events occurred at the *EPAS1* and *EGLN1* genes, as previously reported by multiple studies conducted on high-altitude Himalayan populations (***Beall et al., 2010***; ***Yi et al., 2010***; ***Simonson et al., 2010***; ***Horscroft et al., 2017***; ***Zhang et al., 2021***), as well as to corroborate adaptive evolution of some of the genes pointed out by both *VolcanoFinder* and *Signet* analyses. In fact, several chromosomal intervals associated to these loci presented values of the computed *T* statistic that fall within the top 5% of the related distribution (***Figure 4C, D***, ***Figure 4—figure supplements 1 and 2C, D***, ***Figure 4—figure supplement 3C, D***).

More in detail, in addition to *EPAS1*, genomic windows associated to the *DAPK1*, *GNG7*, *AK5*, *TBC1D1*, *PLCB1*, *RASGRF2*, and *PRKAG2* introgressed loci supported by both *VolcanoFinder* and *Signet* approaches were found to present scores within the top 5% of the *T* distribution, suggesting that their haplotype diversity was appreciably shaped by positive selection. Interestingly, adaptive evolution of the *TBC1D1* and *RASGRF2* genes has been previously proposed by studies conducted on different populations of Tibetan ancestry (***Peng et al., 2011***; ***Zheng et al., 2023***).

Estimating genetic distance between modern and archaic sequences

As a final step for prioritizing the most convincing AI genes supported by *VolcanoFinder*, *Signet*, and *LASSI* approaches, as well as to explicitly test whether the Denisovan human species represented a plausible source of archaic alleles for them, we merged Tibetan WGS data with those from low-altitude Han Chinese (CHB) and YRI populations sequenced by the 1000 Genomes Project (***Auton et al., 2015***), and with the Denisovan genome. We then used the *Haplostrip* algorithm (***Marnetto and Huerta-Sánchez, 2017***) to estimate genetic distance between modern and archaic haplotypes at the candidate AI genes reported in the previous paragraph. We especially considered genomic windows that included Denisovan-like derived alleles and that presented values of the likelihood *T* statistic supporting an adaptive evolution (see Materials and methods).

Among the tested putative AI loci, *TBC1D1*, *RASGRF2*, *PRKAG2*, and *KRAS* were found to present substantial proportions of Tibetan haplotypes that cluster close to the Denisovan sequence, thus showing the lowest numbers of pairwise differences with respect to it as compared with CHB or YRI haplotypes (***Figure 4A, B***, ***Figure 4—figure supplement 2A, B***). More in detail, 61% of the *TBC1D1* Tibetan haplotypes turned out to be the nearest ones to the archaic sequence by entailing only two pairwise differences with respect to it (***Figure 4A***), while 29% of Tibetan haplotypes inferred for the *RASGRF2* gene were even identical to the Denisovan DNA (***Figure 4B***). Interestingly, both these chromosomal intervals were classified by the *LASSI* method as regions whose variation pattern was conformed with the soft selective sweep model, presenting three potential adaptive haplotypes (i.e., those haplotypes that plausibly carry putative advantageous alleles and thus increased in frequency due to positive selection). At each gene, one of these haplotypes was found to contain the Denisovan-like derived alleles that are completely absent in YRI (***Figure 4A, D***). A similar pattern was observed also for the considered *PRKAG2* and *KRAS* genomic windows (***Figure 4—figure supplement 2A, B***).

Overall, the distribution of similarities between modern and archaic haplotypes described for the four identified AI candidate loci appears to be comparable to that obtained for *EPAS1*, with the sole relevant distinction being represented by an even more pronounced differentiation between Tibetan

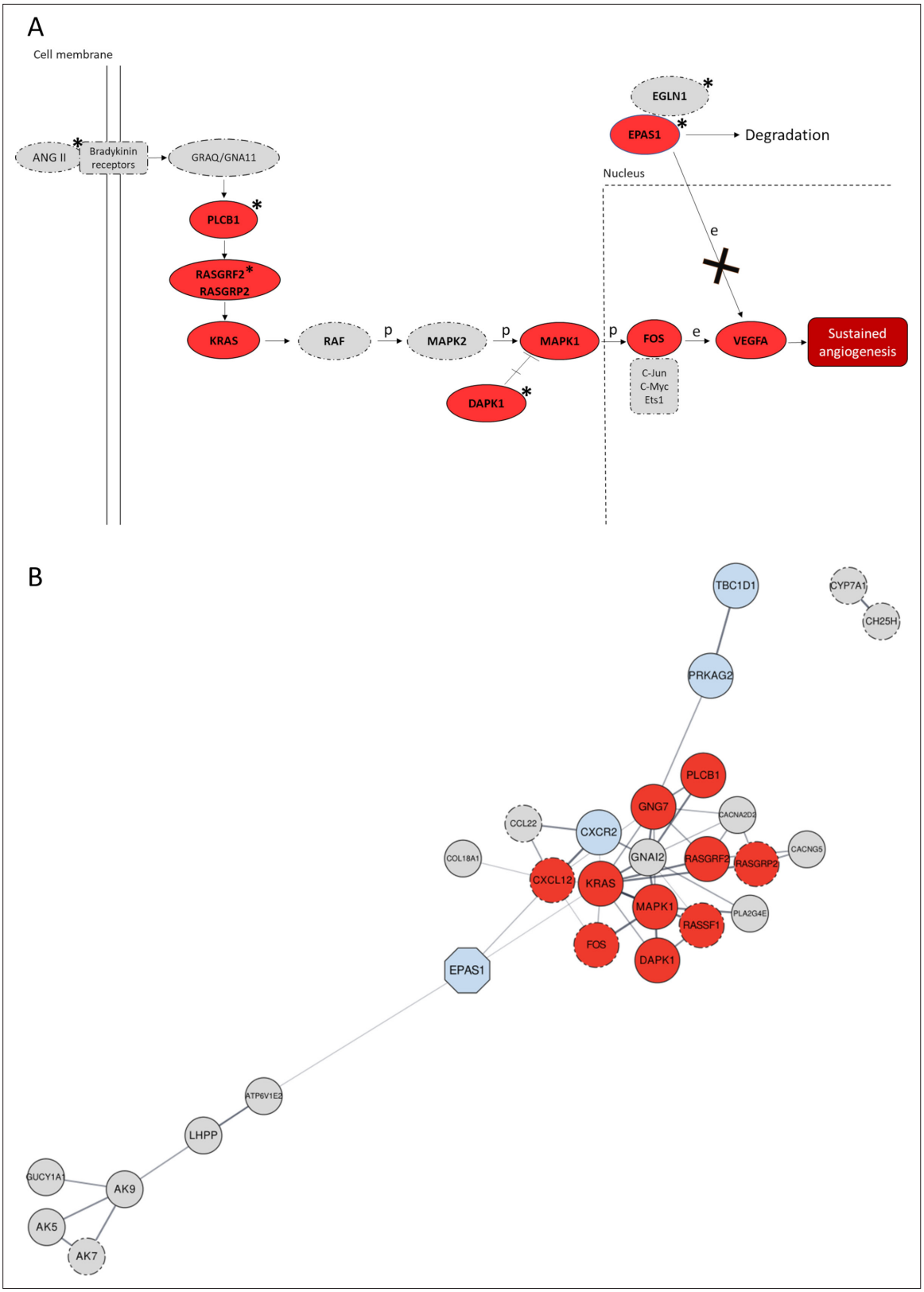


Figure 3. Significant gene networks including Denisovan-like derived alleles according to the *Signet* analysis. (**A**) Schematic representation of the activation of the *RAS*/*MAPK*(*ERK*) axis after interaction of the bradykinin receptors with their ligands (e.g., ANG II) within the framework of the *Pathways in Cancer* network. Genes supported by both *Signet* (i.e., belonging to the significant network associated to *Pathways in cancer*) and *VolcanoFinder* (i.e., including at least a single-nucleotide variant (SNV) showing likelihood ratio (LR) value within top 5% of the obtained results) analyses as potentially

Figure 3 continued on next page

Figure 3 continued

introgressed loci, are highlighted in red and present solid outline. Grey circles with dotted-dashed contour instead indicate genes supported only by *Signet*, while loci marked with stars are those including genomic windows showing *LASSI T* statistic within top 5% of the related distribution. After the interaction between ANG II (active enzyme angiotensin II) and bradykinin receptors, activation of the Ras protein encoded by *KRAS* mediated by RAS-GTPases (e.g., *RASGRF2*) comports a series of phosphorylation reactions that eventually promotes angiogenesis (***Kranenburg et al., 2004***). In detail, phosphorylation of the MAPK1 protein and prevention of MAPK1-DAPK-1-dependent apoptosis leads to increased MAPK1 activity (***Kanehisa and Goto, 2000***; ***Stevens et al., 2007***) that causes improved *FOS* mRNA expression (***Monje et al., 2005***). *FOS* together with other proteins (e.g., Jun) forms the AP-1 transcription factor, which bounds to the *VEGF* promoter region upregulating its expression in endothelial cells (***Catar et al., 2013***) and sustaining angiogenesis when the hypoxia inducible factor 1 (HIF-1) signalling cascade is inhibited (***Lorenzo et al., 2014***). (**B**) Gene network built by setting co-expression as force function and by displaying the entire set of genes identified by the *Signet* algorithm as belonging to significant pathways including Denisovan-like derived alleles. Genes whose variation pattern was supported by both *VolcanoFinder* and *Signet* analyses (e.g., *TBC1D1*) as shaped by archaic introgression are displayed with a solid black outline. The *EPAS1* positive control locus that has been previously proved to have mediated adaptive introgression in Tibetan populations was represented as light-blue octagonal. Genes included in pathways involved in angiogenesis (e.g., *RASGRF2*) and/or activated in hypoxic conditions (e.g., *PRKAG2*) are reported as dark red and light-blue circles, respectively, while the remining fraction of significant genes are represented as light-grey circles. The closeness or the distance between all nodes reflects the tendency to be co-expressed with each other and all the connections inferred are characterized by a confident score ≥0.7.

and CHB patterns plausibly ascribable to the occurrence of a hard (rather than soft) selective sweep at *EPAS1* (***Figure 4—figure supplement 3A***), as previously proposed (***Simonson et al., 2010***; ***Huerta-Sánchez et al., 2014***). In fact, at *EPAS1* the haplotype carrying the lowest number of pairwise differences (N = 3) with respect to the archaic one belongs to the Tibetan population, in which it reached 74% of frequency, being instead absent in all the others modern groups considered (***Figure 4—figure supplement 3A***). Conversely, the *EGLN1* genomic window showing the highest *T* score according to *LASSI* analysis was characterized by an opposite pattern, with CHB haplotypes being overrepresented among those with the lowest number of pairwise differences (N = 3) with respect to the Denisovan genome (***Figure 4—figure supplement 3B***). Moreover, the sole *EGLN1* putative adaptive haplotype inferred by *LASSI* for Tibetans was among those presenting the highest number of differences as compared with the archaic sequence, being characterized exclusively by alleles that are not observed in the Denisovan genome and thus suggesting that natural selection targeted modern rather than archaic *EGLN1* variation (***Figure 4—figure supplement 3B***).

Discussion

In the present study, we aimed at investigating how far gene flow between the Denisovan archaic human species and the ancestors of modern populations settled in high-altitude regions of the Himalayas contributed to the evolution of key adaptive traits of these human groups, in addition to having conferred them reduced susceptibility to chronic mountain sickness (***Huerta-Sánchez et al., 2014***). For this purpose, we used WGS data from individuals of Tibetan ancestry to search for genomic signatures ascribable to AI mediated by weak selective events rather than by hard selective sweeps, under the assumption that soft sweeps and/or processes of polygenic adaptation are more likely to have occurred in such remarkably isolated and small effective population size groups (***Gnecchi-Ruscone et al., 2018***).

By assembling a large genome-wide dataset including both low- and high-altitude populations, we first framed the available WGS data into the landscape of East-Asian genomic variation. This confirmed that the genomes under investigation are well representative of the overall profiles of ancestry components observable in high-altitude Himalayan peoples (***Figure 1A, B***). In fact, the considered individuals were found to show close genetic similarity to other populations of Tibetan ancestry (***Jeong et al., 2014***) and to Sherpa people from Nepal (***Gnecchi-Ruscone et al., 2017***), as well as to appreciably diverge from the cline of variation of lowland East-Asians (***Abdulla et al., 2009***; ***Figure 1C***).

Based on this evidence, we submitted Tibetan WGS to a pipeline of analyses that implemented multiple independent approaches aimed at identifying genomic regions characterized by signatures putatively ascribable to AI events and by tight functional correlations with each other. According to such a rationale, we shortlisted the candidate introgressed loci that most likely contribute to the same adaptive trait by searching for chromosomal intervals including loci simultaneously showing: (1) significant LR scores computed by the *VolcanoFinder* algorithm (***Figure 2A, B***, ***Figure 2—figure***

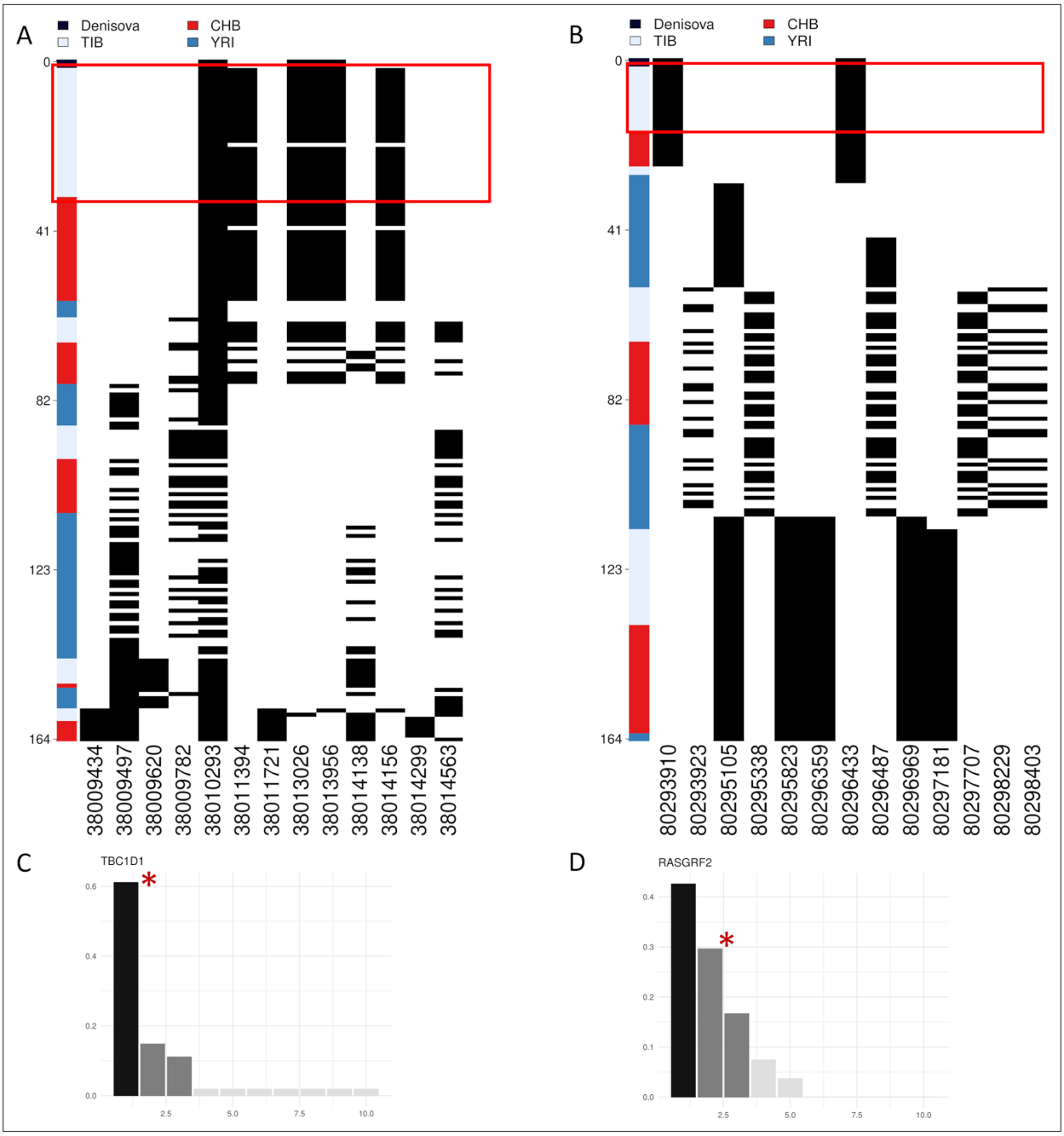


Figure 4. Representation of genetic distances between modern and archaic haplotypes and barplots showing haplotype frequency spectra for *TBC1D1* and *RASGRF2* candidate adaptive introgression (AI) genes. Haplotypes are reported in rows, while derived (i.e., black square) and ancestral (i.e., white square) alleles are displayed in columns. Haplotypes are ranked from top to bottom according to their number of pairwise differences with respect to the Denisovan sequence. (**A**) Heatmap displaying divergence between Tibetan, CHB and YRI *TBC1D1* haplotypes with respect to the Denisovan genome. A total of 33 *TBC1D1* haplotypes (i.e., 61% of the overall haplotypes inferred for such a region) belonging to individuals with Tibetan ancestry are plotted in the upper part of the heatmap thus presenting the smallest number of pairwise differences with respect to the Denisovan sequence. (**B**) Heatmap displaying divergence between Tibetan, CHB and YRI *RASGRF2* haplotypes with respect to the Denisovan genome. A total of 16 Tibetan haplotypes in the *RASGRF2* genomic region present no differences with respect to the Denisovan sequence. As regards barplots, on the *x*-axis are reported the haplotypes detected in the considered genomic windows, while on the *y*-axis is indicated the frequency for each haplotype. The black and dark-grey bars indicate the more frequent haplotypes (i.e., the putative adaptive haplotypes inferred by the *LASSI* method), while red stars mark those haplotypes carrying Denisovan-like derived alleles. (**C**) *TBC1D1* haplotype frequency spectrum. The *TBC1D1* gene presents a haplotype pattern qualitatively comparable to that observed at *EPAS1* (***Figure 4—figure supplement 3A***), with a predominant haplotype carrying archaic derived alleles and reaching elevated frequencies in Tibetan populations. In line with this observation, such a pattern was inferred by *LASSI* as conformed with a non-neutral evolutionary scenario, even if it seems to be characterized by a soft rather than a hard selective sweep due to the occurrence of three sweeping

Figure 4 continued on next page

Figure 4 continued

haplotypes. (**D**) *RASGRF2* haplotype frequency spectrum. A soft selective sweep was inferred also for the considered *RASGRF2* genomic window, although frequencies reached by the sweeping haplotypes turned out to be more similar with each other. The second most represented haplotype was that carrying the archaic derived alleles and, reached a frequency of 29% in the Tibetan group.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Haplotype frequency spectra of the top windows detected as adaptively evolved by *LASSI* in the *EPAS1* and *EGLN1* genomic regions.

Figure supplement 2. Representation of genetic distances between modern and archaic haplotypes.

Figure supplement 3. Representation of genetic distances between modern and archaic haplotypes.

supplements 3 and 4), (2) Denisovan-like derived alleles belonging to significant networks of functionally related genes reconstructed with the *Signet* method and completely absent in populations of African ancestry (***Figure 3A, B***, ***Supplementary file 1f***), (3) signatures ascribable to the action of natural selection as pointed out by *LASSI* analysis (***Figure 4C, D***, ***Figure 4—figure supplement 2C, D***), and (4) haplotypes more similar to the Denisovan ones rather than to those observed in other modern human populations, as depicted by the *Haplostrips* approach (***Figure 4A, B***; ***Figure 4—figure supplement 2A, B***).

Overall, in addition to *EPAS1*, which we considered as a positive control for AI, a total of 18 genes encompassed within the putative AI chromosomal intervals identified by the *VolcanoFinder* method (***Supplementary file 1e***) were found to have been previously proposed as genomic regions impacted by introgression of Denisovan alleles in Tibetan and/or Han Chinese populations (***Huerta-Sánchez et al., 2014***; ***Hu et al., 2017***; ***Browning et al., 2018***; ***Yang et al., 2017***; ***Zhang et al., 2021***). Among them, *PPARA*, *PRKCE*, and *TBC1D1* (***Figure 2A***, ***Figure 2—figure supplement 2A, B***) were also specifically suggested to have played an adaptive role in high-altitude groups from the Tibetan Plateau (***Simonson et al., 2010***; ***Peng et al., 2011***; ***Horscroft et al., 2017***; ***Arciero et al., 2018***; ***Deng et al., 2019***; ***Zhang et al., 2021***; ***Zheng et al., 2023***). Interestingly, *PPARA* encodes for a nuclear transcription factor whose decreased expression in the myocardium of rats exposed to hypoxia seems to contribute to the maintenance of the correct heart contractile function despite such a stressful condition (***Cole et al., 2016***). Similarly, the PRKCE protein kinase C has been demonstrated to exert a cardio-protective role against ischemic injury (***Scruggs et al., 2016***). Moreover, *TBC1D1* encodes for a protein whose serine phosphorylation sites are targeted by AMP-activated protein kinases (AMPK) after the activation of the *AMPK signalling pathway* as a result of the increase cellular AMP/ATP ratio caused by stresses that induce a lower ATP production (e.g., deprivation of oxygen and/or glucose) or that accelerate ATP consumption (e.g., intense muscle contraction) (***Kanehisa and Goto, 2000***; ***Vichaiwong et al., 2010***). In addition, another member of the *AMPK signalling pathway*, *PRKAG2*, was suggested by both *VolcanoFinder* analysis and literature data to present putative introgressed Denisovan alleles in Tibetan populations (***Figure 2B***; ***Zhang et al., 2021***). Mutations at this locus are known to cause the *PRKAG2 cardiac syndrome*, an inherited disease characterized by ventricular pre-excitation, supraventricular arrhythmias, and cardiac hypertrophy (***Zhang et al., 2013***; ***Porto et al., 2016***). Dysregulation of AMPK activity mediated by reduction in *PRKAG2* expression and leading to the impairment of glycogen metabolism in cell cultures has been proposed as a possible cause for the development of this pathological condition (***Zhang et al., 2013***). Conversely, enhanced activation of the *AMPK signalling pathway* during pregnancy coupled with *PRKAG2* overexpression was observed in the placenta of women living at high altitudes when compared with women living in low-altitude regions (***Lorca et al., 2021***). Finally, high-altitude individuals of Tibetan ancestry were found to exhibit reduced incidence of major adverse cardiovascular events with respect to low-altitude controls possibly indicating the involvement of protective cardiac mechanisms in the modulation of high-altitude adaptations as previously proposed (***Kolár and Ostádal, 2004***; ***Mallet et al., 2018***; ***Lei et al., 2024***). We can thus speculate that adaptive evolution at the *PPARA*, *PRKCE*, *TBC1D1*, and *PRKAG2* genomic regions in Tibetans might have contributed to the development of protective mechanisms that reduce cardiovascular risk associated to the hypoxic stress.

In addition to these genes, other two of the identified candidate AI loci (i.e., *RASGRF2* and *KRAS*) have been previously proved to have been targeted by natural selection in Tibetan populations (***Peng et al., 2011***) or to present putative introgressed archaic alleles (***Hu et al., 2017***; ***Browning et al.,***

2018; **Figure 2—figure supplements 3 and 4**). The proteins encoded by these loci are strictly related from a functional perspective, with the Ras protein specific guanine nucleotide releasing factor 2 representing a calcium-regulated nucleotide exchange factor that activates the RAS protein codified by the proto-oncogene *KRAS* (**Kanehisa and Goto, 2000**; **Sayers et al., 2022**).

Despite evidence reported in literature seem to corroborate *VolcanoFinder* results that indicate *PPARA* and *PRKCE* as putatively implicated in AI events experienced by Tibetan ancestors, only *EPAS1*, *TBC1D1*, *RASGRF2*, *PRKAG2*, and *KRAS* loci were finally retained after the adopted filtering procedure to represent the most reliable candidate AI loci.

Archaic introgression at these genomic regions was first confirmed by *Signet* analysis, with *TBC1D1* and *PRKAG2* being included in a significant gene network belonging to the *AMPK signalling pathway*, while *RASGRF2* and *KRAS* participate to that related to the *Ras signalling pathway* (**Supplementary file 1f**). The same analysis pointed to *EPAS1* as a member of a significant network belonging to the *Pathways in cancer*, a complex group of biological functions such as those involved in Ras, MAPK, VEGF, and HIF-1 signalling cascades (**Kanehisa and Goto, 2000**; **Figure 3A** and **Supplementary file 1f**). These findings emphasize a link between biological mechanisms activated within the context of hypoxic tumour microenvironments in different types of cancers and those involved in high-altitude adaptations, especially as concerns functions that might underlie the improved angiogenesis observed in Tibetan and Sherpa individuals (**Gnecchi-Ruscone et al., 2018**; **Calderón - Gerstein and Torres - Samaniego, 2021**). For instance, accumulation of the HIF-1α transcriptional factor in the nucleus of cells close to hypoxic tumour masses comports the activation of diverse biological responses such as the formation of dense capillary structures that permit oxygen and nutrients supplies to cancer cells, thus determining tumour progression and/or treatment failure (**Brahimi-Horn et al., 2007**; **Calderón - Gerstein and Torres - Samaniego, 2021**). In line with this evidence, the *Ras* and *MAPK/ERK signalling pathways* have been proposed to play a significant role in promoting angiogenesis by triggering *VEGF* expression, being possibly implicated in the adaptive response to hypoxia evolved by high-altitude populations (**Figure 3A**; **Kanehisa and Goto, 2000**; **Kranenburg et al., 2004**). Moreover, in the study by **Lorenzo et al., 2014** a gain-of-function mutation in the *EGLN1* gene was observed in Tibetans and was proved to enhance the catalytic activity of the HIF prolyl 4-hydroxylase 2 (*PHD2*) under hypoxic conditions. This alters the binding of HIF-2α (the isoform 2 of the inducible hypoxia transcriptional factor encoded by *EPAS1*) and negatively regulates the activation of the *HIF-1 signalling pathway* during hypoxia, eventually offering protection against the detrimental effects of prolonged polycythaemia. When HIF-2α exerts its transcriptional activities along with p300 protein and HIF-1β, it enhances *VEGF* expression and permits the activation of the *VEGF signalling pathway* (**Kanehisa and Goto, 2000**; **Rashid et al., 2021**). Accordingly, down-regulation of the *HIF-1 signalling pathway* comports the reduction of VEGF mRNA expression (**Greenberger et al., 2008**; **Zhang et al., 2018**). Coupled with these observations, results from the *Signet* analysis suggest that in individuals of Tibetan ancestry, when the *HIF-1 signalling pathway* is likely down-regulated in chronic hypoxic conditions (**Lorenzo et al., 2014**), adaptive changes at the *Ras/MAPK signalling pathways* could represent alternative biological mechanisms that in its place enable to sustain improved angiogenesis and thus permit adequate tissue oxygenation (**Figure 3A**).

The same five candidate genes, in addition to the *EGLN1* locus considered as a negative control for AI, were also confirmed by the *LASSI* method to have adaptively evolved in the studied populations (**Figure 4C, D**, **Figure 4—figure supplement 1A, B**, **Figure 4—figure supplements 2 and 3C, D**). In fact, several genomic windows in their associated chromosomal intervals and/or in their flanking regions presented positive values of the computed likelihood *T* statistic, many of which falling in the top 5% of the related distribution, which indicate their non-neutral evolution. Interestingly, *TBC1D1*, *RASGRF2*, *PRKAG2*, *KRAS*, and *EPAS1* significant genomic windows pointed out by *LASSI* were also found to include Denisovan-like derived alleles that are completely absent in the YRI African population, suggesting that positive selection acted in Tibetans on haplotypes containing archaic introgressed variation. Such a scenario was further supported by the *Haplostrips* analysis, which revealed for all the loci mentioned above patterns of similarity between Tibetan and Denisovan haplotypes that are comparable to that observed for *EPAS1* (**Figure 4A, B**, **Figure 4—figure supplement 2A, B**), with the sole exception being represented by *EGLN1* (**Figure 4—figure supplement 1B**, **Figure 4—figure supplement 3B, D**, **Figure 4—figure supplement 3A**).

Overall, we collected multiple evidence supporting both the archaic origin and the adaptive role of variation at *TBC1D1*, *RASGRF2*, *PRKAG2*, and *KRAS* genes in populations of Tibetan ancestry. Genetic signatures at such loci are especially consistent with the hypothesis of adaptive events mediated by soft selective sweeps and/or polygenic mechanisms that involved haplotypes including both modern and archaic introgressed alleles. Therefore, the results obtained have succeeded in expanding the knowledge about AI events having involved the ancestors of modern high-altitude Himalayan populations and Denisovans and emphasized once more the complexity of the adaptive phenotype evolved by these human groups to cope with challenges imposed by hypobaric hypoxia. Accordingly, they offer new insights for future studies aimed at elucidating the molecular mechanisms by which several genes along with *TBC1D1*, *PRKAG2*, *RASGRF2*, and *KRAS* interact with each other and contribute to mediate physiological adjustments that are crucial for human adaptation to the high-altitude environment.

Materials and methods

Ethics

The University of Bologna Ethics Committee released approval (Prot. 205142, 12/9/2019) for the present study within the framework of the project 'Genetic adaptation and acclimatization to high altitude as experimental models to investigate the biological mechanisms that regulate physiological responses to hypoxia'. However, no new sampling campaign was performed in the context of the present study and all the genomic data analysed were publicly available. The informed consent for the 27 Tibetan WGS analysed here was previously obtained and declared in the *Ethics statement* section of the study by ***Jeong et al., 2018***.

Dataset composition and curation

The dataset used in the present study included WGS data for 27 individuals of Tibetan ancestry recruited from the high-altitude Nepalese districts of Mustang and Ghorka (***Jeong et al., 2018***). Although these subjects reside in Nepal, they have been previously proved to speak Tibetan dialects and to live in communities showing religious and social organizations proper of populations settled on the Tibetan Plateau, being also biologically representative of high-altitude Tibetan people (***Cho et al., 2017***). To filter for high-quality genotypes, the dataset was subjected to QC procedures using the software PLINK v1.9 (***Purcell et al., 2007***). In detail, we retained autosomal SNVs characterized by no significant deviations from the Hardy–Weinberg equilibrium ($p > 5.3 \times 10^{-9}$ after Bonferroni correction for multiple testing), as well as loci/samples showing less than 5% of missing data. Moreover, we removed SNVs with ambiguous A/T or G/C alleles and multiallelic variants, obtaining a dataset composed by 27 individuals and 6,921,628 SNVs. WGS data were finally phased with SHAPEIT2 v2.r904 (***Delaneau et al., 2013***) by applying default parameters and using the 1000 Genomes Project dataset as a reference panel (***Auton et al., 2015***) and HapMap phase 3 recombination maps.

Population structure analyses

To assess representativeness, genetic homogeneity, and ancestry composition of Tibetan WGS included in the dataset, we performed genotype-based population structure analyses. For this purpose, we merged the unphased WGS dataset with genome-wide genotyping data for 34 East-Asian populations (***Gnecchi-Ruscone et al., 2017***; ***Landini et al., 2021***) and we applied the same QC described above. The obtained extended dataset included 231,947 SNVs and was used to assess the degree of recent shared ancestry (i.e., identity-by-descent, IBD) for each pair of individuals. Identity-by-state (IBS) estimates were thus used to calculate an IBD kinship coefficient and a threshold of PI_HAT = 0.270 was considered to remove closely related subjects to the second degree (***Ojeda-Granados et al., 2022***). To discard variants in linkage disequilibrium (LD) with each other we then removed one SNV for each pair showing $r^2 > 0.2$ within sliding windows of 50 SNVs and advancing by 5 SNVs at the time. The obtained LD-pruned dataset was finally filtered for variants with minor allele frequency <0.01 and used to compute PCA utilizing the *smartpca* method implemented in the EIGENSOFT package (***Patterson et al., 2006***), as well as to run the ADMIXTURE algorithm version 1.3.0 (***Alexander et al., 2009***) by testing $K = 2$ to $K = 12$ population clusters. In detail, 25 replicates with different random seeds were run for each K and we retained only those presenting the highest

log-likelihood values. In addition, cross-validation errors were calculated for each *K* to identify the value that best fit the data. Both PCA and ADMIXTURE results were visualized with the R software version 4.0.5. ADMIXTURE results were visualized with the R software version 4.0.5.

Detecting signatures of AI

To identify chromosomal regions showing signatures putatively ascribable to AI events, we submitted the phased WGS dataset to the *VolcanoFinder* pipeline, which relies on the analysis of the allele frequency spectrum of the population that is supposed to have experienced archaic introgression (***Setter et al., 2020***). This model considers three populations: recipient (i.e., the modern population), donor (i.e., the archaic population), and outgroup and assumes the occurrence of the introgression event after which a beneficial haplotype is introduced in the modern gene pool and starts to rise in frequency because of demographic random processes and because of the action of natural selection on it. The influence of these evolutionary forces comports the existence of multiple haplotypes that carry the beneficial archaic allele, thus comporting elevated heterozygosity level tested by the LR statistic. Therefore, the pattern of variability tested by such a statistic deeply differs from that attributable to classical selective sweeps (especially from that associated to a hard selective sweep), allowing to identify weak signatures of adaptation that resulted from both the introgression and the action of natural selection on beneficial standing genetic variation.

The *VolcanoFinder* algorithm was chosen among several approaches developed to detect AI signatures according to the following reasons. First of all, it can test jointly both archaic introgression and adaptive evolution according to a model that differs from those considered by other statistics that are aimed at identifying chromosomal segments showing low divergence with respect to a specific archaic sequence and/or enriched in alleles uniquely shared between the admixed group and the archaic source and characterized by a frequency above a certain threshold in the population under study (***Racimo et al., 2017***). In fact, these methods are especially useful to test an evolutionary scenario conformed to that expected in the case that adaptation was mediated by hard selective sweeps. On the contrary, *VocanoFinder* was proved to have an elevated power in the identification of AI events mediated by more than one predominant haplotype (***Setter et al., 2020***), as expected when soft sweeps/polygenic adaptations occurred. Moreover, *VolcanoFinder* relies on less demanding computational efforts with respect to algorithms that require to be trained on genomic simulations built specifically to reflect the evolutionary history of the population under study (***Gower et al., 2021***; ***Zhang et al., 2023***), but possibly introducing bias in the obtained results if the information that guides simulations is not accurate.

We thus scanned Tibetan WGS data using the *VolcanoFinder* method and calculating two statistics for each polymorphic site: α (subsequently converted in $-\log\alpha$) and LR which are informative, respectively, of: (1) the strength of natural selection and (2) the conformity to the evolutionary model of AI. Since elevated LR scores are assumed to support AI signatures (***Setter et al., 2020***), we filtered the most significant results obtained by focusing on SNVs showing LR values falling in the positive tail (i.e., top 5%) of the distribution built for such a statistic. We then visualized the distribution of both α and LR parameters across the genomic regions including the *EPAS1* and *EGLN1* genes, which we considered, respectively, as positive and negative control loci for AI and by investigating chromosomal intervals spanning 50 kb up- and downstream to these genes. We then defined the new candidate AI genomic regions based on their conformity with the pattern observed for the positive AI control gene (i.e., according to the detection of multiple peaks of LR scores consistently distributed in the entire genomic region considered and coupled with elevated values of the $-\log\alpha$ parameter). Moreover, we relied on evidence advanced by previous studies aimed at detecting archaic introgression signatures from WGS data for individuals with Tibetan and/or Han Chinese ancestries (***Hu et al., 2017***; ***Browning et al., 2018***; ***Zhang et al., 2021***) to filter out loci potentially targeted by natural selection in Tibetans, but with questionable archaic origins.

Identifying gene networks including Denisovan-like derived alleles

To validate archaic introgression signatures inferred with *VolcanoFinder* by using an independent method, we followed the *Signet* approach described by ***Gouy and Excoffier, 2020***, with the aim of identifying biological functions whose underlying genomic regions might have been significantly shaped by Denisovan introgression. The *Signet* approach consists in crosschecking the information

contained in the input dataset with that collected in reference databases of functional pathways, such as the KEGG (available at https://www.kegg.jp/), using a simulated annealing algorithm approach to define the High Scoring Subnetworks within each biological pathway (**Gouy and Excoffier, 2020**). In detail, we used the *Signet* algorithm to reconstruct network of genes that participate to the same biological pathway and that also included Denisovan-like derived alleles observable in the Tibetan population but not in an outgroup population of African ancestry, by assuming that only Eurasian *H. sapiens* populations experienced Denisovan admixture. To do so, we first compared the Tibetan and Denisovan genomes to assess which SNVs were present in both modern and archaic sequences. These loci were further compared with the ancestral reconstructed refence human genome sequence to discard those presenting an ancestral state (i.e., that we have in common with several primate species). Moreover, we further filtered the considered variants by retaining only those alleles that were completely absent (i.e., with frequency equal to zero) in the YRI population sequenced by the 1000 Genomes Project (**Auton et al., 2015**). We then calculated the average frequency in the Tibetan population of the Denisovan-like derived alleles observable in each gene and we used the obtained genomic distribution to inform *Signet*. We performed five independent runs of the *Signet* algorithm to check for consistency of the significant gene networks and functional pathways identified and we finally depicted the confirmed results using Cytoscape v3.9.1 (**Shannon et al., 2003**).

Testing adaptive evolution of candidate introgressed loci

To confirm signatures ascribable to the action of natural selection at the putative introgressed loci supported by both *VolcanoFinder* and *Signet* analyses, we applied the *LASSI* likelihood method (**Harris and DeGiorgio, 2020**) on the available Tibetan WGS data. Such an approach detects and distinguishes genomic regions that have experienced different types of selective events (i.e., strong and weak ones) and has been demonstrated to be more powerful in doing so than other haplotype-based approaches (**Harris and DeGiorgio, 2020**). The rationale behind the *LASSI* approach is based on the recognition of the modification resulted in the haplotype frequency spectrum of a given genomic region after the action of natural selection on it. For instance, according to the hard sweep model, when in the population arise a beneficial mutation with a very strong impact on a given phenotypic trait, the haplotype frequency spectrum of such a genomic region will be characterized by a single haplotype with an extremely elevated frequency in the population (i.e., the sweeping haplotype), while the other haplotypes whether they exist, are found at very low frequencies. Consequentially, when the selection acts on few new variants with a lower impact on a trait or on standing genetic variation, the resulted haplotype spectrum will be characterized by the existence of two or few more haplotypes that reach moderate frequencies in the population. Conversely, the variability pattern associated to the haplotype frequency spectrum expected under neutrality will be characterized by a series of different haplotypes at low frequencies in the population.

Specifically, we calculated for each genomic window the likelihood *T* statistic, which measures the conformity of variability patterns of the analysed region to those expected according to a haplotype frequency spectrum under adaptive rather than neutral evolution. In addition, the *LASSI* algorithm calculated the parameter *m* (i.e., the number of sweeping haplotypes) for each genomic region, thus classifying them as affected by hard sweeps (when $m = 1$) or soft sweeps (when $m > 1$). In detail, *T* scores significantly different from zero indicate the conformity with a non-neutral evolutionary scenario, with ever higher likelihood scores being indicative of increasingly robust evidence for a selective event (**Harris and DeGiorgio, 2020**).

The method requires to fix a custom value for the length of the considered genomic windows, which are measured in terms of the number of SNVs included in them, and to move windows by 1 SNV after each computation. Therefore, we selected this fixed-length value (i.e., 13 SNVs) by estimating the average number of SNVs included into a haplotype block as defined for the population under study by using the `--blocks` function implemented in PLINK v1.9 (**Purcell et al., 2007**). Moreover, by following the indications by **Harris and DeGiorgio, 2020** of choosing values for the fixed number of haplotypes in the spectrum (i.e., *K* values) <10 for increasing the power of the *T* statistic, we set it at seven. Finally, we choose the likelihood model 3 to calculate the *T* statistic and we applied the *LASSI* algorithm to the phased WGS dataset.

We then focused on the genomic windows showing *T* scores falling in the positive tail (i.e., top 5%) of the obtained distribution and we crosschecked these results with those significant ones pointed out

by *VolcanoFinder* and *Signet* analyses to shortlist genomic regions having plausibly experienced both archaic introgression and adaptive evolution.

Haplostrip analysis

To explicitly test whether the putative adaptive introgressed loci pointed out by *VolcanoFinder*, *Signet*, and *LASSI* analyses present variation patterns compatible with a scenario of introgression from the Denisovan archaic human species, we estimated genetic distance between modern and archaic haplotypes inferred for those genomic windows supported by all the methods mentioned above, as well as for *EPAS1* and *EGLN1* for the sake of comparison with established positive and negative control genes that have been previously proved to be involved or not in AI events. For this purpose, we used the *Haplostrip* pipeline, as described in previous studies (**Huerta-Sánchez et al., 2014**; **Marnetto and Huerta-Sánchez, 2017**). Moreover, since the *EGLN1* gene did not include any Denisovan-like variant as identified according to the criteria described in the previous paragraphs, we choose to build the *Haplostrips* heatmap by considering the *EGLN1* genomic window associated with the highest value of the *LASSI* statistic.

In detail, we merged the 27 Tibetan whole genomes under study (**Jeong et al., 2018**) with 27 CHB, 27 YRI WGSs (**Auton et al., 2015**) and with the Denisovan genome (**Meyer et al., 2013**) (downloadable at http://cdna.eva.mpg.de/neandertal/altai/Denisovan/). The CHB population, which is known to share a relatively recent common ancestry with Tibetans, was used as a 'negative low-altitude control' (i.e., as a group whose ancestors experienced Denisovan introgression, but did not evolve high-altitude adaptation). YRI individuals were instead used as the outgroup population (i.e., a population that presumably did not experience Denisovan admixture), as previously proposed (**Zhang et al., 2021**). We then phased the assembled dataset with SHAPEIT2 v2.r904 (**Delaneau et al., 2013**), as described in the *Dataset composition and curation* section and we run the *Haplostrip* algorithm.

Acknowledgements

We acknowledge support from the Fondazione Cassa di Risparmio in Bologna through the project 'Genetic adaptation and acclimatization to high altitude as experimental models to investigate the biological mechanisms that regulate physiological responses to hypoxia', which was granted to MS (n. 2019.0552).

Additional information

Funding

Funder	Grant reference number	Author
Fondazione Cassa di Risparmio in Bologna	2019.0552	Marco Sazzini

The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

Author contributions

Giulia Ferraretti, Paolo Abondio, Data curation, Software, Formal analysis, Investigation, Writing – original draft; Marta Alberti, Software, Formal analysis; Agnese Dezi, Phurba T Sherpa, Paolo Cocco, Massimiliano Tiriticco, Marco Di Marcello, Guido Alberto Gnecchi-Ruscone, Luca Natali, Angela Corcelli, Giorgio Marinelli, Davide Peluzzi, Data curation, Writing – review and editing; Stefania Sarno, Formal analysis, Writing – original draft; Marco Sazzini, Conceptualization, Resources, Supervision, Funding acquisition, Writing – review and editing

Author ORCIDs

Marco Sazzini https://orcid.org/0000-0001-5382-7827

Ethics

The University of Bologna Ethics Committee released approval (Prot. 205142, 12/9/2019) for the present study within the framework of the project 'Genetic adaptation and acclimatization to high altitude as experimental models to investigate the biological mechanisms that regulate physiological responses to hypoxia'. However, no new sampling campaign was performed in the context of the present study and all the genomic data analysed were publicly available. The informed consent for the 27 Tibetan WGS analysed here was previously obtained and declared in the Ethics statement section of the study by Jeong et al., 2018.

Peer review material

Reviewer #1 (Public review): https://doi.org/10.7554/eLife.89815.3.sa1
Reviewer #2 (Public review): https://doi.org/10.7554/eLife.89815.3.sa2
Author response https://doi.org/10.7554/eLife.89815.3.sa3

Additional files

Supplementary files

- Supplementary file 1. Supplementary tables 1a-1f. (**a**) Populations included in the extended dataset. (**b**) Single-nucleotide variants (SNVs) associated with values falling in top 5% of the distribution of likelihood ratio (LR) statistic calculated by *VolcanoFinder*. (**c**) SNVs associated with values falling in top 5% of the distribution of the LR statistic and comprised in the genomic region of the *EPAS1* gene (i.e., 2:46474546–2:46663836). (**d**) SNVs associated with values falling in top 5% of the distribution of the LR statistic and comprised in the genomic region of the *EGLN1* gene (i.e., 1:231449502–1:231608033). (**e**) Adaptive introgressed genes confirmed by *VolcanoFinder* and previous studies. (**f**) Gene networks including Denisovan-like derived alleles identified according to the *Signet* approach.
- MDAR checklist

Data availability

The current manuscript is a computational study, so no data have been generated for this manuscript. The dataset used has been generated by ***Jeong et al., 2018***. The code and the software used have been developed by ***Marnetto and Huerta-Sánchez, 2017***; ***Setter et al., 2020***; ***Gouy and Excoffier, 2020***; ***Harris and DeGiorgio, 2020***.

The following previously published dataset was used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Jeong C, Witonsky DB, Basnyat B, Neupane M, Beall CM, Childs G, Craig SR, Novembre J, Di Rienzo A	2018	Tibetan/Sherpa Sequence Reads	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA420511/	NCBI BioProject, PRJNA420511

References

Abdulla MA, Ahmed I, Assawamakin A, Bhak J, Brahmachari SK, Calacal GC, Chaurasia A, Chen CH, Chen J, Chen YT, Chu J, Cutiongco-de la Paz EMC, De Ungria MCA, Delfin FC, Edo J, Fuchareon S, Ghang H, Gojobori T, Han J, Ho SF, et al. 2009. Mapping human genetic diversity in Asia. *Science* **326**:1541–1545. DOI: https://doi.org/10.1126/science.1177074, PMID: 20007900

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**:1655–1664. DOI: https://doi.org/10.1101/gr.094052.109, PMID: 19648217

Arciero E, Kraaijenbrink T, Haber M, Mezzavilla M, Ayub Q, Wang W, Pingcuo Z, Yang H, Wang J, Jobling MA, van Driem G, Xue Y, de Knijff P, Tyler-Smith C. 2018. Demographic history and genetic adaptation in the himalayan region inferred from genome-wide snp genotypes of 49 populations. *Molecular Biology and Evolution* **35**:1916–1933. DOI: https://doi.org/10.1093/molbev/msy094, PMID: 29796643

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR, 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* **526**:68–74. DOI: https://doi.org/10.1038/nature15393, PMID: 26432245

Beall CM. 2007. Two routes to functional adaptation: tibetan and andean high-altitude natives. *PNAS* **104**:8655–8660. DOI: https://doi.org/10.1073/pnas.0701985104

Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, Li C, Li JC, Liang Y, McCormack M, Montgomery HE, Pan H, Robbins PA, Shianna KV, Tam SC, Tsering N, Veeramah KR, Wang W, Wangdui P, Weale ME, et al. 2010. Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. *PNAS* **107**:11459–11464. DOI: https://doi.org/10.1073/pnas.1002443107, PMID: 20534544

Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, Mei R, Scherer SW, Julian CG, Wilson MJ, López Herráez D, Brutsaert T, Parra EJ, Moore LG, Shriver MD. 2010. Identifying signatures of natural selection in tibetan and andean populations using dense genome scan data. *PLOS Genetics* **6**:e1001116. DOI: https://doi.org/10.1371/journal.pgen.1001116, PMID: 20838600

Brahimi-Horn MC, Chiche J, Pouysségur J. 2007. Hypoxia and cancer. *Journal of Molecular Medicine* **85**:1301–1307. DOI: https://doi.org/10.1007/s00109-007-0281-3, PMID: 18026916

Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM. 2018. Analysis of human sequence data reveals two pulses of archaic denisovan admixture. *Cell* **173**:53–61. DOI: https://doi.org/10.1016/j.cell.2018.02.031, PMID: 29551270

Calderón - Gerstein WS, Torres - Samaniego G. 2021. High altitude and cancer: an old controversy. *Respiratory Physiology & Neurobiology* **289**:103655. DOI: https://doi.org/10.1016/j.resp.2021.103655

Catar R, Witowski J, Wagner P, Annett Schramm I, Kawka E, Philippe A, Dragun D, Jörres A. 2013. The proto-oncogene C-Fos transcriptionally regulates VEGF production during peritoneal inflammation. *Kidney International* **84**:1119–1128. DOI: https://doi.org/10.1038/ki.2013.217, PMID: 23760290

Chen X, Li X, Zhang W, He J, Xu B, Lei B, Wang Z, Cates C, Rousselle T, Li J. 2018. Activation of AMPK inhibits inflammatory response during hypoxia and reoxygenation through modulating JNK-mediated NF-κB pathway. *Metabolism* **83**:256–270. DOI: https://doi.org/10.1016/j.metabol.2018.03.004, PMID: 29526538

Cho JI, Basnyat B, Jeong C, Di Rienzo A, Childs G, Craig SR, Sun J, Beall CM. 2017. Ethnically Tibetan women in Nepal with low hemoglobin concentration have better reproductive outcomes. *Evolution, Medicine, and Public Health* **2017**:82–96. DOI: https://doi.org/10.1093/emph/eox008, PMID: 28567284

Cole MA, Abd Jamil AH, Heather LC, Murray AJ, Sutton ER, Slingo M, Sebag-Montefiore L, Tan SC, Aksentijević D, Gildea OS, Stuckey DJ, Yeoh KK, Carr CA, Evans RD, Aasum E, Schofield CJ, Ratcliffe PJ, Neubauer S, Robbins PA, Clarke K. 2016. On the pivotal role of PPARα in adaptation of the heart to hypoxia and why fat in the diet increases hypoxic injury. *FASEB Journal* **30**:2684–2697. DOI: https://doi.org/10.1096/fj.201500094R, PMID: 27103577

Dannemann M, Racimo F. 2018. Something old, something borrowed: admixture and adaptation in. *Current Opinion in Genetics and Development* **53**:009. DOI: https://doi.org/10.1016/j.gde.2018.05

Delaneau O, Zagury JF, Marchini J. 2013. Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods* **10**:5–6. DOI: https://doi.org/10.1038/nmeth.2307, PMID: 23269371

Deng L, Zhang C, Yuan K, Gao Y, Pan Y, Ge X, He Y, Yuan Y, Lu Y, Zhang X, Chen H, Lou H, Wang X, Lu D, Liu J, Tian L, Feng Q, Khan A, Yang Y, Jin Z-B, et al. 2019. Prioritizing natural-selection signals from the deep-sequencing genomic data suggests multi-variant adaptation in Tibetan highlanders. *National Science Review* **6**:1201–1222. DOI: https://doi.org/10.1093/nsr/nwz108, PMID: 34691999

Dengler F. 2020. Activation of AMPK under Hypoxia: many roads leading to rome. *International Journal of Molecular Sciences* **21**:2428. DOI: https://doi.org/10.3390/ijms21072428

Enard D, Petrov DA. 2018. Evidence that RNA viruses drove adaptive introgression between neanderthals and modern humans. *Cell* **175**:360–371.. DOI: https://doi.org/10.1016/j.cell.2018.08.034, PMID: 30290142

Gittelman RM, Schraiber JG, Vernot B, Mikacenic C, Wurfel MM, Akey JM. 2016. Archaic hominin admixture facilitated adaptation to out-of-africa environments. *Current Biology* **26**:3375–3382. DOI: https://doi.org/10.1016/j.cub.2016.10.041, PMID: 27839976

Gnecchi-Ruscone GA, Jeong C, De Fanti S, Sarno S, Trancucci M, Gentilini D, Di Blasio AM, Sherpa MG, Sherpa PT, Marinelli G, Di Marcello M, Natali L, Peluzzi D, Pettener D, Di Rienzo A, Luiselli D, Sazzini M. 2017. The genomic landscape of Nepalese Tibeto-Burmans reveals new insights into the recent peopling of Southern Himalayas. *Scientific Reports* **7**:15512. DOI: https://doi.org/10.1038/s41598-017-15862-z, PMID: 29138459

Gnecchi-Ruscone GA, Abondio P, De Fanti S, Sarno S, Sherpa MG, Sherpa PT, Marinelli G, Natali L, Di Marcello M, Peluzzi D, Luiselli D, Pettener D, Sazzini M. 2018. Evidence of polygenic adaptation to high altitude from tibetan and sherpa genomes. *Genome Biology and Evolution* **10**:2919–2930. DOI: https://doi.org/10.1093/gbe/evy233, PMID: 30335146

Gouy A, Excoffier L. 2020. Polygenic patterns of adaptive introgression in modern humans are mainly shaped by response to pathogens. *Molecular Biology and Evolution* **37**:1420–1433. DOI: https://doi.org/10.1093/molbev/msz306, PMID: 31935281

Gower G, Picazo PI, Fumagalli M, Racimo F. 2021. Detecting adaptive introgression in human evolution using convolutional neural networks. *eLife* **10**:e64669. DOI: https://doi.org/10.7554/eLife.64669, PMID: 34032215

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MHY, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, et al. 2010. A draft sequence of the Neandertal genome. *Science* **328**:710–722. DOI: https://doi.org/10.1126/science.1188021, PMID: 20448178

Greenberger LM, Horak ID, Filpula D, Sapra P, Westergaard M, Frydenlund HF, Albaek C, Schrøder H, Ørum H. 2008. A RNA antagonist of hypoxia-inducible factor-1alpha, EZN-2968, inhibits tumor cell growth. *Molecular Cancer Therapeutics* **7**:3598–3608. DOI: https://doi.org/10.1158/1535-7163.MCT-08-0510, PMID: 18974394

Hamid I, Korunes KL, Beleza S, Goldberg A. 2021. Rapid adaptation to malaria facilitated by admixture in the human population of Cabo Verde. *eLife* **10**:e63177. DOI: https://doi.org/10.7554/eLife.63177, PMID: 33393457

Harris AM, DeGiorgio M. 2020. A likelihood approach for uncovering selective sweep signatures from haplotype data. *Molecular Biology and Evolution* **37**:3023–3046. DOI: https://doi.org/10.1093/molbev/msaa115, PMID: 32392293

Horscroft JA, Kotwica AO, Laner V, West JA, Hennis PJ, Levett DZH, Howard DJ, Fernandez BO, Burgess SL, Ament Z, Gilbert-Kawai ET, Vercueil A, Landis BD, Mitchell K, Mythen MG, Branco C, Johnson RS, Feelisch M, Montgomery HE, Griffin JL, et al. 2017. Metabolic basis to Sherpa altitude adaptation. *PNAS* **114**:6382–6387. DOI: https://doi.org/10.1073/pnas.1700527114, PMID: 28533386

Hu H, Petousi N, Glusman G, Yu Y, Bohlender R, Tashi T, Downie JM, Roach JC, Cole AM, Lorenzo FR, Rogers AR, Brunkow ME, Cavalleri G, Hood L, Alpatty SM, Prchal JT, Jorde LB, Robbins PA, Simonson TS, Huff CD. 2017. Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLOS Genetics* **13**:e1006675. DOI: https://doi.org/10.1371/journal.pgen.1006675, PMID: 28448578

Huerta-Sánchez E, Jin X, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Somel M, Ni P, Wang B, Ou X, Luosang J, Cuo ZXP, Li K, Gao G, Yin Y, Wang W, Zhang X, Xu X, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**:194–197. DOI: https://doi.org/10.1038/nature13408, PMID: 25043035

Jeong C, Alkorta-Aranburu G, Basnyat B, Neupane M, Witonsky DB, Pritchard JK, Beall CM, Di Rienzo A. 2014. Admixture facilitates genetic adaptations to high altitude in Tibet. *Nature Communications* **5**:1–7. DOI: https://doi.org/10.1038/ncomms4281, PMID: 24513612

Jeong C, Witonsky DB, Basnyat B, Neupane M, Beall CM, Childs G, Craig SR, Novembre J, Di Rienzo A. 2018. Detecting past and ongoing natural selection among ethnically Tibetan women at high altitude in Nepal. *PLOS Genetics* **14**:e1007650. DOI: https://doi.org/10.1371/journal.pgen.1007650, PMID: 30188897

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**:27–30. DOI: https://doi.org/10.1093/nar/28.1.27

Kolár F, Ostádal B. 2004. Molecular mechanisms of cardiac protection by adaptation to chronic hypoxia. *Physiological Research* **53**:S3–S13 PMID: 15119931.

Kranenburg O, Gebbink MFBG, Voest EE. 2004. Stimulation of angiogenesis by Ras proteins. *Biochimica et Biophysica Acta* **1654**:23–37. DOI: https://doi.org/10.1016/j.bbcan.2003.09.004, PMID: 14984765

Landini A, Yu S, Gnecchi-Ruscone GA, Abondio P, Ojeda-Granados C, Sarno S, De Fanti S, Gentilini D, Di Blasio AM, Jin H, Nguyen TT, Romeo G, Prata C, Bortolini E, Luiselli D, Pettener D, Sazzini M. 2021. Genomic adaptations to cereal-based diets contribute to mitigate metabolic risk in some human populations of East Asian ancestry. *Evolutionary Applications* **14**:297–313. DOI: https://doi.org/10.1111/eva.13090, PMID: 33664777

Laurent AR, J. M. J, Subhash K, Alasdair M, Klara D, Loren G, Farbod B, Baback G, Ma L, P. F. A, Joshua K, Mary C, Derek M, Raja R, Meral B, Marsh SGE, Martin M, Lisbeth AG, Sofia T, Peter P. 2011. The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science (New York, N.Y.)* **334**:89–94. DOI: https://doi.org/10.1126/science.1209202

Lei L, Liu M, Ma D, Lei X, Zeng S, Li P, Huang K, Lyu J, Lei Q. 2024. Cardioprotective effects of high-altitude adaptation in cardiac surgical patients: a retrospective cohort study with propensity score matching. *Frontiers in Cardiovascular Medicine* **11**:1347552. DOI: https://doi.org/10.3389/fcvm.2024.1347552, PMID: 38628317

Liu C-C, Witonsky D, Gosling A, Lee JH, Ringbauer H, Hagan R, Patel N, Stahl R, Novembre J, Aldenderfer M, Warinner C, Di Rienzo A, Jeong C. 2022. Ancient genomes from the Himalayas illuminate the genetic history of Tibetans and their Tibeto-Burman speaking neighbors. *Nature Communications* **13**:1203. DOI: https://doi.org/10.1038/s41467-022-28827-2, PMID: 35260549

Lorca RA, Houck JA, Laurent LC, Matarazzo CJ, Baker K, Horii M, Nelson KK, Bales ES, Euser AG, Parast MM, Moore LG, Julian CG. 2021. High altitude regulates the expression of AMPK pathways in human placenta. *Placenta* **104**:267–276. DOI: https://doi.org/10.1016/j.placenta.2021.01.010, PMID: 33472134

Lorenzo FR, Huff C, Myllymäki M, Olenchock B, Swierczek S, Tashi T, Gordeuk V, Wuren T, Ri-Li G, McClain DA, Khan TM, Koul PA, Guchhait P, Salama ME, Xing J, Semenza GL, Liberzon E, Wilson A, Simonson TS, Jorde LB, et al. 2014. A genetic mechanism for Tibetan high-altitude adaptation. *Nature Genetics* **46**:951–956. DOI: https://doi.org/10.1038/ng.3067, PMID: 25129147

Mallet RT, Manukhina EB, Ruelas SS, Caffrey JL, Downey HF. 2018. Cardioprotection by intermittent hypoxia conditioning: evidence, mechanisms, and therapeutic potential. *American Journal of Physiology. Heart and Circulatory Physiology* **315**:H216–H232. DOI: https://doi.org/10.1152/ajpheart.00060.2018, PMID: 29652543

Marnetto D, Huerta-Sánchez E. 2017. *Haplostrips* : revealing population structure through haplotype visualization . *Methods in Ecology and Evolution* **8**:1389–1392. DOI: https://doi.org/10.1111/2041-210X.12747

Maxwell PH, Ratcliffe PJ. 2002. Oxygen sensors and angiogenesis. *Seminars in Cell & Developmental Biology* **13**:29–37. DOI: https://doi.org/10.1006/scdb.2001.0287, PMID: 11969369

McArthur E, Rinker DC, Capra JA. 2021. Quantifying the contribution of Neanderthal introgression to the heritability of complex traits. *Nature Communications* **12**:4481. DOI: https://doi.org/10.1038/s41467-021-24582-y, PMID: 34294692

McCoy RC, Wakefield J, Akey JM. 2017. Impacts of Neanderthal-introgressed sequences on the landscape of human gene expression. *Cell* **168**:916–927.. DOI: https://doi.org/10.1016/j.cell.2017.01.038, PMID: 28235201

Meyer M, Kircher M, Gansauge M, Li H, Mallick S, Schraiber JG, Jay F, Prüfer K, De C, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Siebauer M, Green RE, Bryc K, Briggs AW, Dabney J, Eichler EE. 2013. A high

coverage genome sequence from an archaic. Semantic Scholar. DOI: https://doi.org/10.1126/science.1224344, PMID: Semantic Scholar

Monje P, Hernández-Losa J, Lyons RJ, Castellone MD, Gutkind JS. 2005. Regulation of the transcriptional activity of c-Fos by ERK. A novel role for the prolyl isomerase PIN1. *The Journal of Biological Chemistry* **280**:35081–35084. DOI: https://doi.org/10.1074/jbc.C500353200, PMID: 16123044

Ojeda-Granados C, Abondio P, Setti A, Sarno S, Gnecchi-Ruscone GA, González-Orozco E, De Fanti S, Jiménez-Kaufmann A, Rangel-Villalobos H, Moreno-Estrada A, Sazzini M. 2022. Dietary, cultural, and pathogens-related selective pressures shaped differential adaptive evolution among native mexican populations. *Molecular Biology and Evolution* **39**:msab290. DOI: https://doi.org/10.1093/molbev/msab290, PMID: 34597392

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLOS Genetics* **2**:e190. DOI: https://doi.org/10.1371/journal.pgen.0020190, PMID: 17194218

Peng Y, Yang Z, Zhang H, Cui C, Qi X, Luo X, Tao X, Wu T, Chen H, Shi H, Su B. 2011. Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. *Molecular Biology and Evolution* **28**:1075–1081. DOI: https://doi.org/10.1093/molbev/msq290, PMID: 21030426

Porto AG, Brun F, Severini GM, Losurdo P, Fabris E, Taylor MRG, Mestroni L, Sinagra G. 2016. Clinical spectrum of PRKAG2 syndrome. *Circulation. Arrhythmia and Electrophysiology* **9**:e003121. DOI: https://doi.org/10.1161/CIRCEP.115.003121, PMID: 26729852

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, Li H, Mallick S, Dannemann M, Fu Q, Kircher M, Kuhlwilm M, Lachmann M, Meyer M, Ongyerth M, Siebauer M, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**:43–49. DOI: https://doi.org/10.1038/nature12886, PMID: 24352235

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**:559–575. DOI: https://doi.org/10.1086/519795, PMID: 17701901

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. 2015. Evidence for archaic adaptive introgression in humans. *Nature Reviews. Genetics* **16**:359–371. DOI: https://doi.org/10.1038/nrg3936, PMID: 25963373

Racimo F, Marnetto D, Huerta-Sánchez E. 2017. Signatures of archaic adaptive introgression in present-day human populations. *Molecular Biology and Evolution* **34**:296–317. DOI: https://doi.org/10.1093/molbev/msw216, PMID: 27756828

Rashid M, Zadeh LR, Baradaran B, Molavi O, Ghesmati Z, Sabzichi M, Ramezani F. 2021. Up-down regulation of HIF-1α in cancer progression. *Gene* **798**:145796. DOI: https://doi.org/10.1016/j.gene.2021.145796

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**:1053–1060. DOI: https://doi.org/10.1038/nature09710, PMID: 21179161

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**:354–357. DOI: https://doi.org/10.1038/nature12961, PMID: 24476815

Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, et al. 2022. Database resources of the national center for biotechnology information. *Nucleic Acids Research* **50**:D20–D26. DOI: https://doi.org/10.1093/nar/gkab1112, PMID: 34850941

Scruggs SB, Wang D, Ping P. 2016. PRKCE gene encoding protein kinase C-epsilon-Dual roles at sarcomeres and mitochondria in cardiomyocytes. *Gene* **590**:90–96. DOI: https://doi.org/10.1016/j.gene.2016.06.016, PMID: 27312950

Setter D, Mousset S, Cheng X, Nielsen R, DeGiorgio M, Hermisson J. 2020. VolcanoFinder: Genomic scans for adaptive introgression. *PLOS Genetics* **16**:e1008867. DOI: https://doi.org/10.1371/journal.pgen.1008867

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**:2498–2504. DOI: https://doi.org/10.1101/gr.1239303, PMID: 14597658

Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, Prchal JT, Ge R. 2010. Genetic evidence for high-altitude adaptation in Tibet. *Science* **329**:72–75. DOI: https://doi.org/10.1126/science.1189406, PMID: 20466884

Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, Chisholm RL, Crosslin DR, Hebbring SJ, Jarvik GP, Kullo IJ, Li R, Pathak J, Ritchie MD, Roden DM, Verma SS, Tromp G, Prato JD, Bush WS, Akey JM, Denny JC, et al. 2016. The phenotypic legacy of admixture between modern humans and Neandertals. *Science* **351**:737–741. DOI: https://doi.org/10.1126/science.aad2149, PMID: 26912863

Stevens C, Lin Y, Sanchez M, Amin E, Copson E, White H, Durston V, Eccles DM, Hupp T. 2007. A germ line mutation in the death domain of DAPK-1 inactivates ERK-induced apoptosis. *The Journal of Biological Chemistry* **282**:13791–13803. DOI: https://doi.org/10.1074/jbc.M605649200, PMID: 17244621

Sugden AM. 2018. Earliest modern humans out of Africa. *Science* **359**:407.. DOI: https://doi.org/10.1126/science.359.6374.407-p

Vahdati AR, Weissmann JD, Timmermann A, Ponce de León M, Zollikofer CPE. 2022. Exploring Late Pleistocene hominin dispersals, coexistence and extinction with agent-based multi-factor models. *Quaternary Science Reviews* **279**:107391. DOI: https://doi.org/10.1016/j.quascirev.2022.107391

Vernot B, Akey JM. 2014. Resurrecting surviving neandertal lineages from modern human genomes. *Science* **343**:1017–1021. DOI: https://doi.org/10.1126/science.1245938

Vichaiwong K, Purohit S, An D, Toyoda T, Jessen N, Hirshman MF, Goodyear LJ. 2010. Contraction regulates site-specific phosphorylation of TBC1D1 in skeletal muscle. *The Biochemical Journal* **431**:311–320. DOI: https://doi.org/10.1042/BJ20101100, PMID: 20701589

Wang Y, Zou X, Wang M, Yuan D, Yang L, Zeng Y, Cheng F, Tang R, He G. 2022. The genomic history of southwestern Chinese populations demonstrated massive population migration and admixture among proto-Hmong-Mien speakers and incoming migrants. *Molecular Genetics and Genomics* **297**:241–262. DOI: https://doi.org/10.1007/s00438-021-01837-3, PMID: 35031862

Xu S, Li S, Yang Y, Tan J, Lou H, Jin W, Yang L, Pan X, Wang J, Shen Y, Wu B, Wang H, Jin L. 2011. A genome-wide search for signals of high-altitude adaptation in Tibetans. *Molecular Biology and Evolution* **28**:1003–1011. DOI: https://doi.org/10.1093/molbev/msq277, PMID: 20961960

Yang J, Jin ZB, Chen J, Huang XF, Li XM, Liang YB, Mao JY, Chen X, Zheng Z, Bakshi A, Zheng DD, Zheng MQ, Wray NR, Visscher PM, Lu F, Qu J. 2017. Genetic signatures of high-altitude adaptation in Tibetans. *PNAS* **114**:4189–4194. DOI: https://doi.org/10.1073/pnas.1617042114, PMID: 28373541

Yang X-Y, Rakha A, Chen W, Hou J, Qi X-B, Shen Q-K, Dai S-S, Sulaiman X, Abdulloevich NT, Afanasevna ME, Ibrohimovich KB, Chen X, Yang W-K, Adnan A, Zhao R-H, Yao Y-G, Su B, Peng M-S, Zhang Y-P. 2021. Tracing the genetic legacy of the tibetan empire in the balti. *Molecular Biology and Evolution* **38**:1529–1536. DOI: https://doi.org/10.1093/molbev/msaa313, PMID: 33283852

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, Zheng H, Liu T, He W, Li K, Luo R, Nie X, Wu H, Zhao M, Cao H, Zou J, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**:75–78. DOI: https://doi.org/10.1126/science.1190371, PMID: 20595611

Zhang B, Xu R, Zhang J, Zhao X, Wu H, Ma L, Hu J, Zhang J, Ye Z, Zheng X, Qin Y. 2013. Identification and functional analysis of a novel PRKAG2 mutation responsible for Chinese PRKAG2 cardiac syndrome reveal an important role of non-CBS domains in regulating the AMPK pathway. *Journal of Cardiology* **62**:241–248. DOI: https://doi.org/10.1016/j.jjcc.2013.04.010, PMID: 23778007

Zhang C, Lu Y, Feng Q, Wang X, Lou H, Liu J, Ning Z, Yuan K, Wang Y, Zhou Y, Deng L, Liu L, Yang Y, Li S, Ma L, Zhang Z, Jin L, Su B, Kang L, Xu S. 2017. Differentiated demographic histories and local adaptations between Sherpas and Tibetans. *Genome Biology* **18**:115. DOI: https://doi.org/10.1186/s13059-017-1242-y, PMID: 28619099

Zhang J, Xu J, Dong Y, Huang B. 2018. Down-regulation of HIF-1α inhibits the proliferation, migration, and invasion of gastric cancer by inhibiting PI3K/AKT pathway and VEGF expression. *Bioscience Reports* **38**:BSR20180741. DOI: https://doi.org/10.1042/BSR20180741

Zhang X, Witt KE, Bañuelos MM, Ko A, Yuan K, Xu S, Nielsen R, Huerta-Sanchez E. 2021. The history and evolution of the Denisovan- *EPAS1* haplotype in Tibetans . *PNAS* **118**:1–9. DOI: https://doi.org/10.1073/pnas.2020803118

Zhang X, Kim B, Singh A, Sankararaman S, Durvasula A, Lohmueller KE. 2023. MaLAdapt reveals novel targets of adaptive introgression from neanderthals and denisovans in worldwide human populations. *Molecular Biology and Evolution* **40**:msad001. DOI: https://doi.org/10.1093/molbev/msad001, PMID: 36617238

Zheng W, He Y, Guo Y, Yue T, Zhang H, Li J, Zhou B, Zeng X, Li L, Wang B, Cao J, Chen L, Li C, Li H, Cui C, Bai C, Qi X, Su B. 2023. Large-scale genome sequencing redefines the genetic footprints of high-altitude adaptation in Tibetans. *Genome Biology* **24**:73. DOI: https://doi.org/10.1186/s13059-023-02912-1, PMID: 37055782

communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-025-07813-6

Convergent evolution of complex adaptive traits modulates angiogenesis in high-altitude Andean and Himalayan human populations

Giulia Ferraretti[1,18], Aina Rill[1,14,18], Paolo Abondio[2,15,18], Kyra Smith[1], Claudia Ojeda-Granados[1,16], Sara De Fanti[3], Marta Alberti[1], Massimo Izzi[4], Phurba T. Sherpa[5], Paolo Cocco[6], Massimiliano Tiriticco[6], Marco Di Marcello[6], Agnese Dezi[7], Guido Alberto Gnecchi-Ruscone[8,17], Luca Natali[6,9], Angela Corcelli[10], Giorgio Marinelli[6], Paolo Garagnani[11,12], Davide Peluzzi[6], Donata Luiselli[2], Davide Pettener[1], Stefania Sarno[1,19] & Marco Sazzini[1,13,19] ✉

Convergent adaptations represent paradigmatic examples of the capacity of natural selection to influence organisms' biology. However, the possibility to investigate the genetic determinants underpinning convergent complex adaptive traits has been offered only recently by methods for inferring polygenic adaptations from genomic data. Relying on this approach, we demonstrate how high-altitude Andean human groups experienced pervasive selective events at angiogenic pathways, which resemble those previously attested for Himalayan populations despite partial convergence at the single-gene level was observed. This provides additional evidence for the drivers of convergent evolution of enhanced blood perfusion in populations exposed to hypobaric hypoxia for thousands of years.

Convergent evolution refers to an evolutionary scenario in which different species or populations independently develop the same (or similar) biological trait instead of inheriting it from a common ancestor. In particular, convergent adaptation occurs when distantly related species/populations, which live in distinct geographical areas or epochs, occupy comparable ecological settings and are subjected to the same selective pressures, thus being similarly targeted by natural selection[1]. Several cases of convergent evolution have been described so far in a number of animal and plant taxa[1–4], with fewer examples (e.g., skin pigmentation[5], malaria resistance[6], lactose tolerance[7] and metabolic adaptations to cold climates[7]) being instead reported in the human species when comparing adaptive traits among populations of different ancestry. Nevertheless, experimental approaches to accurately characterise genomic variability within and between species/populations, as well as inferential statistical methods suitable to investigate how far phenotypic convergence is achieved by means of genetic convergence, have been only recently conceived. This is especially the case of

[1]Laboratory of Molecular Anthropology & Centre for Genome Biology, Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy. [2]Department of Cultural Heritage, Ravenna Campus, University of Bologna, Ravenna, Italy. [3]IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy. [4]Complex Operative Unit of Endocrinology and Diabetes Care, Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy. [5]Mount Everest Summiters Club, Rolwaling, Dolakha, Nepal. [6]Explora Nunaat International, Montorio al Vomano, Teramo, Italy. [7]Department of Precision and Regenerative Medicine and Ionian Area, University of Bari Aldo Moro, Bari, Italy. [8]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [9]Italian Institute of Human Paleontology, Rome, Italy. [10]Department of Translational Biomedicine and Neuroscience, University of Bari Aldo Moro, Bari, Italy. [11]IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy. [12]Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy. [13]Interdepartmental Centre Alma Mater Research Institute on Global Changes and Climate Change, University of Bologna, Bologna, Italy. [14]Present address: Josep Carreras Leukaemia Research Institute, PhD Programme in Biomedicine, University of Barcelona, Barcelona, Spain. [15]Present address: Department of Biology, University of Rome Tor Vergata, Rome, Italy. [16]Present address: Department of Medical and Surgical Sciences and Advanced Technologies "GF Ingrassia", University of Catania, Catania, Italy. [17]Present address: Archaeo- and Palaeogenetics, Institute for Archaeological Sciences, Department of Geosciences & Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany. [18]These authors contributed equally: Giulia Ferraretti, Aina Rill, Paolo Abondio. [19]These authors jointly supervised this work: Stefania Sarno, Marco Sazzini. ✉e-mail: marco.sazzini2@unibo.it

complex (i.e., polygenic) adaptive traits, whose evolution is regulated by changes at several genes that are functionally related and contribute to the modulation of the same biological function[8]. Development of methods able to test the occurrence of selective events under a realistic approximation of a polygenic adaptation model has thus opened new possibilities to assess whether natural selection acted on the same genetic variants or genes or functional pathways in species/populations showing shared or similar adaptive traits[9].

To this end, human adaptation to the high-altitude environment represents a valuable case study because the main selective pressure acting on populations dwelling at altitude (i.e., hypobaric hypoxia) cannot be mitigated by cultural adaptations and thus acts with the same intensity on human groups living at comparable altitudes irrespectively to their ancestry, geographical locations and socio-cultural contexts. We can thus hypothesise that Himalayan and Andean high-altitude populations, which represent the main human groups whose ancestors have had to cope with such a selective pressure, experienced convergent evolution at least to some extent. Despite this assumption, limited overlapping of potential adaptive loci (e.g., entailing variants at the *EGLN1*, *EDNRA* and *EPAS1* genes) has been described so far between them[10–13]. This may be partially due to the limitations imposed by studies that to date focused mainly on selective sweeps. Indeed, these adaptive mechanisms are mediated by the action of natural selection on single/few genetic variants at the same gene, which implies that these loci can exert a large effect on a given biological trait[8]. That being so, selective sweeps have low probability to occur in human populations, who have long maintained low effective population sizes and have accumulated low genetic variability and thus need several thousands of years to evolve[14]. Coupled with the considerably more recent human colonisation of the high-altitude ecological niche in the Andes than in the Himalayas, as supported by estimates at around 9 thousand years ago (kya) for the divergence between Andean and lowland South American populations[15] and by archaeological evidence for occupation of the Tibetan Plateau since Palaeolithic times[16], this suggests selective sweeps might be not the predominant drivers of adaptation of Andean people to such a challenging environment.

Consistent with this scenario, Andeans indeed evolved less effective genetically regulated physiological adjustments than populations of Tibetan/Sherpa ancestry, especially to counteract long-term detrimental effects of hypobaric hypoxia[17,18]. When compared with acclimatized Native American lowlanders, they exhibit traits contributing to enhanced efficiency of respiratory functions and oxygen utilisation, such as larger lung volumes and chest dimensions, narrower alveolar to arterial oxygen gradients, less hypoxia-induced pulmonary vasoconstriction, greater uterine artery blood flow during pregnancy, less altitude-related decrement in maximal exercise oxygen consumption and increased cardiac oxygen utilisation[17,18]. However, when genetic determinants of these physiological characteristics have been searched for, most of the inferred selective sweeps failed to be replicated by multiple studies, suggesting that their relationships with the observed adaptive traits are far from being fully elucidated[15,18–21].

Results and discussion

To test whether polygenic adaptations played a role in shaping the Andean adaptive phenotype, and to extend to complex traits the investigation of the genetic bases of potential convergent evolution between Andean and Himalayan populations, we assembled a dataset representative of genomic variation at several high-altitude ethnic groups from the Andean region (Supplementary Table 1). After stringent quality control (QC) filtering (see 'Methods'), 7,966,198 Single Nucleotide Variants (SNVs) from 24 Bolivian Aymara whole genome sequences (WGS)[15] were used to perform multiple selection scans (i.e., H12, nSL and LASSI) that collectively are able to test for the occurrence of a range of selective events (see 'Methods' for further details concerning the rationale behind the choice of these inferential approaches). The obtained results then informed gene-network analyses aimed at identifying sets of functionally related loci enriched for signatures of natural selection. Furthermore, we applied the same pipeline of analyses on: (i) imputed genotype data for 8,024,216 SNVs from 130 individuals belonging to additional Bolivian Aymara groups, as well as from Aymara, Quechua, and Uros high-altitude Peruvian populations[22,23] used as validation datasets and (ii) a control dataset including 24 individuals from three different Amazonian low-altitude groups (i.e., Ashaninka, Cashibo, and Shipibo)[23] to filter out adaptive loci shared between high- and low-altitude South American populations, with the aim of focusing exclusively on those typical of Andean peoples.

Identification of Native American low-altitude control populations

Setting the considered Bolivian Aymara WGS into the genomic landscape of indigenous groups from Central and South America enabled us to confirm them as well representative of the genetic variation observable across Andean high-altitude populations (Fig. 1a, Supplementary Information). Comparable results were obtained also when considering imputed data for Andeans, thus providing evidence for consistency between non-imputed and imputed datasets (Supplementary Fig. 1). Such a consistency was further attested by the high and significant correlation observed between pre- and post-imputation identity-by-state (IBS) matrices of individual pairwise differences in the genome-wide proportion of shared alleles (Mantel correlation = 0.95, $P < 10^{-4}$).

The outgroup-f3 statistic was then computed on the assembled reference dataset to select the most suitable low-altitude control populations of South American ancestry to be contrasted with Andean groups in terms of evolved biological adaptations (see 'Methods' for a detailed description of the adopted selection criteria). According to this approach, while each Andean population was confirmed to share the greatest genetic similarity with the other Andean groups, Amazonian populations from Peru appeared as the next most affine ones to Andeans (Fig. 1b, Supplementary Fig. 2 and Supplementary Information). Among these putative control populations, eight Ashaninka, six Cashibo and 10 Shipibo subjects were specifically selected to constitute the low-altitude control group according to their clustering patterns and negligible Andean-specific genetic component, as pointed out by CHROMOPAINTER/fineSTRUCTURE and ADMIXTURE estimates respectively of haplotypes sharing and individual ancestry proportions (Fig. 1c and Supplementary Information).

Genomic signatures ascribable to polygenic adaptive events in Andean populations

At first, genome-wide distributions of nSL and H12 statistics, which are informative of both strong and weak selective events, were calculated for Andean and control groups relying on both WGS and imputed data. Then, we used these distributions as input for the *signet* algorithm to infer gene networks made up of loci belonging to the same functional pathway and that have been concurrently targeted by natural selection (see 'Methods'). By considering only pathways confirmed by each selection statistic and by both WGS and imputed datasets, we shortlisted the most plausible biological functions having evolved adaptively in the considered populations.

Significant gene-networks identified for the low-altitude control group were found to participate primarily to immune-related pathways (e.g., *Staphylococcus aureus infection*, *Human papillomavirus infection* and *Arachidonic acid metabolism*), suggesting the evolution of adaptations able to mediate immune and inflammatory responses to pathogens (Supplementary Table 2). Similar selective pressures seem to have acted also on Andean groups, as attested by their putative adaptive loci belonging to the *Purine metabolism* and *Herpes simplex virus 1 infection* pathways (Fig. 2a and Supplementary Table 3).

Conversely, adaptive events plausibly contributing to the response to stresses imposed by the high-altitude environment were inferred only in Andeans and were mediated by genes involved in the *Focal adhesion* pathway, which regulates cells adhesion to extracellular matrix and subsequent signal transduction at the intracellular level (Fig. 2a, b and Supplementary Table 3). Most of the loci constituting the significant networks identified by analysing both Aymara WGS and Andean imputed datasets

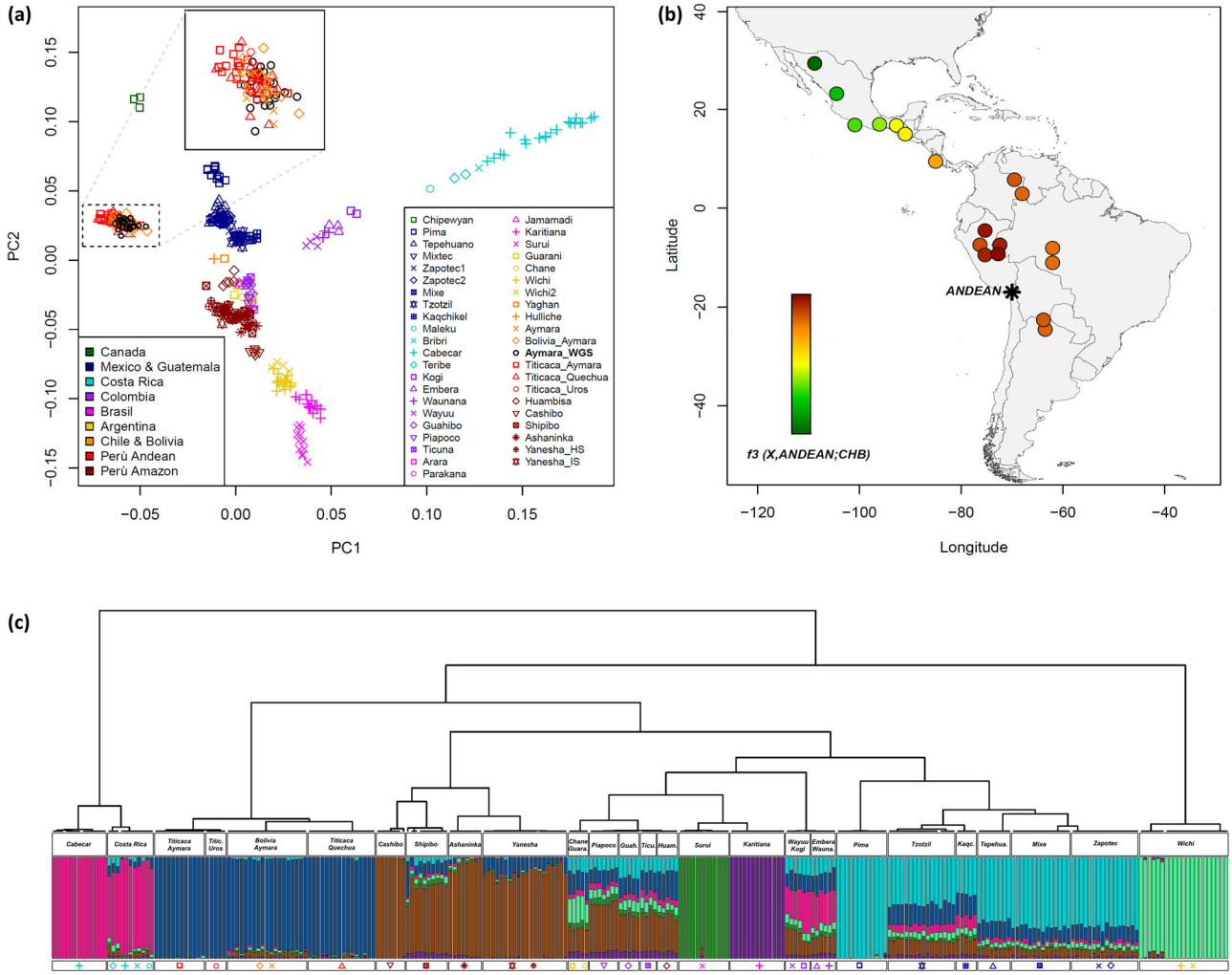


Fig. 1 | Population structure analyses performed to investigate genetic relationships between high-altitude Andean groups and low-altitude Central and South American populations. a PCA representing PC1 versus PC2 computed for the 43 un-admixed Native American groups reported in the legend at the bottom right of the plot. Individuals are colour-coded according to their country of origin as reported in the bottom-left legend. The box at the top of the plot details the position of Bolivian Aymara WGS in the considered genetic space with respect to the other high-altitude Andean populations. **b** Heat map of values for the outgroup-f3 statistic representing the estimated amount of shared genetic drift between the Andean population cluster (marked by a star-like symbol) and each of the considered populations from Central and Southern America (indicated in their approximate geographic locations) A gradient ranging from green to red is specifically used to indicate lower-to-higher levels of genetic affinity as reported by the corresponding colour scale. **c** Fine-scale patterns of genetic clustering and proportions of ancestry components inferred at the individual level. The fineSTRUCTURE hierarchical clustering dendrogram obtained for Andean, Central and South American individuals included in the assembled dataset is reported, along with ancestry components inferred for each subject with ADMIXTURE analysis at $K = 8$. For each genetic cluster pinpointed by the fineSTRUCTURE analysis, the corresponding individuals and their admixture proportions are specifically detailed, along with the symbols used in the PCA plot for the populations they belong to. The source data used to generate the figures are reported in Supplementary Data 2—Tables 1–3.

are indeed known to play a pivotal role in angiogenetic processes (Fig. 2a, b). Especially five of these genes (i.e., *LAMA3*, *COL4A1*, *ITGB6*, *PRKCB* and *PTK2*) represent the most robust candidate targets of positive selection due to patterns ascribable to polygenic adaptive evolution identified according to both nSL-based and H12-based network analyses (Fig. 2b). Among them, *ITGB6*, *PRKCB* and *PTK2* have been proved to be the main regulators of vascular sprouting, which implies the transformation of blood vessels endothelial cells into cells able to migrate and proliferate out of the vessel to create new vascular structures[24–31]. In fact, *PTK2* inactivation has been demonstrated to prevent the beginning of the angiogenetic process, leading to a lack of development of vascular tissue in embryos and thus to their premature death[32,33]. Although overall adaptive evolution of loci at the *Focal adhesion* pathway was highlighted by significant networks obtained according to both nSL and H12 statistics, the remaining 28 genes included in these networks were pointed out as potentially adaptive by a single statistic. Nevertheless, their functional relationships with the highly confirmed loci described above have been well established and they are known to contribute to the regulation of blood vessels formation as well (Fig. 2a, b). For instance, multiple studies collected evidence supporting the interaction among *PTK2*, *VEGFA*, *PDGFA* and *PDGFD* as a key driver of blood vessels formation in placenta and some variants at these loci are found to be associated with reduced functionality of maternal-foetal circulation, ischaemic placental disease and impaired foetal growth[34,35].

To further validate these findings, we run the *signet* algorithm also on genome-wide distributions of values obtained by applying the LASSI method respectively on the WGS, imputed and low-altitude control datasets. This inferential approach was indeed proven to outperform H12 and nSL statistics in the detection of weak selective events, being also able to accurately distinguish them from hard selective sweeps[36] (see 'Methods' for further details). After having filtered out putative adaptive gene networks shared among high-altitude and low-altitude groups, LASSI-based network

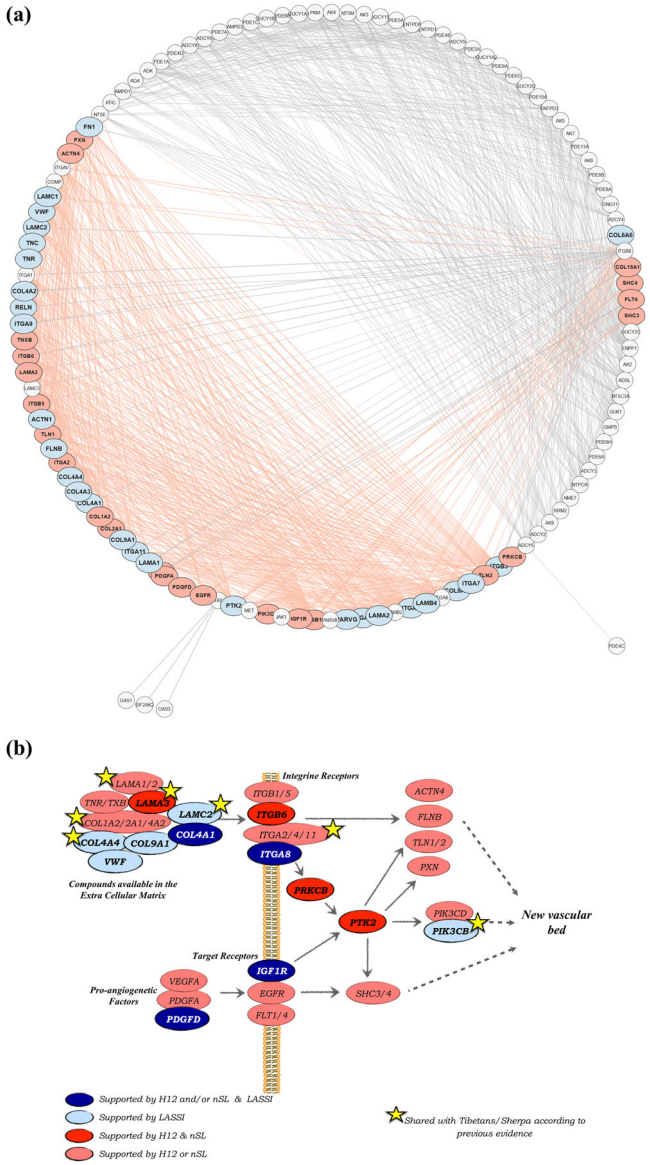


Fig. 2 | Functional relationships between genes putatively mediating polygenic adaptations evolved by Andean populations and their role in angiogenic processes. a Circular network build with Cytoscape v3.10.3 and displaying protein-protein functional interactions inferred for the entire set of adaptive genes belonging to significant pathways identified by H12-based and nSL-based *signet* analyses according to both WGS and imputed datasets. Genes pointed out by LASSI-based *signet* analyses as participating to the *Focal Adhesion/PI3K-Akt signalling* pathways are also reported. Gene products that establish similar connections in the network are placed next to each other, comporting the delineation of two well-distinguishable functional clusters. In detail, non-angiogenesis related genes (e.g., belonging to the *Herpes simplex virus 1 infection* pathway) are reported as light-grey circles, while angiogenesis-related genes (i.e., belonging to the *Focal adhesion/PI3K-Akt signalling* pathways) are displayed with pink (when supported by H12 and/or nSL) and light blue (when supported by LASSI) ellipses. The source data used to generate the plot is reported in Supplementary Data 2—Table 4. **b** Scheme displaying functional interactions between *PI3K-Akt signalling* and *Focal adhesion* adaptive genes contributing to improved angiogenesis in Andean populations according to the KEGG database. During the early phases of angiogenesis molecular compounds made available in the extracellular matrix, such as those building the basement vascular membrane (e.g., *COL4A1*) and angiogenetic factors (e.g., *VEGFA* and *PDGFD*), interact with integrin and target receptors (e.g., *IGF1R*) stimulating protein kinases (e.g., *PRKCB* and *PTK2*), which subsequentially activate those signalling cascades essential for migration of endothelial cells and formation of new vascular structures. The yellow stars mark those genes previously identified as loci putatively mediating polygenic adaptation to hypobaric hypoxia in Tibetan/Sherpa populations[52,61,62], thus providing evidence for partial genetic convergence between Andean and Himalayan adaptive traits in addition to the remarkable convergence observed at the biological function/pathway level.

analyses pointed out 39 genes plausibly targeted by positive selection and supported by results for both WGS and imputed datasets (Supplementary Fig. 3). Among them, 27 loci were found to belong to functional pathways related to angiogenesis promotion, cell proliferation and vasodilatation (Supplementary Fig. 3), with four of them (i.e., *COL4A1*, *IGF1R*, *PDGFD*, *ITGA8*) being included in the previously described significant networks belonging to the *Focal adhesion* pathway pointed out by H12-based and nSL-based analyses (Fig. 2b). Another significant *Focal adhesion* network was supported only by the Aymara WGS dataset (Fig. 2b and Supplementary Table 4), with some genes (i.e., *COL4A1*, *COL4A4*, *COL9A1*, *LAMC2*, *VWF* and *ITGA8*) participating also to a network belonging to a functionally related pathway (i.e., *PI3K-Akt signalling*) detected according to imputed data (Supplementary Fig. 3 and Supplementary Table 4). These results contribute to corroborate the identification of the *Focal adhesion pathway* and of associated molecular cascades as complex biological functions that have been shaped by the action of natural selection in Andean highlanders.

Among the adaptive genes validated by such an approach, *COL4A1* encodes a type IV collagen alpha protein that is one of the main constituents of the vascular basement membrane[29,37], whose degradation and remodelling represent the very first steps of angiogenesis induced by *VEGF*, *FGF* and *PDGF* growth factors[37]. Interestingly, mutant *COL4A1* mice present retinal and subretinal neovascular lesions with respect to wild-type ones[38], while human *COL4A1* and *COL4A2* variants have been associated with cerebrovascular disease[29] and intracerebral haemorrhages due to systemic small-vessel disease[39], suggesting that *COL4A* proteins play a crucial role in modulating the formation of new vascular structures. The *IGF1R* gene instead encodes for the insulin like growth factor 1 receptor that binds *IGF1* angiogenetic factors and is known to be overexpressed in most malignant tissues where it acts as an anti-apoptotic agent by enhancing cell survival[29,40]. In particular, remodelling of tumour blood vessels mediated by *IGF1* was observed in the melanoma mouse model[40], with *IGF1* and *IGF1R* deficiency being also associated to severe postnatal growth retardation in mice[41–43]. Similarly, homozygous partial *IGF1* deletion in humans causes intrauterine growth retardation, which is accompanied by low average placental weight[44]. Finally, the platelet-derived growth factor *PDGF-D* represents another angiogenetic factor whose activity promotes tumour vasculogenesis in mice due to *VEGF* upregulation[45]. Moreover, *PDGFD* down-regulation was also observed in the placenta of women with preeclampsia, a syndrome characterised by high resistance utero-placental circulation, placental hypoperfusion and elevated expression of antiangiogenetic factors[46].

In line with this evidence, two additional significant gene networks participating to the *Apelin signalling* and *Vascular smooth muscle contraction* pathways were proposed to have adaptively evolved in Andeans, but not in the low-altitude control group, according to both WGS and imputed data. However, these signatures did not reach statistical significance in H12/nSL-based network analyses, being detected only by LASSI-based ones and including the *GNG7*, *PRKCE*, *RYR2* and *RYR3* loci (Supplementary Table 4 and Supplementary Fig. 4). In particular, *RYR2* and *RYR3* codify for two ryanodine receptors (RyR): the former is expressed in the sarcoplasmic reticulum of the cardiac muscle[29] and is essential for contraction regulation mediated by calcium release[47,48], while the latter is mainly found in skeletal muscle, as well as in the heart Purkinje tissue, thus possibly contributing to electric stability of these fibres in such an organ[49]. When compared with acclimatised lowlanders, native Andean individuals are found to maintain heart rate unaltered during apnoea[50], similarly to what observed for Himalayan people[51] and suggesting that adaptive modulation of hearth frequency and cardiac output might result in a decreased risk of developing arrhythmias in these high-altitude populations.

Convergent adaptive evolution between Andean and Himalayan populations

Interestingly, when we previously applied a gene network-based approach to WGS from Tibetan and Sherpa populations, several genes included in Andean significant networks obtained from H12/nSL analyses (e.g., *LAMA1, COL4A2* and *ITGA2*) were similarly found to participate to functional pathways targeted by natural selection (Fig. 2b, Supplementary Table 5)[52]. This pattern was basically confirmed by results from LASSI-based analyses. First, this approach recapitulated previous evidence supporting convergent adaptations between Andean and Tibetan highlanders involving the *EPAS1*[10] and *EGLN1*[13] genes, with genomic windows at these loci being classified as targets of selective sweeps and showing likelihood values falling in the top 1% of the distribution obtained for the Aymara WGS dataset.

As regards instead genes included in significant networks, *LAMC2* was already proposed to have contributed to polygenic adaptive events also in Tibetan/Sherpa populations[52] (Fig. 2b, Supplementary Table 5). This gene codifies for the laminin gamma 2 subunit, which is an epithelium-specific basement membrane protein highly expressed in different types of cancers and tumour metastasis[53–55]. Silencing *LAMC2* expression was proved to cause cell cycle arrest and significantly suppresses migration and invasion of tumour cells, as well as in vivo angiogenesis of the malignant tissue due to partially blocked activation of the *EGF* receptor and its downstream pathway[54,55]. Moreover, *LAMC2* expression in the placenta predominantly interests syncytiotrophoblasts and cytotrophoblasts villous and was found to be substantially upregulated in Placental Accreta Spectrum (PAS) tissues, stimulating trophoblast over-invasion via *PI3K/Akt/MMP2/9* signalling pathways[56]. In fact, PAS is a condition causing pathologic adherence of the placenta due to abnormal trophoblast neovascularization and invasion into the uterine wall[57,58]. This invasion has been related to improved stimulation of pro-angiogenetic factors not only in the trophoblast embryotic tissue, but also in the maternal basal plate of the placenta[58,59]. We can thus hypothesise that adaptive evolution at *LAMC2* and functionally related loci observed in the considered high-altitude populations might result in promoting angiogenesis in the placental tissues.

Other genes that we propose to have been targeted by natural selection in Andeans (i.e., *COL4A4*, *PLCE1*, *PIK3CB* and *PRKCE*) were previously described as plausible adaptive loci in Tibetan groups[13,60–62] (Fig. 2b and Supplementary Table 5). Again some of them are known to be associated to inhibition of apoptosis and angiogenesis in tumours (*PLCE1*)[63] or to cardiovascular traits (*PRKCE*)[64]. In particular, *PRKCE* has been demonstrated to exert a cardio-protective role against ischaemic injury[64] and presents variation patterns ascribable to both adaptive evolution and archaic introgression in Tibetan populations[12,60,65]. Other genes instead seem to be involved in the modulation of decidua or in the decidualization phase in preparation to pregnancy[66,67]. For instance, *COL4A4* expression, along with that of other proteins belonging to the collagen type IV family, increases during decidualization of the human endometrium[67] participating to the regulation of those morphological and functional changes (e.g., increased vascular permeability, oedema, invasion of leucocytes, vascular remodelling and angiogenesis) that are essential for the establishment of a successful pregnancy[68,69]. Similarly, *PIK3CB* turned out to be up-regulated in the decidua of early-onset pre-eclampsia versus normal pregnancies, plausibly playing a role in abnormal placental development[66].

According to this picture, the obtained results revealed that changes at functional pathways enabling generalised and/or placental improvement of angiogenesis, as well as cardiovascular protection, seem to have characterised the evolutionary history of both Andean and Himalayan populations[51,52,65]. Nevertheless, an incomplete genetic convergence between these high-altitude peoples was observed. This is not unexpected within the framework of an evolutionary scenario largely characterised by polygenic adaptive events, in which single loci play a negligible role with respect to their overall synergic interactions. Overall, we provided evidence for the same biological functions having adaptively evolved in these populations, plausibly in response to the selective pressure imposed by hypobaric hypoxia.

Putative adaptive genes associated with complex traits and eQTLs

To finally test whether the identified candidate adaptive genes are associated to complex physiological and/or pathological traits possibly implicated in the adaptive responses evolved by the considered high-altitude populations, we first consulted the Common Metabolic Disease Knowledge Portal (CMDKP) (Supplementary Table 6) (see 'Methods' for further details). Furthermore, we searched for expression quantitative trait loci (eQTLs) that could be able to modulate the expression of these genes in specific tissues. For this purpose, we relied on information stored in the GTEx portal and we verified if eQTLs at each gene are located in genomic windows showing large positive values of the likelihood statistic calculated by LASSI (i.e., those chromosomal regions representing the most robust targets of natural selection). The proportions of eQTLs affecting the expression of candidate genes in tissues potentially involved in the modulation of biological adaptations to high altitudes (i.e., HA-eQTLs) were reported in Supplementary Table 7, while HA-eQTLs falling in LASSI candidate adaptive windows were shortlisted in Supplementary Table 8.

According to such an approach, the most compelling results were obtained for the *COL4A1*, *IGF1R*, *PDGFD, ITGA8, RYR3, LAMC2, COL4A4*, *PLCE1* and *PRKCE* genes, which turned out to be associated to haematological (e.g., haemoglobin concentration and haematocrit) and/or cardiovascular (e.g., heart rate and stroke) traits. Moreover, they included several HA-eQTLs that have been proved to influence gene expression in the aorta, heart, lung, skeletal muscle, blood and uterus (Supplementary Tables 6–8 and Fig. 3). Among them, *LAMC2*, *PLCE1* and *PRKCE* showed the greatest proportion of HA-eQTLs with respect to the total number of annotated eQTLs (Supplementary Table 7). Remarkably, 100 eQTLs in the *LAMC2* gene, 44 in *PLCE1* and 22 in *PRKCE* were included in LASSI candidate adaptive windows, modulating their expression respectively in the uterus/aorta/atrial appendage, in lungs/artery tibial/atrial appendage and in the skeletal muscle and whole blood tissues (Fig. 3, Supplementary Table 8). Similarly, 51 eQTLs on the *RYR3* gene, 11 on *COL4A1*, eight on *IGF1R* and four on *PDGFD* and *ITGA8* were included in adaptive genomic windows and regulate the expression of these loci in skeletal muscle, lungs, aorta, heart and blood tissues (Fig. 3, Supplementary Table 8).

Overall, this evidence further reinforces the functional links observed between the inferred candidate adaptive genes and complex mechanisms that have been long proposed to be finely regulated in both Andean and Tibetan populations[70–72]. Additionally, they highlight possible associations of these adaptive changes with decreased risk of native Andean and Tibetan/Sherpa populations of developing pathological conditions that are instead more frequent in lowlanders exposed to the high-altitude stresses[50–52,65].

Conclusions

The identified genetic signatures might contribute to the increase in the density of blood vessels that results in an improved blood flow and oxygen delivery to tissues despite the hypoxic stress, especially leading to increased uterine blood flow during pregnancy, which is an adaptive trait qualitatively (but not quantitatively) similar between Andean and Himalayan groups[52,73,74]. Although such a trait appears to be more enhanced in Tibetan/Sherpa populations with respect to Andean ones[13,17,75], the fact that several candidate loci identified in the present study (e.g., *PTK2*, *VEGFA*, *PDGFA*, *PDGFD, COL4A4, PIK3CB, IGF1R* and *LAMC2*) play a pivotal role in the development of vascular tissue specifically in placenta and embryo, and thus in the establishment of efficient maternal-foetal circulation[32–35,56,66,67], seems to support such an observation. Moreover, it might suggest that convergent adaptive evolution succeeded in optimising angiogenesis in these human groups mainly in early life and/or at the reproductive stage, thus leading to improved foetal development during intrauterine life (e.g., in terms of maturation of the respiratory system), which represents a key aspect to reduce neonatal mortality at high altitudes where efficiency of respiratory functions is crucial to ensure the individual's survival. In addition, results obtained for the *PRKCE*, *RYR2* and *RYR3* genes

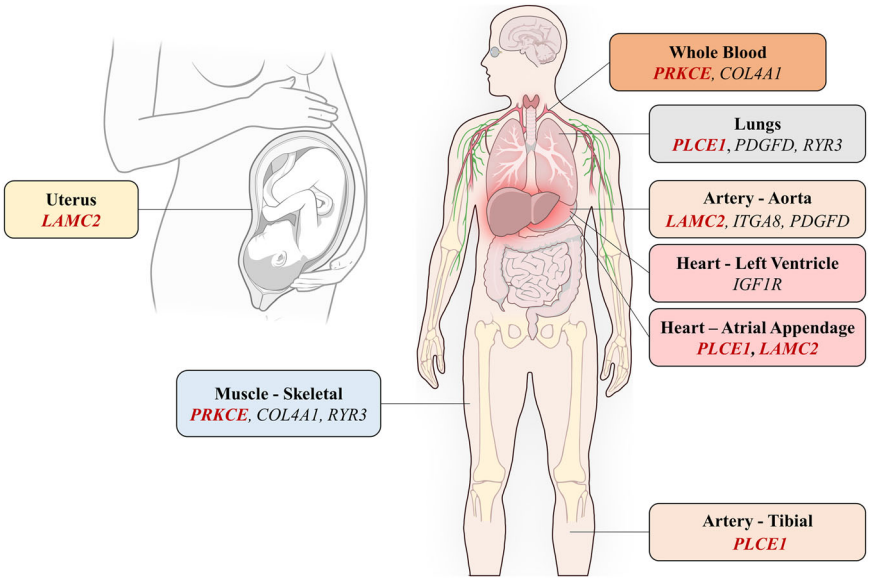


Fig. 3 | Candidate genes presenting eQTLs in putative adaptive genomic windows identified by the LASSI method and able to influence their expression in specific tissues. Schematic representation of the body districts in which the expression of the *COL4A1*, *IGF1R*, *PDGFD*, *ITGA8*, *RYR3*, *LAMC2*, *PLCE1* and *PRKCE* candidate adaptive genes is modulated by eQTLs annotated in the GTEx portal. Genes reported in bold/red are those previously proposed to play an adaptive a role also in populations of Tibetan/Sherpa ancestry[52,60,62,65]. Illustration from NIAID NIH BIOART Source (https://bioart.niaid.nih.gov/bioart/519; https://bioart.niaid.nih.gov/bioart/420).

suggest that also some cardiovascular traits might have been shaped by natural selection in both Andean and Himalayan peoples[50,51]. That being so, these findings highlighted polygenic mechanisms mediated by adaptive evolution of several focal adhesion and/or functionally related genes as the possible drivers of enhanced angiogenesis, oxygen transport and cardiovascular protection in Andean human groups similarly to what was previously observed in Tibetan/Sherpa populations, thus providing additional evidence for the molecular bases of their convergent adaptation to hypobaric hypoxia.

Methods

Assembled datasets

A WGS dataset including 9,439,267 SNVs for 24 Bolivian Aymara individuals who were born and live at around 3900 m of altitude was assembled from previously generated data[15] and used to perform selection scans. Genome-wide genotype data were also collected for 141 individuals who were born and live between 3800 and 4200 m above sea level and belong to different Andean high-altitude populations that have been characterised by a largely shared genetic history[22,23]. A written informed consent for data treatment was signed by each participant as attested in previous works[15,22,23] and all ethical regulations relevant to human research participants were followed. Moreover, on 09/12/2019 the Ethics Committee of the University of Bologna released approval (prot. 205142) for the present study within the framework of the project 'Genetic adaptation and acclimatisation to high altitude as experimental models to investigate the biological mechanisms that regulate physiological responses to hypoxia' funded by Fondazione Cassa di Risparmio in Bologna (grant n. 2019.0552). In detail, 713,014 genetic variants were retrieved for 26 Bolivian Aymaras, as well as for 21, 22 and nine subjects respectively from Aymara, Quechua and Uros populations settled in the Peruvian Titicaca Lake area[23]. Genotypes for 364,470 loci were collected also for 40 Quechua and 23 Aymara additional individuals[22].

Data curation and imputation of the validation dataset

WGS data were submitted to QC procedures to filter for high-quality genotypes and to exclude possible closely related individuals. Specifically, the PLINK package v2.0[76] was used to retain only SNVs on autosomal chromosomes, variants/samples showing missing data <5% and loci with non-significant deviations from the Hardy-Weinberg equilibrium ($p > 1.059 \times 10^{-9}$). We also removed variants showing ambiguous A/T or G/C alleles and we calculated pairwise identity-by-descent (IBD) statistics by estimating the genome-wide proportions of shared alleles for each pair of individuals. Then, to exclude subjects related to the second degree, we filtered out one individual from each pair showing an IBD coefficient >0.270, as previously proposed for populations of Native American ancestry[77] (Supplementary Information). This led to the generation of a high-quality WGS dataset including 7,966,198 SNVs and 24 individuals.

The described QC filters were applied also to the two assembled genome-wide high-altitude Andean datasets[22,23] making available for validation analyses 681,932 and 364,413 variants, respectively. These data were merged by selecting only the 210,540 variants shared between both datasets and were phased by means of the Eagle algorithm v2.4[78] implemented in the Michigan Imputation Server (MIS) v1.2.4[79] using the 1000 Genomes Project Phase 3 and WGS datasets as reference panels. The MIS Minimac4 algorithm was then used to impute genotypes for 10,715,833 variants, which were submitted to the previously described QC procedures. In addition, the *bcftools* package v1.8[80] was used to retain only loci showing a squared Pearson correlation coefficient (r^2) between imputed genotype probability and true genotype call >0.95[81]. Finally, IBD kinship coefficients were calculated for each pair of individuals in the imputed dataset using the same approach described for WGS data. According to these post-imputation QC steps, a high-density imputed dataset including 130 Andean individuals characterised for 8,024,216 variants was obtained.

Assessing samples' genetic ancestry and consistency between non-imputed and imputed data

To verify that the collected Bolivian Aymara WGS were representative of the genetic variation observable across Andean high-altitude populations, we first explored their distribution within the overall genomic landscape of Native American populations. For this purpose, the high-quality WGS dataset was merged with a reference panel made up of genome-wide genotypes from un-admixed individuals from Central and South American indigenous populations[23] and was submitted to Principal Component Analysis (PCA). Concurrently, to test also for consistency between non-imputed and imputed data, we performed PCA again by substituting the imputed data to the original ones as concerns the Andean populations included in the used Native American reference panel. In doing so, the *smartpca* method implemented in the EIGENSOFT package v6.0.1[82] was used after having filtered the dataset for variants in high linkage disequilibrium (LD). LD pruning was obtained by considering sliding windows of 50 nucleotides sites in size and advancing by five loci at the time, as well as by removing variants for each pair showing $r^2 > 0.2$. Moreover, IBS-based matrices of individual pairwise differences in the genome-wide proportion of shared alleles were computed on both the non-imputed and imputed LD-pruned Andean datasets. These matrices were then compared by means of

the Mantel correlation test[83] using functions implemented in the R *vegan* package and empirically assessing statistical significance by running >10,000 permutations.

Selecting low-altitude control populations

To assess whether selection signatures detected in Andean people are informative about adaptations in response to high-altitude selective pressures, we selected low-altitude control populations of South American ancestry from the previously described reference panel of Native American groups[23] to be used for replicating selection scans. In doing so, our rationale was to identify South American groups that: (i) share an ancient common origin with Andeans, (ii) do not show evidence of remarkable admixture with Andeans occurred after the divergence from their common ancestors, (iii) did not experience high-altitude-related selective pressures during their evolutionary history. For this purpose, we calculated outgroup-*f3* statistic in the form of *f3*(X, Andean; CHB) to formally assess the extent of shared drift between Andean populations and the other Central and South American populations (X). This was done using the *qp3Pop* function implemented in the ADMIXTOOLS package v3.0[84] by including only groups with more than five individuals and first by considering Andean populations separately (i.e., Aymara, Bolivian Aymara, Titicaca Aymara, Titicaca Quechua, and Titicaca Uros, respectively), and then by considering them as a whole (i.e., ANDEAN group).

Moreover, to accurately select the individuals to be included in the control group, we further investigated the genetic relationships between Andeans and the putative control populations pointed out by outgroup-*f3* analyses. To this end, we inferred individual-based genetic relationships by reconstructing haplotype sharing patterns by means of the CHROMOPAINTER/fineSTRUCTURE clustering approach[85] and by estimating ancestry proportions for each subject using the ADMIXTURE model-based clustering algorithm[86] (Supplementary Information). According to these procedures, we retained 24 individuals from low-altitude populations (i.e., eight Ashaninka, six Cashibo and 10 Shipibo), which have been characterised for 612,305 SNVs.

Inferring polygenic adaptive events from WGS and imputed data

A combination of selection scans (i.e., nSL, H12 and LASSI methods)[36,87,88] suitable to detect both strong and weak selective events was applied to WGS, imputed and low-altitude control datasets. Collectively, these statistics were chosen to investigate genomic signatures left by the action of natural selection on the haplotype diversity of the examined populations, assuming that those potentially ascribable to polygenic adaptive mechanisms are represented by subtle changes in their haplotype structure and/or frequency spectrum rather than outstanding shifts in the frequency of single genetic variants. In detail, the nSL statistic was computed for each variant using *selscan* v.1.1.0b[89] and a 20,000 bp threshold for gap scale, 200,000 bp as the maximum gap length and 4500 consecutive loci as the maximum extension parameter[88]. The H12 statistic was instead calculated with an ad hoc script using sliding windows of 200,000 bp and advancing by one variant at the time as previously suggested[87]. Unstandardised values for each statistic were then normalised across multiple frequency bins by subtracting to each of them the bin average score and by dividing the resulting value by the related standard deviation. Finally, also the LASSI likelihood statistic was computed across sliding windows, but after having estimated the average number of SNVs included in all the haplotype blocks inferred for each dataset. For this purpose, we run the *--blocks* function implemented in the PLINK package v2.0[76] separately on WGS, imputed and low-altitude control datasets. Then, the LASSI algorithm was used to calculate a distorted haplotype frequency spectrum for each window and to associate to them a likelihood statistic informative of its conformity to a model of adaptive evolution (i.e., hard/soft sweep model) or neutral evolution. Calculation of the m parameter indicating the number of sweeping haplotypes (i.e., those carrying putative adaptive variants) was used to discriminate between hard (i.e., $m = 1$) and soft selective sweeps (i.e., $m > 1$). Such a method was used to validate results from H12/nSL analyses, because it has been proved to outperform them in the identification of weak selective events according to simulations performed by considering a vast range of selection coefficients and demographic models[36].

In order to test a realistic approximation of a polygenic adaptation model, we retained as the representative scores of a given gene, the highest normalised nSL and H12 values, as well as the highest LASSI score computed among those obtained for all variants/windows annotated on that gene. In doing so, the overall distribution of the LASSI statistic was filtered to select windows showing $m \geq 2$ to consider only weak selective events as potential contributors to polygenic adaptations. The resulting genome-wide distributions of selection scores were then submitted to gene-network analyses using the R *signet* algorithm[9] and by setting 20,000 iterations to generate null distributions of the highest scoring subnetwork (HSS) for each gene-network of a given size to be compared with the obtained HSS. This provided a p-value for all the identified gene networks and those <0.05 after controlling the false discovery rate (FDR) were used to point out combinations of multiple genes that participate to the same functional pathway (according to information from the KEGG database[24]) and that have been targeted by natural selection.

The most plausible biological functions having evolved adaptively in the Andean high-altitude groups were finally identified as those (i) supported by at least two out of the three selection scans implemented (i.e., nSL-based, H12-based and LASSI-based significant networks), (ii) replicated by applying the same pipeline of analyses to both WGS and imputed data, and (iii) not represented in significant networks obtained from the analysis of the low-altitude control dataset. Finally, to evaluate convergent evolution between Andean and Himalayan populations at a single-gene level, loci included in the supported networks were crosschecked with lists of genes whose variation patterns were previously proposed to have been shaped by the action of natural selection in Tibetan/Sherpa peoples[13,52,65].

Investigating association between candidate adaptive loci and complex traits/eQTLs

The CMDKP database (https://hugeamp.org/) was used to validate associations between biological functions/traits that have been previously proposed to be involved in adaptive responses to high-altitude selective pressures[51,52,65,70–72] and the candidate adaptive genes identified by gene-network analyses.

Furthermore, we used the GTEx portal (https://gtexportal.org/home/) to search for eQTLs at the identified candidate adaptive loci and we retained only those variants affecting their expression in tissues and/or districts of the body that have been previously proposed to be involved in the modulation of high-altitude adaptations[17,90].

Statistics and reproducibility

Statistical analyses were performed using the softwares PLINK v2.0, SHAPEIT2 v2.r790, bcftools v 1.8, Eagle v2.4, MIS v1.2.4, EIGENSOFT v6.0.1, ADMIXTOOLS v3.0, CHROMOPAINTERv2, fineSTRUCTURE fs2.1, ADMIXTURE v.1.22, selscan v.1.1.0b, R v.3.6.3, LASSI GitHub repository available at https://github.com/mdegiorgio/LASSI.git. Phenotypic associations, eQTLs and pathways information in the present study were obtained consulting the CMDKP database (https://hugeamp.org/), the GTEx portal (https://gtexportal.org/home/) and the KEGG database (http://www.genome.ad.jp/kegg/). The sample sizes of each population used in the analyses are reported in Supplementary Table 1. The source data behind the graphs in the paper are listed in Supplementary Data 2.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

WGS data that support the findings described in the present study are available at the NCBI Sequence Read Archive (accession n. PRJNA470966)[15]. Genotype data used for validation analyses are instead

available via figshare (https://figshare.com/articles/dataset/South_American_dataset_Gnecchi-Ruscone_et_al_2019_/7667174)[23]. We provided all the Supplementary Tables cited in the paper in Supplementary Data 1. The source data behind the graphs in the main text are listed in Supplementary Data 2.

Code availability

All the codes used in the present study are available as functions that were either implemented in published R packages, GitHub public repositories or have been previously provided at https://github.com/paoloabondio/Ojeda-Granados_et_al_2021.

Received: 11 March 2024; Accepted: 25 February 2025;
Published online: 06 March 2025

References

1. Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764 (2013).
2. Scott, R. J. & Spielman, M. Genomic imprinting in plants and mammals: how life history constrains convergence. *Cytogenet Genome Res.* **113**, 53–67 (2006).
3. Wen, W., Alseekh, S. & Fernie, A. R. Conservation and diversification of flavonoid metabolism in the plant kingdom. *Curr. Opin. Plant Biol.* **55**, 100–108 (2020).
4. Chemes, L. B., de Prat-Gay, G. & Sánchez, I. E. Convergent evolution and mimicry of protein linear motifs in host–pathogen interactions. *Curr. Opin. Struct. Biol.* **32**, 91–101 (2015).
5. Edwards, M. et al. Association of the OCA2 polymorphism His615Arg with melanin content in East Asian populations: further evidence of convergent evolution of skin pigmentation. *PLoS Genet*. **6**. https://doi.org/10.1371/journal.pgen.1000867 (2010).
6. Kwiatkowski, D. P. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet*. **77**, 171–192 (2005).
7. Balentine, C. M. & Bolnick, D. A. Parallel evolution in human populations: a biocultural perspective. *Evolut. Anthropol.: Issues, N., Rev.* **31**, 302–316 (2022).
8. Pritchard, J. K. & Di Rienzo, A. Adaptation—not by sweeps alone. *Nat. Rev. Genet.* **11**, 665–667 (2010).
9. Gouy, A., Daub, J. T. & Excoffier, L. Detecting gene subnetworks under selection in biological pathways. *Nucleic Acids Res*. **45**. https://doi.org/10.1093/nar/gkx626 (2017).
10. Lawrence E. S., et al. Humangenetics functional EPAS1/HIF2A missense variant is associated with hematocrit in Andean highlanders. **10**. https://www.science.org (2024).
11. Yu, J. J. et al. Time domains of hypoxia responses and—omics insights. *Front. Physiol*. **13**, 885295 (2022).
12. Zhang, X. et al. The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans. *Proc. Natl. Acad. Sci. USA.* **118**, 1–9 (2021).
13. Bigham, A. et al. Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet.* **6**, e1001116 (2010).
14. Hernandez, R. D. et al. Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924 (2011).
15. Lindo, J. et al. The genetic prehistory of the Andean highlands 7000 years BP though European contact. *Sci. Adv*. **4**, eaau4921 (2018).
16. Zhang, P. et al. Denisovans and Homo sapiens on the Tibetan Plateau: dispersals and adaptations. *Trends Ecol. Evol*. **37**, 257–267 (2022).
17. Beall, C. M. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc. Natl. Acad. Sci. USA*. 104(Suppl 1), 8655–8660 (2007).
18. Julian, C. G. & Moore, L. G. Human genetic adaptation to high altitude: evidence from the Andes. *Genes*. **10**, 150 (2019).
19. Bigham, A. W. Genetics of human origin and evolution: high-altitude adaptations. *Curr. Opin. Genet Dev*. **41**, 8–13 (2016).
20. Crawford, J. E. et al. Natural selection on genes related to cardiovascular health in high-altitude adapted Andeans. *Am. J. Hum. Genet.* **101**, 752–767 (2017).
21. Jacovas, V. C. et al. Selection scan reveals three new loci related to high altitude adaptation in Native Andeans. *Sci. Rep*. **8**. https://doi.org/10.1038/s41598-018-31100-6 (2018).
22. Reich, D. et al. Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
23. Gnecchi-Ruscone, G. A. et al. Dissecting the pre-Columbian genomic ancestry of Native Americans along the Andes–Amazonia divide. *Mol. Biol. Evol.* **36**, 1254–1269 (2019).
24. Kanehisa, M. & Goto S. KEGG: kyoto encyclopedia of genes and genomes. **28**. http://www.genome.ad.jp/kegg/ (2000).
25. Jufri, N. F., Mohamedali, A., Avolio, A. & Baker, M. S. Mechanical stretch: physiological and pathological implications for human vascular endothelial cells. *Vasc Cell*. **7**. https://doi.org/10.1186/s13221-015-0033-z (2015).
26. Cao, D. et al. Retinoic acid-related orphan receptor C regulates proliferation, glycolysis, and chemoresistance via the PD-L1/ITGB6/STAT3 signaling axis in bladder cancer. *Cancer Res.* **79**, 2604–2618 (2019).
27. Daniunaite, K. et al. Promoter methylation of prkcb, adamts12, and naalad2 is specific to prostate cancer and predicts biochemical disease recurrence. *Int. J. Mol. Sci*. **22**. https://doi.org/10.3390/ijms22116091 (2021).
28. You, B. et al. Extracellular vesicles rich in HAX1 promote angiogenesis by modulating ITGB6 translation. *J. Extracell. Vesicles*. **11**. https://doi.org/10.1002/jev2.12221 (2022).
29. Sayers, E. W. et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **50**, D20–D26 (2022).
30. Wang, S. et al. The angiogenic genes predict prognosis and immune characteristics in esophageal squamous cell carcinoma: evidence from multi-omics and experimental verification. *Front. Oncol*. **12**, 961634 (2022).
31. Chen, Z. et al. Microfibril-associated glycoprotein-2 promoted fracture healing via integrin αvβ3/PTK2/AKT Signaling. *Lab. Investig*. **103**. https://doi.org/10.1016/j.labinv.2023.100121 (2023).
32. Shen, T. L. et al. Conditional knockout of focal adhesion kinase in endothelial cells reveals its role in angiogenesis and vascular development in late embryogenesis. *J. Cell Biol.* **169**, 941–952 (2005).
33. Adams, R. H. & Alitalo, K. Molecular regulation of angiogenesis and lymphangiogenesis. *Nat. Rev. Mol. Cell Biol.* **8**, 464–478 (2007).
34. Khankin, E. V., Royle, C. & Karumanchi, S. A. Placental vasculature in health and disease. *Semin. Thromb. Hemost.* **36**, 309–320 (2010).
35. Russell, M. W. et al. Damaging variants in proangiogenic genes impair growth in fetuses with cardiac defects. *J. Pediatr.* **213**, 103–109 (2019).
36. Harris, A. M. & DeGiorgio, M. A likelihood approach for uncovering selective sweep signatures from haplotype data. *Mol. Biol. Evol.* **37**, 3023–3046 (2020).
37. Kalluri, R. Basement membranes: structure, assembly and role in tumour angiogenesis. *Nat. Rev. Cancer* **3**, 422–433 (2003).
38. Alavi, M. V. et al. Col4a1 mutations cause progressive retinal neovascular defects and retinopathy. *Sci Rep*. **6**, 18602 (2016).
39. Kuo, D. S., Labelle-Dumai,s C. & Gould, D. B. Col4a1 and col4a2 mutations and disease: Insights into pathogenic mechanisms and potential therapeutic targets. *Hum. Mol. Genet*. **21**. https://doi.org/10.1093/hmg/dds346 (2012).
40. Xu, G. et al. The evolution of acquired resistance to BRAFV600E kinase inhibitor is sustained by IGF1-driven tumor vascular remodeling. *J. Investig. Dermatol.* **142**, 445–458 (2022).
41. Baker, J., Liu, J. P., Robertson, E. J. & Efstratiadis, A. Role of insulin-like growth factors in embryonic and postnatal growth. *Cell* **75**, 73–82 (1993).

42. Liu, J. P., Baker, J., Perkins, A. S., Roberteon, E. J. & Efetratiadie A. Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf.1) and type 1 IGF receptor (Igflr). *Cell* **75**, 59–72 (1993).
43. Liu, J. L., Yakar, S. & LeRoith, D. Mice deficient in liver production of insulin-like growth factor I display sexual dimorphism in growth hormone-stimulated postnatal growth. *Endocrinology* **141**, 4436–4441 (2000).
44. Woods, K. A., Camacho-Hübner, C., Savage, M. O. & Clark, A. J. Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor 1 gene. *N. Engl. J. Med*. **335**, 1363–1367 (1996).
45. Li, H., Fredriksson, L., Li, X. & Eriksson, U. PDGF-D is a potent transforming and angiogenic growth factor. *Oncogene* **22**, 1501–1510 (2003).
46. Sitras, V. et al. Differential placental gene expression in severe preeclampsia. *Placenta* **30**, 424–433 (2009).
47. Demydenko, K., Ekhteraei-Tousi, S. & Roderick, H. L. Inositol 1,4,5-trisphosphate receptors in cardiomyocyte physiology and disease. *Philos. Trans. R. Soc. B.: Biol. Sci*. **377**. https://doi.org/10.1098/rstb.2021.0319 (2022).
48. Lehnart, S. E., Wehrens, X. H. T., Kushnir, A. & Marks, A. R. Cardiac ryanodine receptor function and regulation in heart disease. *Ann. N. Y. Acad. Sci.* **1015**, 144–159 (2004).
49. Daniels, R. E. et al. Cardiac expression of ryanodine receptor subtype 3; a strategic component in the intracellular Ca2+ release system of Purkinje fibers in large mammalian heart. *J. Mol. Cell Cardiol.* **104**, 31–42 (2017).
50. Busch, S. A. et al. Global REACH: assessment of brady-arrhythmias in Andeans and Lowlanders during apnea at 4330 m. *Front. Physiol*. **10**, 1603 (2020).
51. Bjertness, E., Wu, T., Stigum, H. & Nafstad, P. Acute mountain sickness, arterial oxygen saturation and heart rate among Tibetan students who reascend to Lhasa after 7 years at low altitude: a prospective cohort study. *BMJ Open.* **7**, e016460 (2017).
52. Gnecchi-Ruscone, G. A. et al. Evidence of polygenic adaptation to high altitude from Tibetan and Sherpa genomes. *Genome Biol. Evol.* **10**, 2919–2930 (2018).
53. Smith, S. C. et al. Profiling bladder cancer organ site-specific metastasis identifies LAMC2 as a novel biomarker of hematogenous dissemination. *Am. J. Pathol.* **174**, 371–379 (2009).
54. Pei, Y. F., Liu, J., Cheng, J., Wu, W. D. & Liu, X. Q. Silencing of LAMC2 reverses epithelial-mesenchymal transition and inhibits angiogenesis in cholangiocarcinoma via inactivation of the epidermal growth factor receptor signaling pathway. *Am. J. Pathol.* **189**, 1637–1653 (2019).
55. Garg, M. et al. Laminin-5γ-2 (LAMC2) Is highly expressed in anaplastic thyroid carcinoma and is associated with tumor progression, migration, and invasion by modulating signaling of EGFR. *J. Clin. Endocrinol. Metab.* **99**, E62–E72 (2014).
56. Wang, R. et al. Overexpressed LAMC2 promotes trophoblast over-invasion through the PI3K/Akt/MMP2/9 pathway in placenta accreta spectrum. *J. Obstet. Gynaecol. Res.* **49**, 548–559 (2023).
57. Cahill, A. G., Beigi, R., Heine, R. P., Silver, R. M. & Wax, J. R. Placenta accreta spectrum. *Am. J. Obstet. Gynecol.* **219**, B2–B16 (2018).
58. Bartels, H. C., Postle, J. D., Downey, P. & Brennan D. J. Placenta accreta spectrum: a review of pathology, molecular biology, and biomarkers. *Dis. Markers*. **2018**, 1507674 (2018).
59. Goh, W., Yamamoto, S. Y., Thompson, K. S. & Bryant-Greenwood, G. D. Relaxin, its receptor (RXFP1), and insulin-like peptide 4 expression through gestation and in placenta accreta. *Reprod. Sci.* **20**, 968–980 (2013).
60. Hu, H. et al. Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLoS Genet.* **13**, 1–22 (2017).
61. Arciero, E. et al. Demographic history and genetic adaptation in the Himalayan region inferred from genome-wide SNP genotypes of 49 populations. *Mol. Biol. Evol.* **35**, 1916–1933 (2018).
62. Deng, L. et al. Prioritizing natural-selection signals from the deep-sequencing genomic data suggests multi-variant adaptation in Tibetan highlanders. *Natl. Sci. Rev.* **6**, 1201–1222 (2019).
63. Chen, Y. et al. Epigenetically upregulated oncoprotein PLCE1 drives esophageal carcinoma angiogenesis and proliferation via activating the PI-PLCε-NF-κB signaling pathway and VEGF-C/ Bcl-2 expression. *Mol. Cancer***18**. https://doi.org/10.1186/s12943-018-0930-x (2019).
64. Scruggs, S. B., Wang, D. & Ping, P. PRKCE gene encoding protein kinase C-epsilon—dual roles at sarcomeres and mitochondria in cardiomyocytes. *Gene* **590**, 90–96 (2016).
65. Ferraretti, G. et al. Archaic introgression contributed to shape the adaptive modulation of angiogenesis and cardiovascular traits in human high-altitude populations from the Himalayas. *Elife*. **12**, RP89815 (2024).
66. Tong, J., Niu, Y., Chen, Z. J. & Zhang, C. Comparison of the transcriptional profile in the decidua of early-onset and late-onset pre-eclampsia. *J. Obstet. Gynaecol. Res.* **46**, 1055–1066 (2020).
67. Oefner, C. M. et al. Collagen type IV at the fetal-maternal interface. *Placenta* **36**, 59–68 (2015).
68. Okada, H., Tsuzuki, T. & Murata, H. Decidualization of the human endometrium. *Reprod. Med. Biol.* **17**, 220–227 (2018).
69. Plaisier, M. Decidualisation and angiogenesis. *Best Pract. Res. Clin. Obstet. Gynaecol.* **25**, 259–271 (2011).
70. Manier, G., Guenard, H., Castaing, Y., Varene, N. & Vargas, E. Pulmonary gas exchange in Andean natives with excessive polycythemia-effect of hemodilution. *J. Appl. Physiol.* **65**, 2107–2117 (1988).
71. Beall, C. M. et al. Hemoglobin concentration of high-altitude Tibetans and Bolivian Aymara. *Am. J. Phys. Anthropol.* **106**, 385–400 (1998).
72. Villafuerte, F. C., Cárdenas, R. & Monge-C, C. Optimal hemoglobin concentration and high altitude: a theoretical approach for Andean men at rest. *J. Appl Physiol.* **96**, 1581–1588 (2004).
73. Wu, D. et al. How placenta promotes the successful reproduction in high-altitude populations: a transcriptome comparison between adaptation and acclimatization. *Mol. Biol. Evol*. **39**. https://doi.org/10.1093/molbev/msac120 (2022).
74. Sharma, V., Varshney, R. & Sethy, N. K. Human adaptation to high altitude: a review of convergence between genomic and proteomic signatures. *Hum. Genom*. **16**. https://doi.org/10.1186/s40246-022-00395-y (2022).
75. O'Brien, K. A. et al. Genomic selection signals in Andean highlanders reveal adaptive placental metabolic phenotypes that are disrupted in preeclampsia. *Hypertension*https://doi.org/10.1161/HYPERTENSIONAHA.123.21748 (2023).
76. Chang, C. C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*. **4**. https://doi.org/10.1186/s13742-015-0047-8 (2015).
77. Ojeda-Granados, C. et al. Dietary, cultural, and pathogens-related selective pressures shaped differential adaptive evolution among native Mexican populations. *Mol. Biol Evol*. **39**. https://doi.org/10.1093/molbev/msab290 (2022).
78. Loh, P. R., Palamara, P. F. & Price, A. L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat. Genet.* **48**, 811–816 (2016).
79. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
80. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
81. Palmer, C. & Pe'er I. Bias characterization in probabilistic genotype data and improved signal detection with multiple imputation. *PLoS Genet*. **12**. https://doi.org/10.1371/journal.pgen.1006091 (2016).
82. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, 2074–2093 (2006).

83. Mantel, N. *Cancer Research*. **27**. http://aacrjournals.org/cancerres/article-pdf/27/2_Part_1/209/2382183/cr0272p10209.pdf (1967).
84. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
85. Lawson, D. J., Hellenthal, G., Myers, S. & Falush D. Inference of population structure using dense haplotype data. *PLoS Genet*. **8**. https://doi.org/10.1371/journal.pgen.1002453 (2012).
86. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
87. Garud, N. R., Messer, P. W., Buzbas, E. O. & Petrov, D. A. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. *PLoS Genet.* **11**, 1–32 (2015).
88. Ferrer-Admetlla, A., Liang, M., Korneliussen, T. & Nielsen, R. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol. Biol. Evol.* **31**, 1275–1291 (2014).
89. Szpiech, Z. A. & Hernandez, R. D. Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824–2827 (2014).
90. Gilbert-Kawai, E. T., Milledge, J. S., Grocott, M. P. W. & Martin, D. S. King of the mountains: Tibetan and sherpa physiological adaptations for life at high altitude. *Physiology* **29**, 388–402 (2014).

Acknowledgements

We acknowledge support from the Fondazione Cassa di Risparmio in Bologna through the project 'Genetic adaptation and acclimatisation to high altitude as experimental models to investigate the biological mechanisms that regulate physiological responses to hypoxia', which was granted to M.S. (n. 2019.0552). We also would like to thank John Lindo for having kindly shared Aymara WGS data without which this work would have not been possible.

Author contributions

Data curation, formal analysis, investigation, writing original draft, G.F., A.I., S.S.; software, formal analysis, P.A., K.S., M.A., C.O.-G.; data curation, review and editing original draft, S.D.F., M.I., A.D., P.T.S., P.C., M.T., M.D.M., G.A.G.-R., L.N., A.C., G.M., D.P.; resources, P.G., D.L., D.P.; conceptualisation, resources, supervision, funding acquisition, methodology, writing, review and editing original draft, M.S.

Competing interests

The authors declare no competing interests.

Ethics

A written informed consent for data treatment was signed by each participant as attested in previous works[15,22,23] and all ethical regulations relevant to human research participants were followed. Moreover, on 09/12/2019 the Ethics Committee of the University of Bologna released approval (prot. 205142) for the present study within the framework of the project 'Genetic adaptation and acclimatisation to high altitude as experimental models to investigate the biological mechanisms that regulate physiological responses to hypoxia' funded by Fondazione Cassa di Risparmio in Bologna (grant n. 2019.0552).

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-07813-6.

Correspondence and requests for materials should be addressed to Marco Sazzini.

Peer review information *Communications Biology* thanks Wanjun Gu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Ophelia Bu. A peer review file is available.

Complex adaptive evolution at thyroid hormone, insulin and glycerolipid pathways improved energy metabolism in Eurasian populations from high-latitude cold environments

Giulia Ferraretti[1], Marta Alberti[1], Rosita Cicolini[1], Sabrina Pognant Viù[1], Stefania Sarno[1], Marco Sazzini[1,2].

[1] *Laboratory of Molecular Anthropology & Centre for Genome Biology, Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna Italy.*
[2] *Interdepartmental Centre Alma Mater Research Institute on Global Challenges and Climate Change, University of Bologna, Bologna, Italy.*

Eurasian populations from high-latitudes represent valuable case-studies to investigate the genetic bases of some of the most effective biological adaptations evolved by modern humans to cope with cold climates. Nevertheless, by relying on traditional models describing the action of natural selection (i.e., selective sweeps), studies conducted so far succeeded in identifying a limited fraction of these adaptive events. To overcome this issue, we brought together diverse inferential methods to pinpoint combinations of genes presenting both signatures of adaptive evolution and functional relationships supporting their synergic role in modulating a biological trait, as expected under a mechanism of polygenic adaptation. This approach was applied on whole genome sequence data available for the Yakut ethnic group from Northeastern Siberia and results were compared with those obtained for a Russian population showing relatively shared ancestry with them to distinguish between Yakut-specific adaptive events and adaptations plausibly shared among other high-latitude populations. Multiple genes contributing to biological functions known to be modulated during cold exposure, such as thyroid hormone/insulin signalling, brown adipose tissue differentiation, and glycerolipid metabolism, showed patterns of variation conform to adaptive evolution. Changes at these loci support enhanced heat production in brown adipocytes and responsiveness to insulin in other peripheral tissues, with regulation of glucose uptake and lipid metabolism being found to have been influenced also by introgression of Neanderthal/Denisovan alleles. These findings provided new insights concerning the genetic bases of complex adaptive traits evolved by the ancestors of high-latitude Eurasian populations in response to selective pressures imposed by cold environments.

Introduction

The Yakutia region in Northeastern Siberia represents one of the most remote and extreme environments ever inhabited by modern human populations, being especially characterized by long, harsh winter seasons with temperatures dropping at –71°C (Chevychelov and Bosikov 2010). Archaeological evidence indicates that *Homo sapiens* resided in this area long before the Last Glacial Maximum (LGM) in the late Pleistocene, as demonstrated by fossil remains of a hunted mammoth in the Yenisei Bay area dated to approximately 45,000 years ago (Pitulko et al. 2016). Nowadays, the Yakut (Sakha) people, a Turkic-speaking indigenous human group, constitute the majority of the local population (Puzyrev et al. 2003; Levy et al. 2013). Given the extreme climatic conditions they have to cope with, high-latitude Siberian populations, such as the Yakuts, provide a valuable case study for investigating the genetic bases of some of the most effective complex biological adaptations evolved by modern humans in response to cold-induced selective pressures. Nevertheless, previous studies that addressed this research question relied mainly on the analysis of genome-wide data, being thus limited to the identification of adaptive events ascribable to traditional models

describing the action of natural selection, such as hard and soft selective sweeps (Cardona et al. 2014; Hallmark et al. 2019). In fact, to our knowledge no evidence has been provided so far concerning the genetic determinants of complex (i.e., polygenic) adaptations to extreme cold environments evolved by high-latitude Siberian human groups. This is plausibly due to both the scarcity of methods capable to formally test a realistic approximation of the polygenic adaptation model and the only recent availability of whole genome sequence (WGS) data for these populations (Bergström et al. 2020), which represent the sole kind of data suitable to investigate the tangled gene-gene interactions underlying the modulation of complex adaptive traits. Notably, polygenic adaptations are increasingly supposed to represent the vast majority of the adaptive responses evolved by modern human populations (Pritchard and Di Rienzo 2010), especially by those having maintained low effective population size trough time due prolonged isolation (Hernandez et al. 2011; Gnecchi-Ruscone et al. 2018; Ferraretti et al. 2024; Ferraretti et al. 2025), which is a condition previously proposed for the Yakuts (Kharkov et al. 2008; Bergström et al. 2020).

Another significant gap in the current literature on this topic concerns the role of alleles/haplotypes introduced in the *H. sapiens* gene pool via archaic introgression events in shaping variation at genomic regions plausibly involved in the regulation of adaptive responses to cold. Interestingly, the study by Bergström et al. (2020) primary pointed out traces of archaic introgression from both Neanderthal and Denisovan species in Yakut genomes, even though without testing for adaptive introgression events, with the percentage of ancestry attributable to Denisovans being one of the most elevated among those estimated for a large panel of Eurasian populations (Bergström et al. 2020).

Specifically, to investigate the genomic architecture of polygenic adaptations evolved by Yakut ancestors, we first combined two methodologies capable of recognizing different patterns of haplotype variation informative of a variety of selective events and of detecting biological functions that have been pervasively shaped by action of natural selection (Gouy et al. 2017; Harris and DeGiorgio 2020). We also validated the obtained findings by implementing a supervised machine-learning method capable of distinguishing genomic regions characterized by adaptive evolution from neutral background by testing more sophisticated evolutionary models (Mughal and DeGiorgio 2019). We applied such an approach on the sole available WGS data for Yakut people (Bergström et al. 2020), who are considered here as a proxy of high-latitude Siberian human populations. We then compared the results with those obtained for a control population showing a relatively shared genetic ancestry with Yakuts, but living in less extreme environmental settings and having experienced considerably less isolation. This allowed us to identify genes/functional pathways putatively involved in the evolution of cold adaptations that are either unique to the Yakuts or shared across diverse populations and thus presumably representative of more ancient adaptive events occurred in the ancestral groups that first colonized the Eurasian continent up to these high-latitude regions. We finally assessed the impact of archaic introgression in shaping variation at the identified candidate adaptive loci by applying different methods aimed at detecting introgressed chromosomal segments and by comparing haplotype structure at these regions between modern and archaic genomes (Marnetto and Huerta-Sánchez 2017; Browning et al. 2018). Overall, we succeeded in identifying some of the genetic determinants of

polygenic adaptations that have been likely evolved by human populations living at high latitudes in response to the selective pressures imposed by extreme cold environments.

Results

Genetic relationships between Yakut and populations of Central/East Asian and European ancestry

After merging WGS data collected for the Yakut population and a large panel of reference human groups (see Materials and Methods), we obtained an extended unphased dataset including 593 individuals belonging to 24 populations and characterized for 5,056,113 Single Nucleotide Variants (SNVs) (supplementary table S1, Supplementary Material online). Such a dataset was subjected to stringent quality check (QC) filtering procedures (see Materials and Methods), which encompassed pruning loci in linkage disequilibrium (LD) and retained 345,021 SNVs to be used for performing genotype-based populations structure analyses. Principal Components Analysis (PCA) and multiple ADMIXTURE runs thus enabled to first evaluate the overall homogeneity and representativeness of the considered Yakut samples, as well as their possible genetic relations with populations of Central/East Asian, European and Native American ancestry. In details, PC1 was found to describe the main differences observable among European, Southeast Asian, Northeast Asian and Central Asian populations, with Yakut individuals distributing between these two latter groups and showing appreciable genetic affinity primarily with Northeast Asians, as supported also by PC3 (supplementary fig. S1, Supplementary Material online). According to ADMIXTURE analysis, the configuration that best fitted to the data (K=4) showed a nearly fixed ancestry component (i.e., reaching an average frequency of 0.85) in Yakut individuals (Fig. 1a, supplementary figs. S2 and S3, Supplementary Material online), which was detectable, although with smaller proportions, also in Russian, Central Asian and, especially, Northeast Asian groups (Fig. 1a, supplementary fig. S3, Supplementary Material online). The second most represented ancestry faction in Yakut genomes (i.e., average frequency of 0.123) was found to be the predominant one in European populations, while negligible proportions of typical Native American (i.e., 0.025) and Southeast Asian (i.e., 0.00001) components were observed. Overall, these findings suggest a remarkable degree of genetic differentiation of Yakuts when compared with populations of diverse ancestries, which is potentially ascribable to their prolonged isolation with respect to other Turkic-speaking populations of Southern Siberia, as previously proposed according to analyses conducted on uniparentally inherited genetic loci (Kharkov et al. 2008).

To investigate patterns of fine-scale population structure and to test whether the considered Yakut individuals represent a group sufficiently representative and homogeneous from a genetic perspective to be used to infer the adaptive history of the overall population, we carried out ChromoPainter and fineSTRUCTURE analyses on the phased extended dataset (see Materials and Methods). Differently from PCA and ADMIXTURE analyses, such an approach suggested a more consistent shared genetic ancestry among Yakut and some Central Asian (e.g., Xibo and Uygurs), Southeast Asian (e.g., Yi, Naxi, and Miao) and Russian populations (Fig. 1b, supplementary figs. S4 and S5, Supplementary Material online). In particular, according to the fineSTRUCTURE dendrogram Yakut individuals were found to cluster along

with all Yi and Xibo subjects, as well as with most Uygur samples (supplementary fig. S5, Supplementary Material online). Moreover, this homogeneous clade originated from the same branch of the dendrogram leading to the Naxi, Miao, and Russian clusters (Fig. 1b).

Collectively, results from both genotype-based and haplotype-based population structure analyses revealed a remarkable genetic homogeneity of the sequenced Yakut subjects and allowed us also to pinpoint the Russian cohort of individuals as a suitable control group (in terms of both relatively related ancestry and comparable sample size) to be used for distinguishing between Yakut-specific signatures of adaptive evolution and those instead plausibly shared among other high-latitude Eurasian populations.

Effective population size history, gene flow and divergence time between Yakut and control populations

To further assess how much and how long the ancestors of Yakut and Russian populations shared a common evolutionary history, we used the SMC++ approach to infer their effective population size (N_e) fluctuations and genetic split time (see Materials and Methods) (Terhorst et al. 2017). As regards the Yakuts, a decline in N_e beginning approximately 200,000 years ago was observed, being followed by a population expansion since around 24,000 years ago, which reached a peak 8,000 years before the present (Fig. 1c). The effective population size history of the Russian control group exhibited a similar trajectory, but being characterized by an earlier demographic recovery (starting approximately 35,000 years ago) and slightly higher N_e values, potentially reflecting a pattern of less remarkable isolation with respect to what experienced by Yakut people. These demographic trends are consistent with previous findings (Bergström et al. 2020). When divergence time between the ancestors of these groups was estimated, a value ranging between 7,500 and 8,000 years ago was obtained. However, it is worth noting that the SMC++ algorithm models genetic differentiation between groups as a function of time according to an idealized two populations split scenario with no post-divergence gene flow (Terhorst et al. 2017). Since results of ADMIXTURE analysis (Fig. 1a) and literature data (Pakendorf et al. 2003) suggest that a certain degree of gene flow has been occurred between Central/East Asian populations (including Yakuts) and the Russian control group, we used f3-statistics computation and ALDER approaches to formally investigate potential evidence of admixture between them. The obtained results confirmed the introduction of Central/East Asian-related ancestry in the Russian gen pool dating approximately between 52 and 35 generations ago, which roughly corresponds to a time interval spanning from 1,500 to 1,000 years ago assuming a generation time of 30 years (supplementary table S2, Supplementary Material online). In details, significant f3-test indeed showed that the Russian population can be modelled as a mixture between a European-like source (i.e., CEPH) and different populations from Central and East Asia. Interestingly, the most significant z-score (Z = -27.4) was obtained for trios involving Yakut as one of the two parental groups, for instance implementing the test in the form of f3(Russian; Yakut, CEPH), with a corresponding ALDER admixture date estimate of 47.18+/-4.81 generations ago (p-value = $1x10^{-22}$). That being so, evidence supporting relatively recent admixture between Yakut and control groups pointed to a violation of the SMC++ assumption of no post-divergence gene flow, suggesting that the inferred genetic split time between them is likely underestimated.

Polygenic adaptive events modulated Yakut thyroid functioning, lipid metabolism and insulin signalling

To investigate the genetic bases of polygenic adaptations evolved by the Yakut population, we exploited information for the 6,691,957 SNVs included in the Yakut phased WGS dataset to implement the Likelihood Approach for Selective Sweep identification (*LASSI*)-*signet* pipeline of analyses described in the Materials and Methods section. Furthermore, to pinpoint adaptive events peculiar to the Yakuts, we replicated the same analyses also on the Russian control group, which was characterized for 7,482,199 SNVs, and we searched for private and shared patterns of adaptive evolution.

First and foremost, by considering genomic windows ranking in the top 1% of the overall distribution of the *LASSI* statistic, this approach enabled us to recapitulate traditional selective sweeps occurred at the *THADA* gene (supplementary tables S3 and S4, Supplementary Material online), which were previously assessed for the Yakut population (Cardona et al. 2014; Hallmark et al. 2019) and have been proposed to be involved in adaptive responses to extreme cold environments due to the role played by this locus in the regulation of energy metabolism and in the modulation of Type 2 diabetes susceptibility observed in diverse human populations (Cardona et al. 2014).

When we instead used results obtained by *LASSI* to guide the application of the *signet* algorithm (see Materials and Methods), we were able to detect genomic signatures ascribable to the action of positive selection according to a reliable approximation of a polygenic adaptation model. Among those likely related to adaptive events evolved to cope with extreme cold environments, a significant gene network including the *RCAN2*, *HIF1A*, *ATP1B2*, *PIK3R2*, *RXRA*, *SRC*, and *THRB* loci participating to the *Thyroid hormone signalling* pathway was observed exclusively in the Yakut population (Fig. 2, Supplementary Table S5). These results may indicate adaptations evolved specifically in such human group and modulating thyroid functioning, in accordance with previous evidence advancing the crucial role of thyroid hormones in regulating physiological responses to cold stress (Tsibulnikov et al. 2000). Similarly, we identified a significant Yakut-specific gene network belonging to the *AGE-RAGE signalling pathway in diabetic complications*, which included genes (e.g., *NFKB1* and *PRKCD*) whose functions were previously found to be related to regulation of glucose homeostasis and the development of insulin resistance (Bezy et al. 2011; Chen et al. 2013; Guo et al. 2021). In line with this, also the *HTR1F* gene, which was previously proposed to be related to similar functions (Almaça et al. 2016), was found to be included in the *Serotonergic synapse* significant pathway identified for the Yakuts. Moreover, a gene network observed in both the Yakut and control population and belonging to the *PI3K-Akt signaling* pathway included some loci (e.g., *INSR* and *EFNA5*) involved in the modulation of insulin secretion, binding, and activation of insulin downstream responses (Konstantinova et al. 2007; Chen et al. 2019) (Fig. 2, supplementary tables S5 and S6, Supplementary Material online). Two additional significant gene networks belonging to the *Glycerolipid metabolism* and *cGMP-PKG signaling* pathways were also detected in both the examined populations, suggesting that natural selection anciently targeted these complex biological traits, as well as the genes involved in their modulation, long before the divergence of these human groups from a common ancestral gene pool (Fig. 2). Finally, after an extensive literature review, we assessed that some genes (e.g., *NTRK1*,

GPAT3, *CAMK2D*, and *CLDN10*) identified within multiple significant Yakut-specific networks were previously proved to exhibit altered expression in mice Brown Adipose Tissue (BAT), white adipose tissue, and liver after cold exposure (Shore et al. 2013; Wang and Wahl 2014; Labbé et al. 2015; Ghosh et al. 2021) (supplementary table S7, Supplementary Material online). Among them, six genes (e.g., *ADCY3*, *GPAT3*, *INSR*, *PDE4D*, *PDGFRA*, and *PLCB1*) were found to be included also in significant networks observed for the Russian control population (supplementary table S7, Supplementary Material online).

Archaic introgression similarly shaped adaptive traits in Yakut and control populations

To test if variation patterns at the identified putative adaptive genes have been potentially shaped also by gene flow from archaic human species, we implemented multiple approaches aimed at detecting genomic segments showing alleles introgressed from different archaic sources (see Materials and Methods). To this end, we first applied the *Sprime* method (Browning et al. 2018) on the phased Yakut WGS dataset and we crosschecked the obtained results with those ranking within the top 1% of the distribution of the *LASSI* statistic. Submitting chromosomal intervals putatively impacted by both archaic introgression and adaptive evolution to STRING network analysis first revealed a significant enrichment of genes known to be involved in the regulation of immune functions (enriched Monarch phenotype *respiratory disease biomarker*, False Discovery Rate, FDR = 1.88e-21) (supplementary fig. S6, Supplementary Material online). When we instead crosschecked *Sprime* results with those informative of polygenic adaptive events according to the combined *LASSI-signet* approach (see Materials and Methods), we shortlisted several genes participating to significant networks such as those belonging to the *Glycerolipid metabolism* and *Thyroid hormone signalling* (supplementary tables S5, Supplementary Material online). Interestingly, the expression of some of these genes (i.e., *ADCY3*, *CAMK2D*, *PDE4D*, *KCNN2*, *EGFR*, and *IMPA2*) was previously proposed to be altered in mouse BAT after cold exposure (Shore et al. 2013) (supplementary table S7, Supplementary Material online).

To further validate these results and to explicitly test for the archaic source responsible for the observed introgression signatures, we run the Haplostrips algorithm (Marnetto and Huerta-Sánchez 2017) to compare putative adaptive introgressed haplotypes at Yakut and control groups with those of three different archaic genomes (i.e., Neanderthal Altai, Neanderthal Vindija, and Altai Denisovan) (see Materials and Methods for links at specific repositories), as well as of the Yoruba outgroup population (see Materials and Methods for further details). Overall, such an approach corroborated the occurrence of adaptive introgression events at the *ADCY3*, *CAMK2D*, *PLPP3*, *ATP1B2* and *CD247* loci (Fig. 3a, supplementary figs. S7, S8, S9, and S10, Supplementary Material online). The most compelling results were obtained for the *ADCY3* gene, which turned out to be extremely conserved among Yakut and control populations and was previously proposed to be involved in the modulation of polygenic adaptive events able to improve thermogenesis (Sazzini et al. 2020). In detail, most of the Yakut and control *ADCY3* haplotypes were found to present the lowest number of pair-wise differences with respect to the Denisovan genome when compared with the outgroup population of African ancestry (Fig. 3a). An even more clear picture of tight genetic affinity between Yakut/control and

Denisovan haplotypes was depicted by limiting such a comparison to the *ADCY3* chromosomal intervals pointed out by *LASSI* analysis as targeted by positive selection (Fig. 3b). Similar patterns were observed also for the *PLPP3*, *ATP1B2* and *CD247* genes (supplementary figs. S7, S8, S9, Supplementary Material online). Conversely, Haplostrips analysis failed to confirm signatures of archaic introgression at the *CAMK2D* putative adaptive haplotypes (supplementary figs. S10 and S11, Supplementary Material online), suggesting that introgressed alleles at this locus were not involved in the modulation of adaptive processes.

Putative Yakut adaptive loci are characterized by shifts in haplotype/genotype frequency with respect to the control population

By focusing on putative adaptive events specifically occurred in the Yakut population, we then explored allele composition and frequencies of adaptive haplotypes at genes included in significant networks from *Thyroid hormone signalling* and *AGE-RAGE signalling pathway in diabetic complications*, as well as at adaptive loci belonging to different pathways but known to play a role in these biological functions (see Materials and Methods). For the *THRB*, *SRC*, *RCAN2*, *INSR*, *NFKB1*, *PRKCB*, and *HTR1F* genes, we observed significant differences in the frequency of multiple SNVs (i.e., adjusted FDR ≤ 0.05) between the Yakut and control populations (supplementary table S8, Supplementary Material online). Interestingly, some of these variants have been previously reported in the GTEx database as expression Quantitative Trait Loci (eQTLs) able to regulate the expression of the identified candidate adaptive genes in diverse tissues (e.g., white adipose tissue, skeletal muscle, thyroid and blood) according to studies aimed at examining the physiological responses activated by cold exposure (supplementary table S8, Supplementary Material online) (Ribeiro et al. 2001; Blondin et al. 2015; Xu et al. 2019; Zekri et al. 2021). Overall, the frequencies of putative adaptive haplotypes containing these eQTLs were found to be remarkably higher in the Yakut population (i.e., ranging from 0.79 to 0.92) with respect to the control group (supplementary table S9, Supplementary Material online). In particular, such a pattern was observed for all the tested genes except for *THRB*. However, when considering genotype instead of haplotype frequencies, homozygotes for all these candidate adaptive eQTLs were consistently more represented in Yakuts than in the control population (supplementary fig. S12, supplementary table S10, Supplementary Material online). In detail, candidate adaptive eQTLs at *INSR* and *THRB* have been previously associated, in the homozygous state, to increased expression of these genes in white adipose tissue, while those at *HTR1F* and *SRC* were found to increase expression in white adipose tissues/thyroid and liver, respectively, reducing instead the expression of *SRC* in pancreas (supplementary fig. S13, Supplementary Material online) (info available at https://gtexportal.org/home/). Finally, homozygotes for candidate adaptive eQTLs at *RCAN2* have been demonstrated to decrease expression of such a gene in skeletal muscle, as well as of *PLA2G7* in the blood (info available at https://gtexportal.org/home/) (supplementary fig. S13, Supplementary Material online).

Machine-learning approach further validates signatures of positive selection in Yakut and control genomes

To further validate at the single gene level signatures ascribable to the action of natural selection identified by analyses previously described, as well as to point out additional putative adaptive genes involved in the same/related biological functions, we applied the *Trendsetter* multinomial regression classifier to distinguish between different neutral and adaptive evolutionary scenarios (Mughal and DeGiorgio 2019) (see Materials and Methods). Before applying the algorithm to the Yakut and control WGS data, we trained it using *ad hoc* simulated genomic datasets. In detail, we generated two sets of simulations for each of the possible predicted classes of evolutionary events (i.e., hard/soft sweeps and neutral evolution) respectively to i) train the model and ii) build calibration curves, calculate accuracy, precision, recall and f1 scores for the trained algorithm (see Materials and Methods). Calibration/reliability plots suggested the achievement of a good calibration of the predicted output probabilities for both Yakut and control classifications (supplementary figs. S14 and S15, Supplementary Material online), which turned out to be characterized by high accuracy, with values equal to 0.87 and 0.89, respectively (Fig. 4a-b, supplementary fig. S15, Supplementary Material online). Precision, recall and f1-score values were also elevated for both models (i.e., ranging from 0.80 to 0.94), attesting the capacity of the trained classifiers in distinguishing neutral genomic regions from those that evolved adaptively (Fig. 4a-b, supplementary fig. S15).

According to such an approach, two genes (i.e., *SRC* and *RCAN2*) included in the significant network belonging to the *Thyroid hormone signalling* pathway pointed out by *LASSI-signet* analyses were found to be supported as putative adaptive loci in Yakut population (supplementary fig. S16 and supplementary table S11, Supplementary Material online). More in detail, extremely low probabilities of being classified as neutral regions when compared with those relative to the classes of hard and/or soft selective sweeps were obtained for several genomic windows included in these genes (supplementary table S11, Supplementary Material online). For instance, multiple *SRC* chromosomal intervals presented neutral output probabilities of ~ 0.01%, compared to ~ 0.84% and to ~ 0.13% of being classified as soft and hard selective sweeps, respectively (supplementary fig. S16 and supplementary table S11, Supplementary Material online). Comparable results were also obtained for multiple genes belonging to the *AGE-RAGE signalling pathway in diabetic complications* and to *Serotonergic synapse* pathway (i.e., *HTRF1*, *NFKB1*, *PRKCB*, *PRKCD*) (Fig. 4c-d, supplementary fig. S17 and supplementary table S11, Supplementary Material online). As regards instead putative adaptive genes belonging to *Glycerolipid metabolism*, *PI3K-Akt signalling pathways*, and *cGMP-PKG signalling*, *GPAT3*, *ERBB4*, *PRKCB* and *PRKG1 genes* were found to be supported in the Yakuts by the *Trendsetter* approach (supplementary figs. S16 and S17, supplementary table S11, Supplementary Material online), while *PNLIP*, *MOGAT2*, and *PLPP1* loci were proposed to have adaptively evolved in the control group (supplementary fig. S18, supplementary table S12, Supplementary Material online). Finally, signatures of positive natural selection identified for the *ADCY3* gene were supported as well by the *Trendsetter* method in the control group (supplementary fig. S18, supplementary table S12, Supplementary Material online).

Discussion

In the present study, we analysed WGS data available for Yakut people from Northeastern Siberia (Bergström et al. 2020) with the aim of shedding light on the genetic bases of polygenic adaptions evolved by human groups who have long inhabited extreme cold environments. For this purpose, we first investigated the composition of their genomes in terms of ancestry components, which revealed their appreciable differentiation with respect to other Turkic-speaking populations from Siberia, as well as to the bulk of the considered groups of Southeast Asian, Northeast Asian and Central Asian ancestry (Fig. 1 a). By exploring fine-scale genetic structure of the populations included in the assembled dataset, we then demonstrated that the studied Yakut individuals are genetically homogeneous and quite representative of their population of origin, also presenting a relatively common genetic ancestry with a Russian population that was thus selected as a control group to distinguish between shared and private adaptive events having shaped their biological responses to cold climate-induced selective pressures (Fig. 1b). Based on these preliminary results, we inferred Yakut *Ne* history and genetic split time with respect to the control population (Fig. 1b), also testing and dating possible admixture events having involved the ancestors of these human groups in addition to those of other Central/East Asian populations. Results of f3 and ALDER analyses supported the introduction of Central/East Asian-related ancestry in the Russian genetic background due to gene flow occurred approximately 1,000-1,500 years ago and involving the Yakut population (supplementary table S2, Supplementary Material online). Since the SMC++ approach used to infer populations split time assumes no post-divergence gene flow between them (Terhost et al. 2016), this relatively recent admixture suggests that genetic divergence between Yakut and Russian groups likely reached appreciable levels even before the estimated 8,000 years ago (Fig. 1b). In this regard, we can speculate that progressive differentiation of their gene pools started just after the end of the Late Glacial Maximum (LGM), when human groups that have survived to the glaciation in the tundra and cool steppe environments of Eastern Russia and Southern Siberia began to re-occupy the northernmost territories (Kuzmin 2008).

We then reconstructed and compared the adaptive history of the Yakut and control populations by implementing a combination of statistical methods capable of distinguishing a variety of selective events ascribable to different mechanisms of adaptation (Gouy et al. 2017; Harris and DeGiorgio 2020) and by using an independent supervised machine learning approach (Mughal and DeGiorgio 2019) as a validation step. Overall, the most robust results pointed to genes involved in modulation of thyroid functioning and regulation of adaptive thermogenesis as loci having adaptively evolved specifically in the Yakut population, while variation patterns at insulin resistance/signalling and lipid metabolism pathways appeared to have been similarly shaped by positive selection in both Yakut and control groups. According to the timeframe of progressive differentiation between these populations suggested by SMC++ analysis, such latter findings might represent ancient adaptive responses evolved by the ancestral populations that first colonized the Eurasian continent up to these high-latitude regions, since around 45,000 years ago (Pitulko et al. 2016), with Yakut-specific adaptations being instead plausibly evolved after the LGM. In detail, the combined *LASSI-signet* approach pinpointed multiple networks of genes belonging to several functional pathways as involved in the modulation of complex Yakut-specific adaptive responses to cold climate conditions (Fig. 2,

supplementary table S5, Supplementary Material online). Coupled with previous evidence these findings have led us to focus on *THRB*, *RXRA*, *RCAN2*, and *SRC* genes contributing to *Thyroid hormone signalling*; *PRKG1* to *cGMP-PKG signalling*; *ERBB4* to in *PI3K-Akt signalling*; *NFKB1*, *PRKCD*, and *PRKCB* genes to *AGE-RAGE signalling pathway in diabetic complications*, as well as *HTR1F* in *Serotonergic synapse* pathways as Yakut-specific adaptive loci (Fig. 2, supplementary table S5 and S6, Supplementary Material online). As regards those involved in *Thyroid hormone signalling*, *THRB* encodes for a nuclear receptor for the triiodothyronine (T3) hormone (Sayers et al. 2022), which has been shown to directly regulate adaptive thermogenesis in BAT, as well to induce browning process of white adipocytes by regulating the expression of the UCP1 uncoupling mitochondrial protein (Yau and Yen 2020; Zekri et al. 2021; Machado et al. 2022). *RXRA* was instead proved to interact with thyroid hormone receptors (TRs) forming heterodimers that strongly bind T3 response elements, thus enhancing the transcriptional activity of TRs (Zhang et al. 1992). *RCAN2* was identified as a T3-responsive gene (Miyazaki et al. 1996) and its knockout in mice was found to be related to a reduced body weight and white adipose mass compared to controls, a condition that is significantly enhanced in a regime of high-fat diet (Sun et al. 2011). Finally, the *SRC* proto-oncogene encodes for a tyrosine-protein kinase that i) enhances the expression of the cold-induced neuroprotective gene *RBM3* thus mediating neuroprotective effects of mild hypothermia (Yuan et al. 2021), ii) regulates fatty acid synthesis (Zhao et al. 2024), iii) modulates cold-induced activation of *TRPM8*, a receptor acting upstream of cold-induced thermogenesis, response to cold and thermoception processes (Manolache et al. 2020; Sayers et al. 2022), and iv) causes impairment of glucose-induced insulin secretion in pancreatic β-cells when its expression and/or activity is dysregulated (Sato et al. 2016). Interestingly, short-time sojourners in Antarctica have been reported to present increased production and uptake of thyroid hormones, a condition that results in reduced free T3/T4 circulating levels (Levy et al. 2013). This physiological adjustment, known as “polar T3 syndrome”, has been also observed in the Yakut population when comparing their levels of circulating T3/T4 and THS (thyroid-stimulating hormone) between summer and winter seasons, suggesting enhanced capacity to increase metabolic heat production during seasonal severe cold (Levy et al. 2013). Moreover, Yakut young adults displayed improved BAT thermogenesis in response to mild cooling when compared with control groups, a condition that is accompanied by warmer sternum temperatures, and higher metabolic rates (Levy et al. 2022). Therefore, adaptive evolution at *Thyroid hormone signalling* genes may represent one of the determinants of Yakut enhanced thyroid hormone uptake and the consequent augmented BAT activity, as well as of the fine-tuning of lipid/glucose metabolisms and insulin signalling in different districts of the body. Such a hypothesis is further supported by evidence collected for the other candidate adaptive genes reported in supplementary table S7 (Supplementary Material online), whose expression was proved to be up or down regulated in mice BAT, white adipose tissue, placenta and liver as physiological responses to cold exposure (Shore et al. 2013; Wang and Wahl 2014; Labbé et al. 2015; Ghosh et al. 2021). In line with this, also *PRKG1*, for which however selective events have been demonstrated to have occurred in other Siberian populations by previous studies (Cardona et al. 2014; Hallmark et al. 2019), is essential for cell differentiation in BAT, with its lacking being associated to

lower mitochondrial content and reduced amounts of adipogenic factors in mouse BAT, as well as to lower body temperatures than wild-type mice (Haas et al. 2009). Accordingly, suppression of *ERBB4* has been also proved to inhibit BAT thermogenesis and metabolic function (Yu et al. 2024). When the *Trendesetter* approach was used to validate the obtained results, adaptive evolution in Yakuts was confirmed especially for the *SRC*, *RCAN2*, *PRKG1*, and *ERBB4* genes (supplementary figs. S16 and S17, supplementary table S11, Supplementary Material online). Moreover, homozygous eQTLs associated to *THRB*, *SRC*, and *RCAN2* decreased/increased expression were found to reach outstanding and significantly higher frequency in the Yakut population with respect to the control group (supplementary figs. S12 and S13, supplementary table S10, Supplementary Material online), suggesting that these loci might play a relevant role in the modulation of the described adaptive responses.

As concerns genes included in the *AGE-RAGE signalling pathway in diabetic complications* and *Serotonergic synapse* pathways, variation at *NFKB1* was previously associated with increased risk of developing type 2 diabetes (Coto et al. 2018) and the gene silencing was proved to alleviate oxidative stress injury, insulin resistance, inflammation, and cell apoptosis in gestational hypertension mice (Guo et al. 2021). Furthermore, *PRKCB* activation is related to insulin-resistance mediated atherosclerosis (Li et al. 2013) and its upregulation has been detected in diverse tissues of diabetic individuals suggesting involvement of this protein in the development of insulin resistance (Chen et al. 2013). Accordingly, overexpression of *PRKCD* has been detected in obese humans, as well as in mice liver with hepatic insulin resistance characterized by decreased insulin signalling, enhanced lipogenic gene expression, and hepatosteatosis (Bezy et al. 2011). Finally, the *HTR1F* receptor was shown to regulate the secretion of glucagon in pancreatic alpha cells thought the binding with serotonin (Almaça et al. 2016). In fact, serotonin regulates energy metabolism in several others peripheral tissues, such as BAT and white adipose tissue, skeletal muscle and liver (Choi et al. 2020). Notably, it has been advanced that cold exposure significantly potentiates the maximal insulin responses for glucose uptake in mice BAT, white adipose tissue, and skeletal muscles, possibly due to enhanced responsiveness of peripheral tissues to insulin Vallerand et al. 1987; Labbé et al. 2015). Therefore, the involvement of *NFKB1*, *PRKCB*, *PRKCD*, and *HTR1F* genes in modulating Yakut-specific adaptive traits may constitute evidence for the genetic bases of enhanced insulin response with consequent glucose uptake, which is necessary to sustain the metabolic demand of peripheral tissues in extreme cold environments.

In accordance with this picture, the *INSR* insulin receptor was also found to be included in significant gene networks belonging to the *PI3K-Akt signalling* pathway, but in both Yakut and control populations (Fig. 2, supplementary table S5 and S6, Supplementary Material online). In detail, this gene encodes a member of the receptor tyrosine kinase family of proteins which bind insulin, activating downstream signalling and thus regulating glucose uptake, release, as well as the synthesis and storage of carbohydrates, lipids and proteins (Sayers et al. 2022). Moreover, *INSR* mutations have been associated to inherited severe insulin resistance syndromes, such as type A insulin resistance, Donohue and Rabson-Mendenhall syndromes (Sayers et al. 2022). Finally, *LASSI-signet* analyses similarly pointed out in both Yakut and control groups a significant

gene network participating to *Glycerolipid metabolism*, which includes a series of loci involved in the synthesis and digestion of lipids (e.g., *GPAT3, MGLL* and *PLPP3*) (Kanehisa and Goto 2000; Sayers et al. 2022) (Fig. 2, supplementary table S5 and S6, Supplementary Material online). Again, validation of the identified putative adaptive genes by means of the independent *Trendesetter* method provided support for Yakut-specific adaptive evolution of most of those belonging to the *AGE-RAGE signalling* and *Serotonergic synapse* pathways (i.e., *NFKB1, PRKCB, PRKCD, HTRF1),* as well as for a shared adaptive history between the Yakut and Russian groups as concerns loci that play a role in *Glycerolipid metabolism* (i.e., *GPAT3, PNLIP*, *MOGAT2*, *PLPP1*) (Fig. 4c and 4d, supplementary figs. S16, S17, and S18, supplementary table S11 and S12, Supplementary Material online). Overall, when the occurrence of eQTLs was investigated in these candidate genes, homozygous variants able to alter *INSR*, *NFKB1*, *PRKCB*, and *HTR1F* expression were observed at remarkably higher frequency in Yakuts than in the control population (supplementary figs. S12 and S13, supplementary table S10, Supplementary Material online).

Finally, by considering the bulk of adaptive responses to cold climate conditions inferred for the Yakut and control populations, we also explored the possible contribution of archaic introgression events in shaping the depicted evolutionary scenario. To this end, results from both *Sprime* and Haplostrips analyses showed that gene flow from archaic human species similarly impacted Yakut and control variation patterns at some of the identified candidate adaptive genes (Fig. 3, supplementary figs. S7, S8, and S9, Supplementary Material online), suggesting that introgression likely occurred before the divergence of their gene pools. In detail, results indicative of adaptive introgression events were obtained for genes associated with type 2 diabetes (Grarup et al. 2018), adiposity and insulin resistance (Keele et al. 2018; Tian et al. 2018) (*ADCY3*), lipid metabolism (*PLPP3*) (Sayers et al. 2022), tumours (*ATP1B2*) (Sun et al. 2013; Suguro et al. 2014), and immune functions (*CD247*) (Sayers et al. 2022). Among them, the most compiling pattern was observed for the *ADCY3* gene (Fig. 3 and supplementary table S7, Supplementary Material online), whose adaptive evolution in the Russian population was confirmed also by validation analyses (Fig. 3, supplementary fig S18c, supplementary table S12, Supplementary Material online). Interestingly, this locus has been previously reported to be up-regulated in mouse BAT after cold exposure and to be plausibly involved in the evolution of climate-related adaptations in populations of European ancestry (Sazzini et al. 2020).

In conclusion, the application of an integrated pipeline of analyses aimed at uncovering genetic signatures ascribable to a variety of selective events allowed us to provide new evidence on the genetic bases of polygenic adaptations likely involved in the modulation of physiological and metabolic responses to extreme cold environments evolved by different populations from high-latitude regions of the Eurasian continent. We also propose that some of these responses are compatible with an evolutionary scenario of adaptive introgression from both Denisovan and Neanderthal species, thus contributing to expand our knowledge regarding the biological implications of admixture between archaic and modern humans.

Materials and Methods

Dataset composition and curation

WGS data for the 25 Yakut individuals analysed in the present study were collected among those generated by Bergström et al. (2020). To contextualize Yakut genomic variation within the framework of human genetic diversity at large scale, we also assembled an extended dataset by merging WGS data available in public repositories and representative of populations of East Asian, Central Asian, European and Native American ancestries (Auton et al. 2015; Bergström et al. 2020) (supplementary table S1, Supplementary Material online). We merged these data using functions implemented in the PLINK software version 1.9 (Purcell et al. 2007) by performing QC filtering to reduce the possible bias associated with data quality issues. Specifically, we prepared the dataset to be subjected to genotype-based population structure analyses, such as PCA and ADMIXTURE, by retaining those SNVs that: i) presented less than 5% of missing data, ii) respected the Hardy-Weinberg Equilibrium (HWE) according to the stringent p-value threshold obtained after Bonferroni correction for multiple testing (i.e., SNVs showing p-values < 9.78×10^{-9} were removed), iii) presented a Minor Allele Frequency (MAF) greater than 1%, and iv) were in low LD between each other (i.e., showed a r^2 score minor or equal to 0.2). We also excluded those individuals that presented more than 5% of missing data and identity-by-descent (IBD) kinship scores (PiHat) greater than 0.27 after calculation of the degree of recent shared ancestry for each pair of subjects to remove closely related individuals to the second degree, as suggested for datasets including populations with relatively low effective population size and appreciable inbreeding levels (Ojeda-Granados et al. 2022). In parallel, to carry out haplotype-based population structure analyses (i.e., ChromoPainter and fineSTRUCTURE) we phased a copy of the dataset in which we retained low-frequency variants (i.e., showing MAF ≤ 1%) and SNVs in LD. The phasing procedure was performed using the SHAPEIT software v2.r904 (Delaneau et al. 2013), as well as HapMap phase 3 recombination maps and the 1000 Genomes Project dataset as a reference panel (Auton et al. 2015). Furthermore, in order to avoid loss of data (in terms of number of available SNVs) due to the described merging procedure, for populations submitted to SMC++ analysis and selection scans we directly extracted WGS data for the filtered individuals from both the unphased and phased HGDP datasets (available at https://ngs.sanger.ac.uk/production/hgdp/hgdp_wgs.20190516/) and we applied the following QCs to remove SNVs: i) with more than 5% of missing data, ii) showing significant deviation from the HWE (i.e., variants with p-value < 2.51×10^{-8}), iii) presenting ambiguous A/T or G/C alleles, iv) with mapping quality ≤ 20, and v) with values of sequence depth < 30.

Fine scale population structure analyses and identification of a control population

The extended unphased dataset was first subjected to PCA and ADMIXTURE analyses to check for data consistency after the merging procedure and to investigate broad patterns of population genetic structure. Specifically, we carried out PCA by employing the *smartpca* function implemented in the EIGENSOFT package (Patterson et al. 2006), while ADMIXTURE analysis was run using the ADMIXTURE software

version 1.3.0 (Alexander et al. 2009) by considering a progressively increasing number of hypothetic ancestry components (Ks) ranging from 2 to 13 and by choosing the configuration that best fits to the data based on the lowest 5-fold cross-validation (CV) error associated to each K. ChromoPainter and fineSTRUCTURE analyses were then performed to explore patterns of fine scale population structure and to identify genetically homogeneous groups to be submitted to downstream analyses. For this purpose, we used the phased extended dataset to estimate the mutation rate (mu) and the Ne parameters using the *–in* and *–iM* flags implemented in the ChromoPainterv2 software (Lawson et al. 2012) based on a subset of four representative chromosomes (i.e., 4, 10, 15, and 22) and five representative individuals for each population. We then calculated average values for the estimates obtained for each individual and for each of these chromosomes, and we used them to process data from all the 22 chromosomes included in the dataset. In particular, we treated as *donor* and *recipient* each individual in turn using the *0 0 -a 0 0* option and the obtained co-ancestry matrices were subjected to the fineSTRUCTURE algorithm version 2.1.3 to identify population clusters characterized by high internal genetic homogeneity by relying on haplotype sharing patterns between pairs of individuals. In detail, we corrected the fineSTRUCTURE estimation based on the *c* value previously calculated with ChomoPainter. Moreover, we set the flags *-x*, *-y* and *-z* respectively at 3,000,000, 2,000,000 and 10,000 for the primary step of the analysis. Finally, we run the estimation of the tree topology setting the *–x* and *–t* flags at 1,000,000 iterations. By relying on the obtained results, we retained 25 Russian individuals to be used as a control group characterized by relatively shared ancestry with the Yakuts to perform analyses aimed at inferring their fluctuations in Ne and adaptive events.

Inferring Ne fluctuations, split time and admixture events in Yakut and control populations

To infer Ne history of Yakuts and of the selected control group, their respective unphased datasets were analysed with the SMC++ algorithm (Terhorst et al. 2017). Among the methods that rely on Sequentially Markovian Coalescent (SMC) approaches for estimating the demographic history of populations, SMC++ is indeed preferred to process large datasets since it is capable to analyse more than 20 haploid genomes simultaneously (Sellinger et al. 2021). Initially, each genomic dataset was converted to the required format using the *vcf2smc* function by incorporating six different distinguished lineages and the negative mask file for the HGDP genomes, which was obtained by complementing the positive mask file available at https://ngs.sanger.ac.uk/production/hgdp/hgdp_wgs.20190516/accessibility-mask/). Subsequently, the *estimate* function was applied to fit a population size history to each dataset, by considering a mutation rate of 1.25×10^{-8} per site per generation (Terhorst et al. 2017). The analysis spanned a period between 1,000,000 and 1,000 years ago, corresponding to 34,000 and 34 generations respectively, assuming a generation time of 29 years (Bergström et al. 2020). The Yakut and control datasets were then combined into a single dataset to estimate the split time between their ancestral populations, by using the *split* function implemented in the SMC++ program and by assuming a model with no post-split gene flow between populations. Finally, informed by the results of ADMIXTURE analysis, we used the *qp3pop* function implemented in the

ADMIXTOOLS package (Reich et al. 2009; Patterson et al. 2012) to computed the f3-statistics in the form of *f3(Test; Source1, Source2)*, in order to formally test for the presence of admixture events involving the considered Yakut and control populations. Significant admixture events (i.e., population trios yielding a Z-score smaller than −3) were subsequently dated with ALDER (Loh et al. 2013) by using a generation time of 30 years for the conversion of estimated dates.

Integrating multiple methods to detect genomic signatures of adaptive evolution

To detect genomic signatures ascribable to a wide range of selective events, we assembled a pipeline of analyses by relying on three different likelihood-, network-, and Machine Learning-based approaches, which we combined to pinpoint biological traits plausibly impacted by the action of positive natural selection, including complex ones for which adaptive evolution was enabled by selection having simultaneously occurred on many loci contributing to the same biological function. We first took advantage from the *LASSI* method developed by Harris & DeGiorgio (2020), which can distinguish between strong and weak signatures left by natural selection on the haplotype frequency spectrum calculated for sliding genomic windows. Such a method was chosen among several selection statistics since i) it was demonstrated to have an improved power with respect to traditional selection scans in the identification of selective events under a vast range of selection coefficients and demographic models (Harris and DeGiorgio 2020); ii) it can be used to easily distinguish hard selective sweeps from other selective events by filtering results in function of the *m* parameter (i.e., the number of sweeping haplotypes that are supposed to carry putative adaptive variants, which is equal to one for hard sweeps and ≥ 2 in the other cases). The *LASSI* statistic was computed across genomic windows after having estimated the average number of SNVs included in all the inferred haplotype blocks obtained for the Yakut and control groups. In details, we first run the *--blocks* function implemented in the PLINK v1.90b5.2 package on the unphased datasets to infer haplotype blocks and the number of SNVs included in each of them. We then calculated the average number of SNVs encompassing all the inferred haplotype blocks and we used such a value to divide the genome into sliding windows on which the *LASSI* scan was performed. Furthermore, the overall distribution of the *LASSI* statistic was filtered according to the *m* parameter (i.e., by considering $m \geq 2$) to exclude genomic windows affected by hard selective sweeps and by retaining chromosomal intervals showing the greatest likelihood values associated to each gene. The distribution of selection scores filtered according to these criteria was then used to inform the second step of the pipeline of analyses, which relied on the *signet* method (Gouy et al. 2017). Such an approach enabled us to identify networks of genes contributing to the same biological function and collectively characterized by moderate-to-weak signatures of natural selection, thus being potentially informative of selective events occurred according to a polygenic adaptation model. We run the *signet* algorithm using the R version 3.6.3 and by performing 20,000 iterations to generate the null distributions of the highest scoring subnetwork (HSS). After comparison of HHS scores obtained for each single gene network with the null distribution, networks showing FDR < 0.05 were retained as significant groups of candidate adaptive loci participating to

the same functional pathway according to information stored in the KEGG database (Kanehisa and Goto 2000). Finally, the last step of the implemented pipeline of analyses consisted in the utilization of the *Trendsetter* supervised multinomial regression classifier that evaluated the distribution of diverse selection/summary statistics across genomic windows to predict both strong and weak signatures ascribable to the action of natural selection (Mughal and DeGiorgio 2019). Such an approach was used to further support at the single gene level the weak selection signals identified with the combination of *LASSI* and *signet* methods and/or to point out additional putative adaptive loci showing functional relationships with them. For this purpose, we first trained the classifier using three sets of simulations, which were built for each possible predictable class (i.e., hard sweeps, soft sweeps and neutral evolution; see the following section for further details), and comprising 5,000 simulations each. Moreover, we generated three additional test sets each including 1,000 simulations for each class to build calibration curves and to evaluate the accuracy of the trained classifier by using the packages *calibration* and *metrics* implemented in the Scikit-learn Python library. Finally, we used the trained algorithm on the real Yakut and control datasets to classify each region of the genome in hard sweeps, soft sweeps or neutral evolutionary scenarios.

Generating genomic simulations to train the *Trendsetter* classifier

Simulations used to train the *Trendsetter* classifier were generated with the software *discoal* (Kern and Schrider 2016) by considering *Ne* values estimated with SMC++. For each possible evolutionary scenario, we simulated the total number of haplotypes for each population (i.e., sample size) with a chosen length of 1.1 Mb (Mughal and DeGiorgio 2019), assuming a mutation rate (*theta*) of 1.25 x 10^{-8} per site per generation, and a population recombination rate (*rho*) of 1 x 10^{-8}. The values of both *theta* and *rho* were scaled by the length of the simulated haplotype and by the current effective population size (N_o). In addition to these basic parameters, simulating the neutral evolutionary scenario required population size changes to be specified (i.e., *-en* flag). Soft and hard selective sweeps were instead modelled as stochastic events, with beneficial mutations introduced at the centre of the simulated region and assuming they have been subjected to the action of natural selection with a per-generation selection coefficient (i.e., *s*) randomly drawn from a uniform distribution within the range [0.005 - 0.5] (Mughal and DeGiorgio 2019). For all selection simulations, the timing of fixation of the adaptive allele was randomly selected from a uniform distribution spanning from 0 (i.e., present) to 1,200 (Mughal and DeGiorgio 2019) and to 1,034 (Pitulko et al. 2004) generations in the past for the control and Yakut population, respectively. More in detail, these times were chosen since they identify a reliable approximation for the earliest Out of Africa migrations of modern humans, as well as the most ancient evidence for colonization of the Yakutia region by ancestral *H. sapiens* populations (Pitulko et al. 2004). Furthermore, to simulate soft sweeps from standing genetic variation, the initial frequency of the beneficial allele was randomly sampled from a uniform distribution within the interval [0.01 - 0.10] (Mughal and DeGiorgio 2019).

Detecting archaic adaptive introgression events

To evaluate the role played by events of archaic introgression in shaping the genomic variability of the Yakut population, we applied the *Sprime* method (Browning et al. 2018) to the respective phased dataset. Subsequentially, in order to shortlist those introgression events that have potentially mediated also adaptive processes in Yakut ancestors, we crosschecked results obtained with *Sprime* with those suggestive of adaptive evolution as supported by the selection scans previously described (Gouy et al. 2017; Harris and DeGiorgio 2020). In details, we crosschecked putative introgressed chromosomal segments with genomic windows ranking in the top 1% of values of the *LASSI* statistic (subjected to enrichment analysis using the STRING toll available at https://string-db.org/) and with candidate adaptive genes supported by the combined results of *LASSI-signet* methods and/or by *Trendsetter* analysis. Moreover, we further validated putative archaic introgression signatures detected on Yakut candidate adaptive loci by comparing their haplotype structure with those observable from archaic genomes (i.e., Vindija Neanderthal, Altai Neanderthal and Altai Denisovan genomes available respectively at http://cdna.eva.mpg.de/neandertal/Vindija/VCF/Vindija33.19/, ../Altai/, and ../Denisova/) thought the Haplostrips software (Marnetto and Huerta-Sánchez 2017). In particular, the algorithm was run on Yakuts (i.e., the test population), Russians (i.e., the control population selected to have a relatively shared ancestry with Yakuts), and Yoruba (i.e., the outgroup population), as well as on archaic sequences by treating them in tourn as the possible source of introgression. Moreover, we first run Haplostrips on the entire genes and subsequentially on the overlapped genomic windows associated with positive values of the *LASSI* statistic (i.e., those targeted by natural selection according to Harris and DeGiorgio 2020), as previously described (Ferraretti et al. 2024).

Investigating haplotype and genotype frequencies for candidate adaptive eQTLs

To further validate adaptive events occurred specifically in the Yakut population, we investigated possible differences in allele/haplotype frequencies between this human group and the selected control population. For this purpose, we performed a Fisher exact test using the *--assoc fisher* function implemented in the PLINK package version v1.90b5.2 (Purcell et al. 2007) and by considering all SNVs included in the identified candidate adaptive genes. Loci presenting significant differences in frequency were identified as those associated with corrected FDR ≤ 0.05 by using the *p.adjust* function in the R version 4.3.1. We further shortlisted the variants most plausibly involved in mediating the inferred adaptations by focusing on those annotated in the GTEx portal (https://www.gtexportal.org/home/) as single tissue eQTLs that influence the expression of target genes in different body districts previously proposed to present biological responses to cold exposure in both mouse and human models and/or producing hormones able to modulate such mechanisms (i.e., white adipose tissue, thyroid and skeletal muscle) (Ribeiro et al. 2001; Blondin et al. 2015; Xu et al. 2019; Zekri et al. 2021). In details, we reconstructed haplotypes including the significant eQTLs in the examined populations by: i) extracting genomic regions of interest by considering windows pointed out

by *LASSI* analyses as putative targets of natural selection (i.e., those associated with positive values of the likelihood statistic) (Harris and DeGiorgio 2020) and including significant eQTLs, ii) using the *--blocks* function implemented in PLINK v1.90b5.2 to reconstruct haplotype blocks, iii) filtering variants included in these blocks by retaining only those pairs presenting high LD values (i.e., $r^2 \geq 0.8$). Finally, we extracted data for the SNVs composing each haplotype from phased datasets and we calculated haplotype frequencies with the *GHap* R package (Utsunomiya et al. 2016). Genotype frequencies for eQTLs included in the inferred haplotypes were also obtained with the -- *freqx* function implemented in PLINK v1.90b5.2.

Funding

M.A. was supported by the Ministero della Ricerca PRIN2020_2020TACEZR_005 grant to M.S.

Conflict of Interest

The authors have no conflicts of interest to declare.

Data Availability

WGS data analysed are publicly available at the repositories cited in the text (Auton et al. 2015; Bergström et al. 2020). The codes for running the algorithms cited in the text are available at the related GitHub repositories. The eQTLs information used to perform haplotype/genotype frequency analyses displayed in supplementary figs. S12 and S13 and reported in supplementary table S8 and S9 were downloaded from the GTEx Portal (https://www.gtexportal.org/home/).

Author Contributions

Data curation, formal analysis, investigation, writing original draft, G. F., M. A., S. S.; investigation, writing original draft, R.C., S.P.V.; conceptualization, resources, supervision, funding acquisition, methodology, writing, review and editing original draft, M. S.

Figures

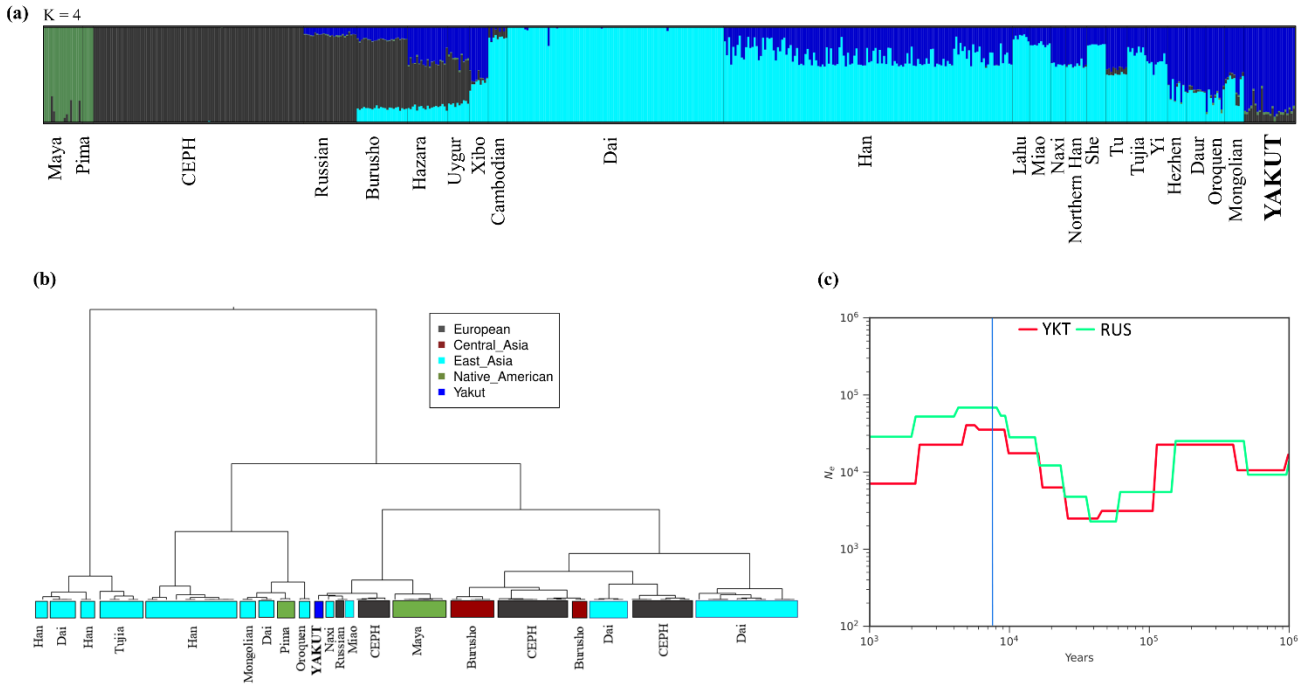


Figure 1. Genetic structure and *Ne* inferences. (a) ADMIXTURE profiles inferred with best predictive accuracy (K=4) for the individuals included in the extended unphased dataset. Samples belonging to the Yakut population are characterized by a predominant ancestry component (i.e., blue), which reaches appreciable proportions also in the majority of the populations from East Asia (e.g., Xibo) Central Asia (e.g., Burusho), and Russia. The ancestry component reaching the greatest frequencies in the majority of the East Asian populations (i.e., light-blue) was instead detectable in minimal proportions in the Yakuts. Additionally, the typical European and Native American components turned out to be recognizable also in the Yakut cluster. **(b)** Dendrogram resulting from ChormoPainter/fineSTRUCTURE analyses performed on the phased extended dataset. The clusters were named according to the population that was mostly represented in each of them, and they are coloured according to the geographic macro-area of origin reported in the legend. The nodes depicted in the plot are characterized by a nodal support ≥ 0.80. The extensive dendrogram displaying all the individuals belonging to a given cluster is reported in Supplementary Figure S5, Supplementay Material online. **(c)** Plot displaying the result of SMC++ estimations of the effective population size (Ne) changes in both Yakut (i.e., YKT, red curve) and Russian (i.e., RUS, green curve) populations. On the x-axis is reported the considered time interval from 10^6 to 10^3 years ago. On the y-axis are instead displayed the inferred Ne values, ranging from 10^7 to 10^2 estimated individuals. From 100k years ago both the curves start to decline, testifying a decreasing in Ne that starts to recover at 40k (RUS) and 30k (YKT) years ago. Such a trend is typical of non-African population since it attests the drop in the effective population size experienced by *H. sapiens* population during the Out of Africa migrations. The vertical blue line indicates the inferred split time between Russians and Yakuts dated between 7,500 and 8,000 years ago. The inferences of Ne for the control population suggests that this human group after the Out of Africa demographic drop experienced a constant increasing in the effective population size until 8k years ago.

Subsequentially the curve starts to slightly decrease settling then at values of 20k estimated individuals at 1k years ago. A similar trend is observed for the Yakuts although especially after the split time, the effective population size markedly dops finally reaching 7k estimated individuals at 1k year ago, suggesting both geographic and genetic isolation of this human group.

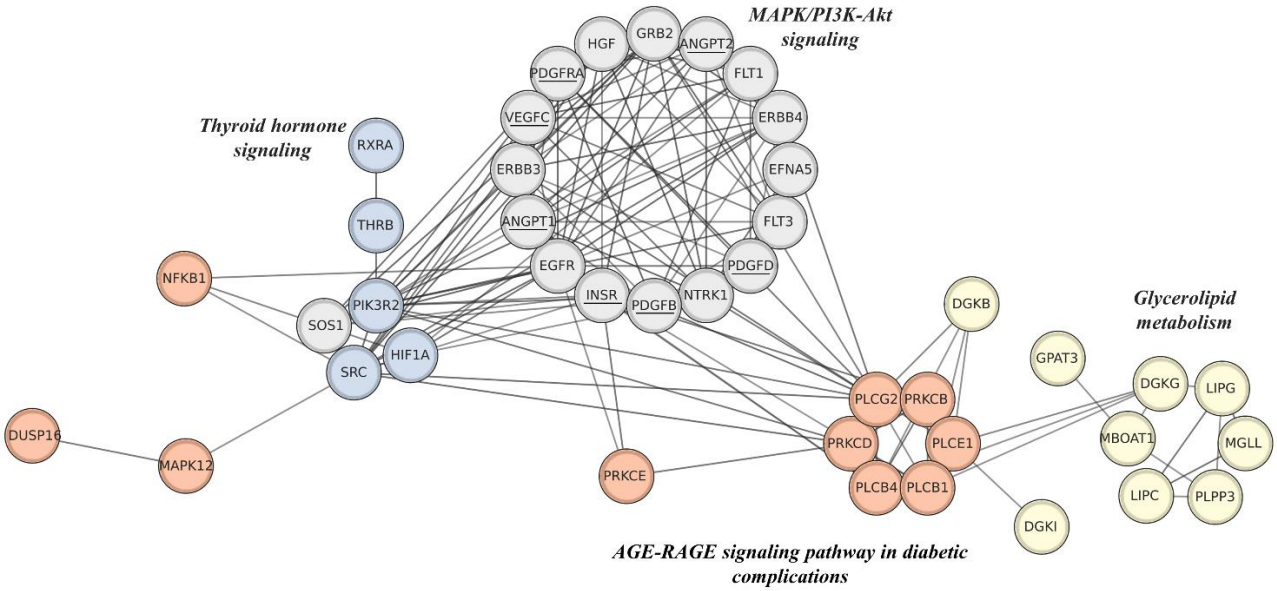


Figure 2. Selective signatures ascribable to the model of polygenic adaptation identified with the LASSI-*signet* approach and plausibly evolved in response to cold-induced selective pressures. Networks of genes belonging to the *Thyroid hormone* pathway and to the *AGE-RAGE signalling pathway in diabetic complications*, whose variation patterns resulted significantly impacted by the action of natural selection exclusively in the Yakut population, are displayed as light blue and orange circles, respectively. Significant gene networks within the *MAPK/PI3K-Akt signalling* and *Glygerolipid metabolism* pathways detected in both Yakut and control human groups are instead shown with grey and yellow circles, respectively. In detail, all the genes reported in the figure and included in the *Glycerolipid metabolism* pathway showed adaptive evolution in both Yakuts and control groups, while only the underlined genes in the *MAPK/PI3K-Akt signalling* presented the same shared pattern. Gene networks were built using the STRING app implemented in Cytoscape version 3.10.3 by setting the STRING high confidence score of 0.7. Genes belonging to the same pathway were displayed as circular networks (i.e., adopting the circular layout in Cytoscape) in order to highlight the functional connections inferred among different significant *signet* networks/pathways.

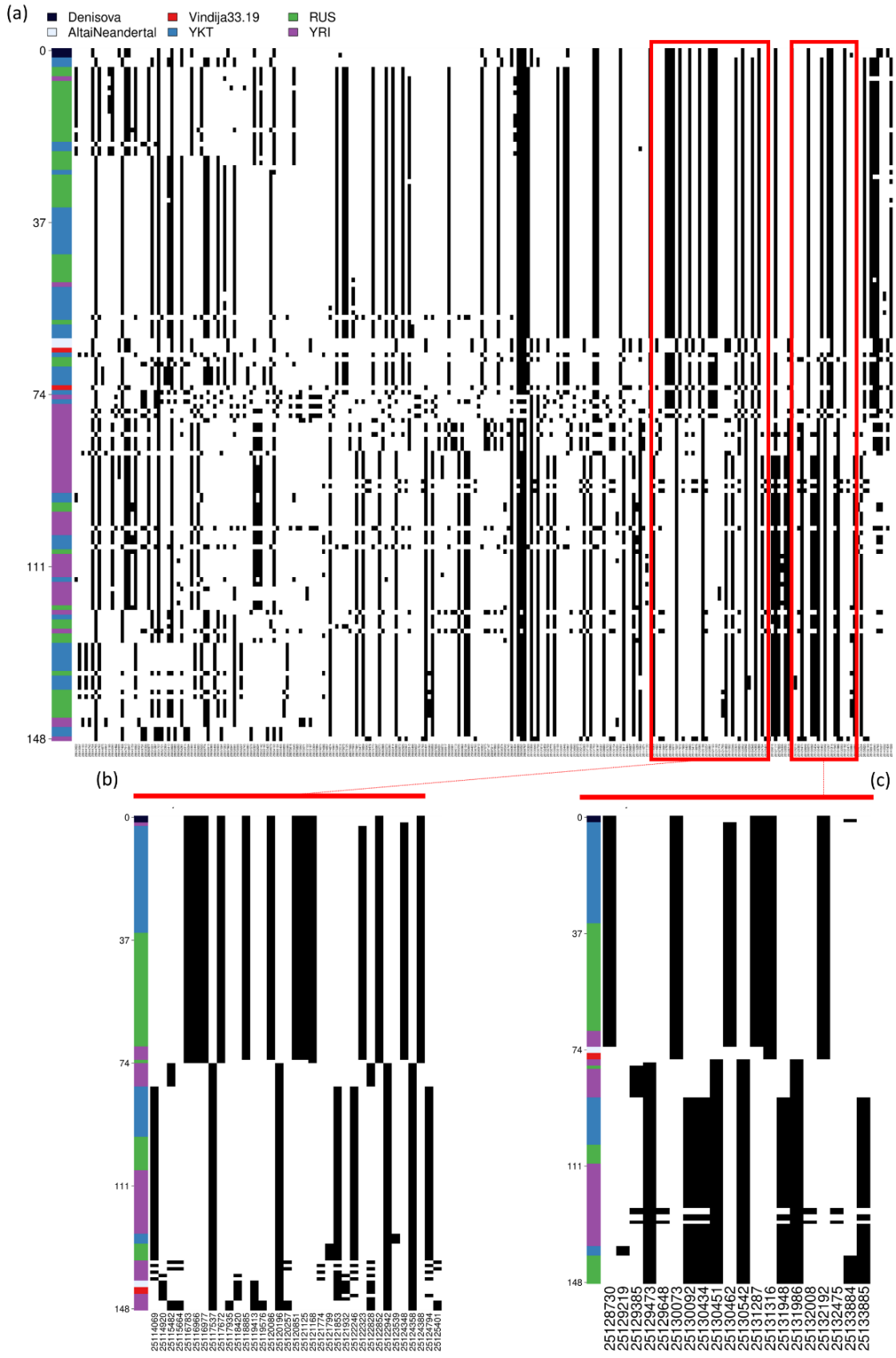


Figure 3. Haplotype structure reconstructed for the adaptive introgressed *ADCY3* gene. (a) Haplostrips plot showing patterns of haplotype diversity in the *ADCY3* gene among modern human populations (i.e., Yakut, Russian, Yoruba) and archaic samples (i.e., Denisovan Altai, Neanderthal Altai and Neanderthal

Vindija). Single haplotypes (reported in rows) are ranked with respect to the number of differences with respect to the Denisovan sequence. SNVs are displayed in columns with the ancestral and derived states reported in white and black colours, respectively. The great majority of haplotypes belonging to both Yakut (i.e., blue) and control (i.e., green) populations were found to present the smallest amount of haplotype pairwise differences with the Denisovan ones (i.e., black), even with respect to Vindija (i.e., red) and Altai (i.e., light blue) Neanderthal samples. Haplotypes of the Yoruba outgroup population (i.e., purple) cluster mainly at the bottom of the plot, attesting a remarkable diversity of this human group compared to the Denisovan sequence and showing opposite haplotype patterns with respect to non-African populations. Red squares localize the positions of overlapped putative adaptive windows identified by *LASSI* in the *ADCY3* gene. **(b-c)** Haplostrips plots build for overlapped genomic windows presenting a variation pattern indicative of adaptive evolution according to the *LASSI* statistic in the Yakut population and falling within the *ADCY3* gene. In both the genomic regions are maintained similar patterns of haplotype variation with respect to those observed across the entire *ADCY3* gene. In detail, the majority of haplotypes belonging to Yakut and control populations are extremely conserved, presenting the smallest amount of pairwise differences with respect to Denisovan haplotypes, thus suggesting the occurrence of archaic adaptive introgression at the *ADCY3* locus in both these human groups.

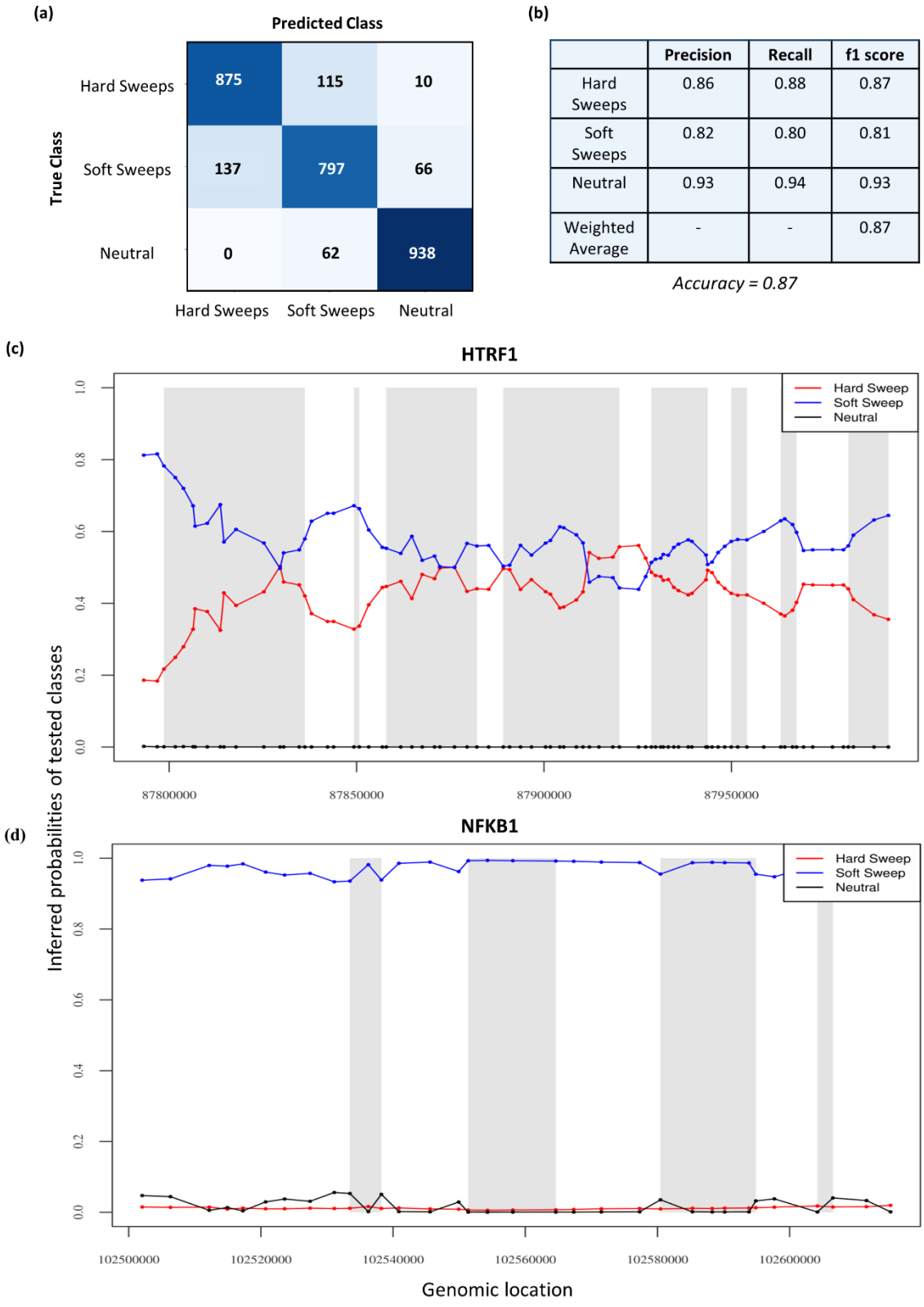


Figure 4. Confusion matrix and selective signatures supported by the *Trendsetter* classifier obtained for the Yakut population. (**a**) Confusion matrix built relaying on results obtained after the application of the trained *Trendsetter* classifier on the validation set of genomic simulations (i.e., comprehending 1,000 simulated replicates for each tested evolutionary scenario). Along the diagonal are reported the numbers of

correct classifications assigned by the algorithm, while in the remaining rows/columns are shown the misclassifications. **(b)** Table displaying precision, recall, and f1-score values calculated for each tested class, as well as the f1-score weighted average and global accuracy values. **(c-d)** Distribution of *Trendsetter* predicted probabilities of being classified as hard sweeps (i.e., red curve), soft sweeps (i.e., blue curve) and neutral regions (i.e., black curve) across the putative adaptive *HTRF1* and *NFKB1* genes in the Yakut population, respectively. For both the plots, the x axis reports the positions of the central SNV in each genomic window identified according to the *Trendsetter* approach on the *HTRF1* and *NFKB1* genes, respectively. On the y-axis are instead shown the values of *Trendsetter* predicted probabilities associated to each of the three tested classes. The grey rectangles in the background represent genomic regions in the considered genes that present values indicative of the action of natural selection according to the *LASSI* statistic. **(c)** The output probabilities inferred for all genomic windows falling in the *HTRF1* gene and associated to the soft selective sweep class are substantially more elevated with respect to the neutral one, suggesting that variation patterns at this locus have been intensively shaped by selective events. More in detail, soft sweeps probabilities are dominant for the entire gene (i.e., ranging from 0.8 to 0.5) except for a region in the middle of it, in which hard sweeps probabilities overcome them. **(d)** Probabilities of soft sweeps classification obtained across the *NFKB1* gene were outstanding with respect to both hard sweeps and neutral scenarios, reaching values constantly higher than 0.90, thus supporting signatures of natural selection for this gene conformed with the soft selective sweeps model.

References

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664.

Almaça J, Molina J, Menegaz D, Pronin AN, Tamayo A, Slepak V, Berggren PO, Caicedo A. 2016. Human Beta Cells Produce and Release Serotonin to Inhibit Glucagon Secretion from Alpha Cells. *Cell Rep* 17:3281–3291.

Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, et al. 2015. A global reference for human genetic variation. *Nature* 526:68–74.

Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, Chen Y, Felkel S, Hallast P, Kamm J, et al. 2020. Insights into human genetic variation and population history from 929 diverse genomes. *Science (1979)* 367.

Bezy O, Tran TT, Pihlajamäki J, Suzuki R, Emanuelli B, Winnay J, Mori MA, Haas J, Biddinger SB, Leitges M, et al. 2011. PKCδ regulates hepatic insulin sensitivity and hepatosteatosis in mice and humans. *Journal of Clinical Investigation* 121:2504–2517.

Blondin DP, Labbé SM, Phoenix S, Guérin B, Turcotte ÉE, Richard D, Carpentier AC, Haman F. 2015. Contributions of white and brown adipose tissues and skeletal muscles to acute cold-induced metabolic responses in healthy men. *Journal of Physiology* 593:701–714.

Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM. 2018. Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell* 173:53-61.e9.

Cardona A, Pagani L, Antao T, Lawson DJ, Eichstaedt CA, Yngvadottir B, Shwe MTT, Wee J, Romero IG, Raj S, et al. 2014. Genome-wide analysis of cold adaptation in indigenous Siberian populations. *PLoS One* 9.

Chen J, Meng Y, Zhou J, Zhuo M, Ling F, Zhang Y, Du H, Wang X. 2013. Identifying candidate genes for type 2 diabetes mellitus and obesity through gene expression profiling in multiple tissues or cells. *J Diabetes Res* 2013.

Chen Y, Huang L, Qi X, Chen C. 2019. Insulin receptor trafficking: Consequences for insulin sensitivity and diabetes. *Int J Mol Sci* 20.

Chevychelov AP, Bosikov NP. 2010. Natural Conditions. In: Troeva EI, Isaev AP, Cherosov MM, Karpov NS, editors. The Far North: Plant Biodiversity and Ecology of Yakutia. Dordrecht: Springer Netherlands. p. 1–23. Available from: https://doi.org/10.1007/978-90-481-3774-9_1

Choi W, Moon JH, Kim H. 2020. Serotonergic regulation of energy metabolism in peripheral tissues. *Journal of Endocrinology* 245:R1–R10.

Coto E, Díaz-Corte C, Tranche S, Gómez J, Alonso B, Iglesias S, Reguero JR, López-Larrea C, Coto-Segura P. 2018. Gene variants in the NF-KB pathway (NFKB1, NFKBIA, NFKBIZ) and their association with type 2 diabetes and impaired renal function. *Hum Immunol* [Internet] 79:494–498. Available from: https://www.sciencedirect.com/science/article/pii/S0198885918300818

Delaneau O, Zagury JF, Marchini J. 2013. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 10:5–6.

Ferraretti G, Abondio P, Alberti M, Dezi A, Sherpa PT, Cocco P, Tiriticco M, Di Marcello M, Gnecchi-Ruscone GA, Natali L, et al. 2024. Archaic introgression contributed to shape the adaptive modulation of angiogenesis and cardiovascular traits in human high-altitude populations from the Himalayas. *Elife* [Internet] 12. Available from: https://elifesciences.org/articles/89815

Ferraretti G, Rill A, Abondio P, Smith K, Ojeda-Granados C, De Fanti S, Alberti M, Izzi M, Sherpa PT, Cocco P, et al. 2025. Convergent evolution of complex adaptive traits modulates angiogenesis in high-altitude Andean and Himalayan human populations. *Commun Biol* [Internet] 8:377. Available from: https://doi.org/10.1038/s42003-025-07813-6

Ghosh S, Park CH, Lee J, Lee N, Zhang R, Huesing C, Reijnders D, Sones J, Münzberg H, Redman L, et al. 2021. Maternal cold exposure induces distinct transcriptome changes in the placenta and fetal brown adipose tissue in mice. *BMC Genomics* 22.

Gnecchi-Ruscone GA, Abondio P, De Fanti S, Sarno S, Sherpa MG, Sherpa PT, Marinelli G, Natali L, Di Marcello M, Peluzzi D, et al. 2018. Evidence of polygenic adaptation to high altitude from Tibetan and Sherpa genomes. *Genome Biol Evol* 10:2919–2930.

Gouy A, Daub JT, Excoffier L. 2017. Detecting gene subnetworks under selection in biological pathways. *Nucleic Acids Res* 45.

Grarup N, Moltke I, Andersen MK, Dalby M, Vitting-Seerup K, Kern T, Mahendran Y, Jørsboe E, Larsen CVL, Dahl-Petersen IK, et al. 2018. Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. *Nat Genet* 50:172–174.

Guo Y, Liu Z, Wang M. 2021. NFKB1-mediated downregulation of microRNA-106a promotes oxidative stress injury and insulin resistance in mice with gestational hypertension. *Cytotechnology* 73:115–126.

Haas B, Mayer P, Jennissen K, Scholz D, Diaz MB, Bloch W, Herzig S, Fässler R, Pfeifer A. 2009. Protein Kinase G Controls Brown Fat Cell Differentiation and Mitochondrial Biogenesis. *Sci Signal* [Internet] 2:ra78–ra78. Available from: https://doi.org/10.1126/scisignal.2000511

Hallmark B, Karafet TM, Hsieh PH, Osipova LP, Watkins JC, Hammer MF. 2019. Genomic evidence of local adaptation to climate and diet in indigenous Siberians. *Mol Biol Evol* 36:315–327.

Harris AM, DeGiorgio M. 2020. A likelihood approach for uncovering selective sweep signatures from haplotype data. *Mol Biol Evol* 37:3023–3046.

Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, Project 1000 Genomes, Sella G, Przeworski M. 2011. Classic Selective Sweeps Were Rare in Recent Human Evolution. *Science (1979)* [Internet] 331:920–924. Available from: https://doi.org/10.1126/science.1198878

Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Available from: http://www.genome.ad.jp/kegg/

Keele GR, Prokop JW, He H, Holl K, Littrell J, Deal A, Francic S, Cui L, Gatti DM, Broman KW, et al. 2018. Genetic Fine-Mapping and Identification of Candidate Genes and Variants for Adiposity Traits in Outbred Rats. *Obesity* 26:213–222.

Kern AD, Schrider DR. 2016. Discoal: Flexible coalescent simulations with selection. *Bioinformatics* 32:3839–3841.

Kharkov VN, Stepanov VA, Medvedeva OF, Spiridonova MG, Maksimova NR, Nogovitsina AN, Puzyrev VP. 2008. The origin of Yakuts: Analysis of the Y-chromosome haplotypes. *Mol Biol* 42:198–208.

Konstantinova I, Nikolova G, Ohara-Imaizumi M, Meda P, Kučera T, Zarbalis K, Wurst W, Nagamatsu S, Lammert E. 2007. EphA-Ephrin-A-Mediated β Cell Communication Regulates Insulin Secretion from Pancreatic Islets. *Cell* [Internet] 129:359–370. Available from: https://www.sciencedirect.com/science/article/pii/S0092867407003686

Kuzmin Y V. 2008. Siberia at the Last Glacial Maximum: Environment and archaeology. *Journal of Archaeological Research* 16:163–221.

Labbé SM, Caron A, Bakan I, Laplante M, Carpentier AC, Lecomte R, Richard D. 2015. In vivo measurement of energy substrate contribution to cold-induced brown adipose tissue thermogenesis. *FASEB Journal* 29:2046–2058.

Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. *PLoS Genet* 8.

Levy SB, Klimova TM, Zakharova RN, Fedorov AI, Fedorova VI, Baltakhinova ME, Bondy M, Atallah D, Thompson-Vasquez J, Dong K, et al. 2022. Brown adipose tissue thermogenesis among young adults in northeastern Siberia and Midwest United States and its relationship with other biological adaptations to cold climates. *American Journal of Human Biology* 34.

Levy SB, Leonard WR, Tarskaia LA, Klimova TM, Fedorova VI, Baltakhinova ME, Krivoshapkin VG, Snodgrass JJ. 2013. Seasonal and socioeconomic influences on thyroid function among the Yakut (Sakha) of Eastern Siberia. *American Journal of Human Biology* 25:814–820.

Li Q, Park K, Li C, Rask-Madsen C, Mima A, Qi W, Mizutani K, Huang P, King GL. 2013. Induction of vascular insulin resistance and endothelin-1 expression and acceleration of atherosclerosis by the overexpression of protein kinase C-β isoform in the endothelium. *Circ Res* 113:418–427.

Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B. 2013. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* 193:1233–1254.

Machado SA, Pasquarelli-do-Nascimento G, da Silva DSS, Farias GR, de Oliveira Santos I, Baptista LB, Magalhães KG. 2022. Browning of the white adipose tissue regulation: new insights into nutritional and metabolic relevance in health and diseases. *Nutr Metab (Lond)* 19.

Manolache A, Selescu T, Maier GL, Mentel M, Ionescu AE, Neacsu C, Babes A, Szedlacsek SE. 2020. Regulation of TRPM8 channel activity by Src-mediated tyrosine phosphorylation. *J Cell Physiol* 235:5192–5203.

Marnetto D, Huerta-Sánchez E. 2017. Haplostrips: revealing population structure through haplotype visualization. *Methods Ecol Evol* 8:1389–1392.

Miyazaki T, Kanou Y, Murata Y, Ohmori S, Niwa T, Maeda K, Yamamura H, Seo H. 1996. Molecular Cloning of a Novel Thyroid Hormone-responsive Gene, ZAKI-4, in Human Skin Fibroblasts*. *Journal of Biological Chemistry* [Internet] 271:14567–14571. Available from: https://www.sciencedirect.com/science/article/pii/S0021925818468378

Mughal MR, DeGiorgio M. 2019. Localizing and classifying adaptive targets with trend filtered regression. *Mol Biol Evol* 36:252–270.

Ojeda-Granados C, Abondio P, Setti A, Sarno S, Gnecchi-Ruscone GA, González-Orozco E, De Fanti S, Jiménez-Kaufmann A, Rangel-Villalobos H, Moreno-Estrada A, et al. 2022. Dietary, Cultural, and Pathogens-Related Selective Pressures Shaped Differential Adaptive Evolution among Native Mexican Populations. *Mol Biol Evol* 39.

Pakendorf B, Wiebe V, Tarskaia LA, Spitsyn VA, Soodyall H, Rodewald A, Stoneking M. 2003. Mitochondrial DNA evidence for admixed origins of Central Siberian populations. *Am J Phys Anthropol* 120:211–224.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient Admixture in Human History. *Genetics* [Internet] 192:1065–1093. Available from: https://doi.org/10.1534/genetics.112.145037

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet* 2:2074–2093.

Pitulko V V, Nikolsky PA, Girya EY, Basilyan AE, Tumskoy VE, Koulakov SA, Astakhov SN, Pavlova EY, Anisimov MA. The Yana RHS Site: Humans in the Arctic Before the Last Glacial Maximum. Available from: https://www.science.org

Pitulko V V., Tikhonov AN, Pavlova EY, Nikolskiy PA, Kuper KE, Polozov RN. 2016. Paleoanthropology: Early human presence in the Arctic: Evidence from 45,000-year-old mammoth remains. *Science (1979)* 351:260–263.

Pritchard JK, Di Rienzo A. 2010. Adaptation - Not by sweeps alone. *Nat Rev Genet* 11:665–667.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.

Puzyrev V, Stepanov V, Golubenko M, Puzyrev K, Maximova N, Kharkov V, Swarovskaya M, Nogovitsina A. 2003. MtDNA and Y-Chromosome Lineages in the Yakut Population. *Russ J Genet* 39:816–822.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. *Nature* 461:489–494.

Ribeiro MO, Carvalho SD, Schultz JJ, Chiellini G, Scanlan TS, Bianco AC, Brent GA. 2001. Thyroid hormone–sympathetic interaction and adaptive thermogenesis are thyroid hormone receptor isoform–specific. *Journal of Clinical Investigation* 108:97–105.

Sato H, Nagashima K, Ogura M, Sato Y, Tahara Y, Ogura K, Yamano G, Sugizaki K, Fujita N, Tatsuoka H, et al. 2016. Src regulates insulin secretion and glucose metabolism by influencing subcellular localization of glucokinase in pancreatic β-cells. *J Diabetes Investig* 7:171–178.

Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, et al. 2022. Database resources of the national center for biotechnology information. *Nucleic Acids Res* 50:D20–D26.

Sazzini M, Abondio P, Sarno S, Gnecchi-Ruscone GA, Ragno M, Giuliani C, De Fanti S, Ojeda-Granados C, Boattini A, Marquis J, et al. 2020. Genomic history of the Italian population recapitulates key evolutionary dynamics of both Continental and Southern Europeans. *BMC Biol* 18.

Sellinger TPP, Abu-Awad D, Tellier A. 2021. Limits and convergence properties of the sequentially Markovian coalescent. *Mol Ecol Resour* 21:2231–2248.

Shore AM, Karamitri A, Kemp P, Speakman JR, Graham NS, Lomax MA. 2013. Cold-Induced Changes in Gene Expression in Brown Adipose Tissue, White Adipose Tissue and Liver. *PLoS One* 8.

Suguro M, Yoshida N, Umino A, Kato H, Tagawa H, Nakagawa M, Fukuhara N, Karnan S, Takeuchi I, Hocking TD, et al. 2014. Clonal heterogeneity of lymphoid malignancies correlates with poor prognosis. *Cancer Sci* 105:897–904.

Sun MZ, Kim JM, Oh MC, Safaee M, Kaur G, Clark AJ, Bloch O, Ivan ME, Kaur R, Oh T, et al. 2013. Na+/K+-ATPase b2-subunit (AMOG) expression abrogates invasion of glioblastoma-derived brain tumor-initiating cells. *Neuro Oncol* 15:1518–1531.

Sun XY, Hayashi Y, Xu S, Kanou Y, Takagishi Y, Tang YP, Murata Y. 2011. Inactivation of the Rcan2 Gene in mice ameliorates the age- and diet-induced obesity by causing a reduction in food intake. *PLoS One* 6.

Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat Genet* 49:303–309.

Tian Y, Peng B, Fu X. 2018. New ADCY3 Variants Dance in Obesity Etiology. *Trends in Endocrinology and Metabolism* 29:361–363.

Tsibulnikov S, Maslov L, Voronkov N, Oeltgen P. 2000. Thyroid hormones and the mechanisms of adaptation to cold. Available from: https://doi.org/10.1007/s42000-020-00200-2

Utsunomiya YT, Milanesi M, Utsunomiya ATH, Ajmone-Marsan P, Garcia JF. 2016. GHap: An R package for genome-wide haplotyping. *Bioinformatics* 32:2861–2862.

Vallerand AL, Perusse F, Bukowiecki LJ. Cold exposure potentiates the effect of insulin on in vivo glucose uptake.

Wang X, Wahl R. 2014. Responses of the insulin signaling pathways in the brown adipose tissue of rats following cold exposure. *PLoS One* 9.

Xu Z, You W, Zhou Y, Chen W, Wang Y, Shan T. 2019. Cold-induced lipid dynamics and transcriptional programs in white adipose tissue. *BMC Biol* 17.

Yau WW, Yen PM. 2020. Thermogenesis in adipose tissue activated by thyroid hormone. *Int J Mol Sci* 21.

Yu Ziwei, Zhang T, Yang X, Xu B, Yu Zhi, An L, Xu T, Jing X, Wang Y, Lu M. 2024. Neuregulin4-ErbB4 signalling pathway is driven by electroacupuncture stimulation to remodel brown adipose tissue innervation. *Diabetes Obes Metab* [Internet] 26:3880–3896. Available from: https://doi.org/10.1111/dom.15735

Yuan X, Zhang J, Ma TT, Zhuang RJ, Lei BB, Wang L, Cheng BF, Wang M, Yang HJ. 2021. Expression regulation of cold-inducible protein RBM3 by FAK/Src signaling for neuroprotection against rotenone under mild hypothermia. *Biochem Biophys Res Commun* 534:240–247.

Zekri Y, Flamant F, Gauthier K. 2021. Central vs. Peripheral action of thyroid hormone in adaptive thermogenesis: A burning topic. *Cells* 10.

Zhang X, Hoffmann B, Tran PB-V, Graupner G, Pfahl M. 1992. Retinoid X receptor is an auxiliary protein for thyroid hormone and retinoic acid receptors. *Nature* [Internet] 355:441–446. Available from: https://doi.org/10.1038/355441a0

Zhao Wentao, Ouyang C, Zhang L, Wang J, Zhang J, Zhang Y, Huang C, Xiao Q, Jiang B, Lin F, et al. 2024. The proto-oncogene tyrosine kinase c-SRC facilitates glioblastoma progression by remodeling fatty acid synthesis. *Nature Communications* 15.